Proceedings of 2020
Asia-Pacific Regional Conference (Online)

December 12-13, 2020
School of Information Management, Wuhan University
Wuhan, China
# Contents

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visualization of Emergency Needs Posted on Social Media by Metaphor Map</td>
<td>1</td>
</tr>
<tr>
<td>Xiaoyue Ma, Kecheng Cui, Chuwei Ji, Pengzhen Xue</td>
<td></td>
</tr>
<tr>
<td>Rong Tang, Jie Jiang</td>
<td></td>
</tr>
<tr>
<td>Factors Influencing the Health Behavior During Public Health Emergency: A Case Study on Norovirus Outbreak in a University</td>
<td>27</td>
</tr>
<tr>
<td>Shijuan Li, Qian Jiang, Pengyi Zhang</td>
<td></td>
</tr>
<tr>
<td>Examining User Perception and Usage of Voice Search</td>
<td>40</td>
</tr>
<tr>
<td>Ning Sa, Xiaojun (Jenny) Yuan</td>
<td></td>
</tr>
<tr>
<td>Examination of Effects of Time Constraint and Task Type on Users' Query Typing Behaviors</td>
<td>48</td>
</tr>
<tr>
<td>Hanqi Peng, Yanyi Qian, Chang Liu</td>
<td></td>
</tr>
<tr>
<td>A Comparison Study of Measures to Quantify the Evolution of Prolific Research Teams</td>
<td>56</td>
</tr>
<tr>
<td>Bentao Zou, Yuefen Wang</td>
<td></td>
</tr>
<tr>
<td>Exploring Significant Characteristics and Models for Classification of Structure Function of Academic Documents</td>
<td>65</td>
</tr>
<tr>
<td>Bowen Ma, Chengzhi Zhang, Yuzhuo Wang</td>
<td></td>
</tr>
<tr>
<td>How Users’ Gaze Behavior Is Related to Their Quality Evaluation of a Health Website Based on HONcode Principles?</td>
<td>75</td>
</tr>
<tr>
<td>Qin Qin, Qing Ke, Jia Tina Du, Yushan Xie</td>
<td></td>
</tr>
<tr>
<td>Exploring Public Response to COVID-19 on Weibo with LDA Topic Modeling and Sentiment Analysis</td>
<td>86</td>
</tr>
<tr>
<td>Runbin Xie, Samuel Kai Wah Chu, Dickson Kak Wah Chiu, Yangshu Wang</td>
<td></td>
</tr>
<tr>
<td>Cross-Language Fake News Detection</td>
<td>100</td>
</tr>
<tr>
<td>Samuel Kai Wah Chu, Runbin Xie, Yanshu Wang</td>
<td></td>
</tr>
<tr>
<td>Exploring Occupation Differences in Reactions to COVID-19 Pandemic on Twitter</td>
<td>110</td>
</tr>
<tr>
<td>Yi Zhao, Haixu Xi, Chengzhi Zhang</td>
<td></td>
</tr>
<tr>
<td>Public Sector Employee Perspective towards Adoption of E-Government in Pakistan: A Proposed Research Agenda</td>
<td>119</td>
</tr>
<tr>
<td>Nisar Ahmad, Muhammad Waqas, Xiaojuan Zhang</td>
<td></td>
</tr>
<tr>
<td>Adoption of E-Services and Quality of Life Among Older Consumers in China</td>
<td>125</td>
</tr>
<tr>
<td>Muhammad Waqas, Nisar Ahmad, Jiang Wu</td>
<td></td>
</tr>
</tbody>
</table>
Factors Influencing User Behavior Intention to Use Mobile Library Application: A Theoretical and Empirical Research based on Grounded Theory ................................................................. 131
Junren Ming, Rong Chen, Ruide Tu

International Cooperation Among Artificial Intelligence Research Teams Based on Regional Cooperation Models ................................................................................................................. 147
Jiajun Cao, Yuefen Wang

Language and Intercultural Information Ethics Concepts: A Preliminary Discussion of Privacy ............................. 159
Yuanye Ma

Examining Cooking Hobbyists’ Information-Seeking Behavior in Different Situations and in Different Stages from a Serious Leisure Perspective ............................................................. 167
Tien-I Tsai, Hsuan-Yi Chen

Quality Assessment for Digital Stories by Young Authors ......................................................................................... 174
Joana K. Y. Tse, Stephanie W. Y. Chan, Samuel K. W. Chu

Analysis of User Social Support Network in Online Tumor Community ......................................................................... 184
Hui Lin, Shijuan Li

Research on Influencing Factors of Personal Information Disclosure Intention of Social Media in China ...... 195
Anrong Fan, Qiao Wu, Xiaofei Yan, Xiaotong Lu, Yue Ma, Xue Xiao

An Empirical Study on the Cueing Effect of the Emotional Post Title in a Virtual Community ....................... 208
Guihua Li, Siyan Lin, Weiping Yu, Sang Guo

Information Search Trail Recommendation Based on Markov Chain Model and Case-based Reasoning .... 228
Afeng Wang, Yiming Zhao, Yijin Chen

Place, Practice, and Flow: Information Practices in the Mahamevnawa Buddhist Monastery ............................... 242
Pethigamage Perera, Michael Olsson
Visualization of Emergency Needs Posted on Social Media by Metaphor Map

Abstract: Thematic analysis based on a social network is one of the effective means in emergency management. To improve users’ understanding, the results of the thematic analysis are often displayed through visualization. However, the previous researches on text theme visualization rarely considered the unified representation of theme structure and theme evolution. Cognitive load leads to difficulty, and it happens due to the separate representation of structure and evolution relationship and it still increases because of the characteristics of urgency, uncertainty, etc. Therefore, a metaphor map is introduced in this study to overcome the limits of previous visualization tools in characterizing the structure and evolutionary relationship of emergency. On the one hand, different elements in the metaphor map represent the information of popularity and structure of the demand, respectively. On the other hand, the visual design based on the metaphor map strengthens the representation of the content and evolution states of the demand themes. At the theoretical level, a visualization method for emergency needs based on the metaphor map is proposed in this study, which enriches the theory of emergency information visualization. At the practical level, this study explores the design of a visualization system based on the metaphor map under crisis scenarios, which enhances the interaction between users and crisis information, and provides references for decision-making such as emergency material scheduling and emergency resource coordination.

Keywords: human-computer interaction, emergency, semantic visualization, social media, theme evolution

1 Introduction

Disaster relief material management is an important part of emergency management. In the emergency process, a series of decision-making activities is required for material management, including demands estimation (type and quantity) (Sun et al., 2018), materials distribution (Ferrer et al., 2018), and transportation planning (Gralla & Goentzel, 2018). For the overall design and implementation of emergency response, it is necessary to determine the required material reserve according to information such as the type of crisis and disaster (Safaei, Farsad, & Paydar, 2020). So far, the main research has focused on how official organizations determine the types and quantity of disaster relief supplies based on present crisis information. However, considering the efficiency of emergency organizations, economic and social factors, and other issues, it is difficult for official organizations to respond to the disaster victims in who need timely attention in case of emergency. Even if some feedback is obtained, it is still difficult to make timely decisions due to communication delay and lack of organizational efficiency. As for incomplete and untimely feedback, social network analysis has become an effective means to compensate for this drawback. The application of relevant data in social media for analysis can effectively assist the emergency personnel in situational awareness and analysis and help the emergency organizations to prepare, respond, and recover from disasters (Kirac & Milburn, 2018). However, the texts in social media used for analysis are short and noisy, which has caused certain
difficulties for managers to analyze and recognize. The results of the thematic analysis are scattered and complex, and it is difficult for the emergency personnel to focus on and dig deep into the important information. Therefore, many studies have focused on semantic visualization and the theme evolution visualization based on social media data in emergencies.

As mentioned above, the demand analysis for emergency supplies in previous studies is mainly based on the inherent information of the crisis event itself (Hu, 2011), which can meet the requirements of most situations, but lacks a certain degree of flexibility. To compensate for the shortcoming, some studies have conducted a content analysis of social media data. Through the analysis of real-time text information, the flexibility of analysis and decision-making is improved to a certain extent. However, few analyze the content and the evolution process of the needs that remain in existing researches, leading to the lack of overall awareness of emergency needs for emergency personnel. Also, because the previous visualization methods used in social media are difficult to effectively show the phase characteristics and relationships in the process of demand evolution, a relatively effective way is also needed to visualize the content and the evolution of the need to improve people's intuitive cognition and data.

Because of the above problems, this study proposes a method for visualizing crisis needs in social media based on a metaphor map in crises. First, this study collects and categorizes data related to needs arising out of a crisis in social media, and designs metaphors for different needs based on themes and contents. Finally, the metaphor map for crisis is constructed based on social media data. The method is verified by experimental evaluation and questionnaire. Specifically, the questions of this study are mainly reflected in the following two aspects:

RQ1: How to analyze the content and theme evolution of emergency needs in social media in a unified and coordinated way?

RQ2: How to realize the holistic visualization of the characteristics and relationships in different stages of the demand evolution process?

Theoretically, this study put forward a method of emergency demand visualization based on a metaphor map in a crisis, which enriches the theory of emergency information visualization and emergency interaction. Practically, this study explores the design of the visualization system based on the metaphor map in a crisis, enhancing the interaction between users and crisis information, providing references for emergency material scheduling, emergency resource coordination, and other emergency decisions.

The purpose of this study is to introduce the concept of the metaphor map into the visual design of social media data related to emergency demands. In particular, the visualization of the demand-related theme in crises can enhance the emergency managers' understanding of the structure and evolution of crisis demands. The visualization method of emergency demand based on the metaphor map proposed in this paper can not only enhance the users' overall cognition of the topic structure, content, and evolution process, but also explore the feasibility of building visualization components based on the metaphor map to improve interface interaction. It provides references for the construction and visualization design of the text thematic metaphor map.

2 Related Works

2.1 Thematic Analysis of Social Media in Emergency

With the rapid development of the Internet and social media, social media technology provides another communication mechanism for the public in emergencies. As a widely used information medium, social media is often used to release disaster response information (Kim, Bae, & Hastak, 2018). Officials have begun to post information on multiple social media platforms (Tang et al., 2015). Social media analysis, an effective emergency response means, has also attracted more and more attention (Kim & Hastak, 2018). At present, many studies have explored how to apply social media in emergency management from different perspectives, including social network structure analysis, short text emotion analysis, semantic perception, and topic evolution analysis.

Social media theme analysis is a common technology for studying information dissemination and emergency response, which can be used to formulate relevant emergency strategies (Xu et al., 2019) and help identify and analyze behavior and information characteristics of the public (Sarkar et al., 2020). For example, during an emergency response, the information disseminated on social networks may contain rumors, false or true but outdated information (Zhao et al., 2014), and emergency managers can use rumor detection mechanisms to help them identify false information. Some studies focus on the classification and evolution of social media text topics, such as the application of document clustering
and topic modeling techniques to classify, annotate, and understand the content generated by a large number of users (Curiskis, Drake, Osborn, & Kennedy, 2020), and the evolution of the theme of the event can be sorted out by building a co-word network (Wei et al., 2020). However, the information flow in social media is very large in the process of collecting relevant data. For a specific event, the messages collected through predefined keywords will contain a lot of irrelevant information. Using distributed learning can achieve good results in identifying informative posts from a large number of non-informative posts (Ghafarian & Yazdi, 2020). An automatic keyword generation mechanism can also be used to enable managers to track event topics more accurately along with events in data collection (Zheng & Sun, 2019). Social media data can also be combined with the content, sentiment, and social network analysis to help emergency responders plan rescue work and find people in need of medical and emotional assistance (Kiatpanont, Tanlamai, & Chongstitvatana, 2016; Wu & Cui, 2018). Social network text mining contributes to finding valuable hidden information to determine the priority of the areas to be rescued and the specific problems that those who have been influenced are experiencing (Deng, Liu, Zhang, Deng, & Ma, 2016). If managers ignore or lack an understanding of how to communicate through social media (Ki & Nekmat, 2014) during an emergency, it may lead to improper handling of the situation.

2.2 Visualization of Social Media Information Semantics

As a source of temporal and spatial information, social media can provide fast, instant real-time information about events, which contributes to better situational awareness and decision-making (Guan & Chen, 2014). Many social media visualization applications rely on geographic and temporal characteristics, and maps are widely applied in emergency management as an effective spatiotemporal information representation tool (Cervone, Sava, Huang, Schnebele, Harrison, & Waters, 2016). As the number and types of big data streams generated an increase in emergencies, so does the need for systems to support improved situational awareness. Some systems begin to use the semantics of data to enhance visualization such as perceiving the emergency from multiple perspectives (Onorati, Diaz, & Carrion, 2019), creating a message flow graph through comments and other features, analyzing the evolution of keywords and emotions over time (Neppalli, Caragea, Squicciarini, Tapia, & Stehle, 2017) and analyzing the structure between nodes through tree visualization (Li, Tian, Xu, McGuffin, & Yuan, 2020). Although this method of analyzing the subject or emotional evolution describes the evolution process, how to visualize the structural features and relationships of each stage without leading to more cognitive difficulty in the evolution process is still one of the main problems of research.

In the information visualization field, metaphor can help users analyze and understand information effectively by mapping the relationship between visual stimuli and semantics. More and more studies have proposed the application of metaphor in the interface design and information visualization, such as specifying the color of semantic resonance to represent data (for example, blue represents ocean data, green represents plants) (Lin, Fortuna, Kulkarni, Stone, & Heer, 2013), using brick wall metaphor in the software to indicate firewall warning (Raja, Hawkey, Hsu, Wang, & Beznosov, 2011), and designing traffic signs (Caldas, Moreira, Rebelo, & Rossetti, 2018). Different from its interface icon design, metaphors provide information for logical reasoning and involve complex information processing activities in information visualization such as identifying specific patterns (Pramanick & Mitra, 2017) and making judgments (Kale et al., 2018). Because people have limited ability to process given information in a limited time (Ziemkiewicz & Kosara, 2008), metaphor can automatically activate a series of available related concepts to make unfamiliar information easy to explain (Landau, et al., 2010). Metaphor relies on the analogy between source and target concepts. The target concept represents some specific, familiar, tangible, and simple general knowledge. In contrast, the source domain expresses abstract, unfamiliar, intangible, or complex references.

Maps, a rich source of visual metaphors, have been used by many researchers to describe virtual spaces and express abstract non-spatial data in recent years (Xin, Ai, & Ai, 2018). This kind of map visualization exploits the map form to represent non-spatial data and takes advantage of the cognitive convenience of maps (Skupin, 2000). People’s more direct sensory experience of spatial objects is employed to represent abstract concepts. Some studies use the self organizing map (SOM) method to cluster semantic data and try to express the differences between semantics in the form of maps (Skupin, 2002), and discuss the mapping research of transforming metaphors into non-geographical visualization (Skupin & Fabrikant, 2003). This kind of cartographic research that transforms metaphors into non-geographical visualization has been applied to the fields of network communication analysis.
(Fowler, et al., 2014) and video content (Ma, Liu, , Zhao, & Wang, 2016). In the field of social media, a metaphor map generation method is used to analyze the evolving themes in real-time Twitter messages (Gansner, Hu, & North, 2013). Later, some studies also proposed a variety of metaphor maps to allow users to analyze the evolution of events on social media (Chen, Chen, Lin, Yuan, Liang, & Zhang, 2017; Chen, Chen, Wang, Liang, Wu, & Yuan, 2019) and the user’s reposting behavior on social media (Chen, Li, Chen, & Yuan, 2020).

2.3 Motivation

In previous studies, the demand analysis for emergency supplies was mainly based on the existing information about emergency events such as determining the type of materials according to the type of the event and determining its quantity according to the scale of the event. In previous studies, the demand analysis for emergency supplies was mainly based on the existing information about emergency events such as determining the type of materials according to the type of the event and determining its quantity according to the scale of the event. Due to the complexity of emergency material management, unsynchronized communication between organizations and the efficiency of collaboration, there are larger errors and delays in the official forecast statistics compared with the real needs of the victims. These defects have been improved after the introduction of social media to provide feedback. Due to the complexity of social media analysis, visualization tools are often used to show the analysis results. But few studies have paid attention to the visual representation of the demand distribution structure and evolution relationship of each stage at the same time. The above problems are specifically reflected as follows: (1) For the demand related statistical analysis in social media, there is no simultaneous analysis of the evolutionary relationship and content of the demand theme. The incomplete information causes relevant emergency personnel to lack an overall understanding of the actual demand situation; (2) There is a lack of visual representation of the demand theme and the evolution process in social media, and the text-based representation method leads to a lack of intuitive awareness of the entire process by the emergency personnel. The existing visualization methods do not consider how to balance the overall representation of specific evolution relationship and demand distribution.

In response to the above problems, on the one hand, combined with demand content analysis, this study analyzes the demand evolution process during the occurrence of a general emergency event based on the theme evolution stage in the social media platform to provide references for decision-making in emergency supplies management. On the other hand, the use of metaphor maps in the research simultaneously visualizes the content and evolution of requirements and carries out related interaction design, which decreases the difficulty of information cognition and enhances the understanding of actual needs in an emergency. This study plans to re-design the stage division method of emergency demand topics that can be used for metaphor map visualization and expands the information interaction based on metaphor map. The visualization method of emergency demand based on social media and the metaphor map proposed in the research is helpful to the cognition of the dynamic evolution process of emergency demand and emergency decision-making based on the overall cognition of demand.

3 Metaphor Map Design Procedure

The theme of communication of emergency needs in social media is sudden, derivative, and easy to spread. This section mainly introduces the design procedure of the thematic metaphor map about emergency needs. Before the construction of the metaphor map, this study first analyzes the evolution characteristics of the emergency demand theme. Based on the analysis of the theme evolution characteristics, we explore the needs to be met by the metaphor map. The last part describes the main ideas of metaphor map design.

3.1 Evolution Characteristics of Emergency Demand Theme

Sina Weibo, a social media platform similar to Twitter, was selected as the source of demand theme analysis data. Each microblog is limited to 140 words. Users can post content and topics independently. Content can be re-posted or commented on by users. The theme evolution in social networks under emergencies has the following characteristics:

Multi-stage. Emergency has obvious characteristics of time sequence changes: brewing, diffusion, outbreak, climax, and extinction. The theme evolves with the development of the event, reflecting the gradual change process between different stages of the event. At each stage, there are needs themes and key users. In the present
study, we improve the division of the stage evolution, which is suitable for the dynamic representation of metaphor. The emergence of abnormal evolution means that the evolution of the demand theme enters another stage.

**The important impact of key users.** A key user refers to a user who has a large number of fans and a large number of related Weibo re-posts and comments. Users with greater influence in social media can easily promote the change and evolution of demand themes.

**Easy to spread.** Themes are easily spread across regions, and the evolution process is extremely uncontrollable.

**The popularity and content change over time.** The process of the theme evolution can be divided into two types as follows. (1) Normal evolution: the evolution of the theme has gone through five stages: brewing, diffusion, outbreak, climax, and extinction, without the occurrence of abnormal situations; 2) Abnormal evolution: a new demand theme has been introduced under the general theme of emergencies, and different demand themes are merged. A single demand theme is divided into two or more themes, or the original demand theme that tends to be less focused has obtained more attention again.

### 3.2 Emergency Theme Modeling and Analysis

Based on the sequence rule of event evolution, this paper analyzes the evolution process of emergency events subject based on DTM (dynamic topic model) model from two key points: theme content and theme heat. In this paper, the themes of multiple evolution directions are merged into one evolution direction according to the time sequence.

DTM is based on the premise that the event keywords will change with the evolution process, and we can assume that the state of keywords depends on the keywords in the previous time slice. The model determines the different stages of events by the state of keywords in different time slices. This model adds time tags and compares the evolution of subject keywords under the two-time slices, which is an effective way to deal with temporal text analysis. Therefore, it can be used to identify the development stages of the emergency demand theme.

In this paper, the incremental clustering algorithm is used to detect the topic, and the time window is divided by the first discretization to realize the smooth transition of temporal association and content in the process of topic evolution. In the process of emergency demand subject detection, the change of the data input sequence has a great impact on detection results. Nevertheless, the Single-Pass algorithm divides the clusters according to the similarity threshold, and the number of demand topics cannot be known or the number of themes is not a fixed value in the dynamic emergency. So it is more suitable for the detection of emergency demand themes.

The theme can be described by extracting the hot words with a high frequency of words. The change of the attention level of words can be calculated by the positive difference of probability in the interval period. The words with lower results are considered to be the more stable needs in the process of emergency event evolution. The words with higher results are considered to be the content of demand that new needs have been generated or have been met. By analyzing the changes in the content of the theme words in the same topic in different periods, the reasons for the evolution of topic heat can be found from the fine-grained level.

The model has a good effect on the evolution of emergency demand themes in social media through preliminary data experiments. However, the research on this part is still in progress. We will provide open experimental results and relevant analysis of time sequence evolution, theme evolution context, topic evolution heat index, and so on.

### 3.3 Design Requirements of Metaphor Map

The purpose of the metaphor map is to enable interactive exploration and analysis of the evolution of emergency demand theme structure, the impact of key users, and demand theme semantics based on social media data. The study summarizes the following requirements regarding the evolution of the visual emergency needs theme:

1. Easily identify the most important demand theme in each phase. If a demand theme has more microblogs related to it or higher topic popularity, it is considered that the material is more urgently needed.

2. Based on the theme stage division, it is easy to identify which topics are emerging, ending, continuing, or standalone. A topic flow is defined as a path across stages between a topic and its related topics. Four terms are involved here:
   - *emerging*: Topics that were not discussed in the previous stage. (i.e., there was no similar demand theme in the previous phase).
   - *ending*: A keyword to the topic will not appear in the next stage (i.e., there is no similar theme in the next stage).
• continuing: Themes that have appeared before and after this phase.
• standalone: A theme that is not related to any topic in the previous or next phase.

3) Skim through the details of the selected stage. This information includes the most frequently occurring keywords and all microblogs related to the selected themes.

4) Pick specific words to determine the evolution of the demand in the whole emergency. Users may be interested in how one or more keywords are displayed throughout the data set. By identifying the topics associated with these keywords, users can understand how demands change in the entire data set.

3.4 Metaphor Map Construction

The metaphor map proposed in the study comprehensively represents the theme structure, impact of key users, and theme semantics related to emergency needs in social media data. These concepts are mapped to different metaphors on the map:

Continent or island: A whole continent represents the same type of demand themes, and different types of demand themes are separated by oceans. The larger the area of the continent, the more users post about this type of demand on social media. The islands show that the types of needs posted by users are relatively rare, and the theme semantics is quite different from other topics, having difficulties in merging with other types of needs.

Country: The continent on the map is composed of multiple countries. The country represents the demand for each specific substance of the same type. The larger the country is, the greater the proportion of the demand will be. The borders between countries are indicated by thicker borders.

Town: Each polygon on the map represents a town. It is mapped to Weibo user nodes with the same or similar demand keywords in the demand topic structure. The towns with greater influence are highlighted with black strokes.

Highway: Highways stand for the needs of microblogs posted by key users because roads can connect cities and towns on the map, and we can use one-way highways to show the impact of key users on other users, such as common users repost and comment on the Weibo of key users. The longer the highway mileage means the greater the semantic gap with the original Weibo in the reposting or commenting process.

First, to characterize the semantic layout of the demand theme at different stages, the weighted values of the different demand themes in each stage are calculated in the study. Second, the area of each continent or island in the process of metaphor map construction and layout is determined by the data set of topic classification. Finally, the location of key nodes is determined according to the area and location of the continent in the map, and the rest of the continent is filled with ordinary nodes, and there may be several key nodes in each country. If there are many roads in a country, the topic related to this type of demand is mainly dominated by opinion leaders (as shown in Figure 2).

4 Evaluation

To verify the usability of the metaphor map for the visualization of emergency demand thematic analysis, this chapter conducts a preliminary usability study on the metaphor map through experiments and questionnaires.
4.1 Experimental Design

The experiment recruited 16 participants (7 women and 9 men) between the ages of 21 and 29. Participants were familiar with Weibo differently: three of the participants used Weibo for 6 to 10 hours per week, three participants for 2 to 5 hours per week, and the rest for less than 2 hours per week. The study was conducted on 2,347 Weibo datasets collected from 1 January to 25 February of 2020 (the search keyword is “Snow Storm”). The obtained data was filtered and classified. Since the metaphor map has not yet become a widely used tool for visualization, the experiment did not set up a control group. After receiving the relevant training, the participants accepted and started the experimental task:

1) which requirement topic appears most frequently in the first stage?
2) in the second stage, what are the requirements keywords for the least frequent topics?
3) what types of requirements are emerging in phase 3?
4) what requirements exist in almost all the time span?
5) which needs throughout the time span are dominated by opinion leaders?
6) which two demand themes have the highest similarity

Participants scored a Likert score with a total score of 20 according to four indicators of the NASA task load index (Hart & Staveland, 1988): performance, effort, frustration, and effectiveness of the tool (the higher the score, the better the performance). Scores above 18.0 are considered excellent, 15.0–17.9 above average, 12.0–14.9 Medium, below 12.0 points poor. Each session lasts 20 minutes. After the mission, each participant needs to complete a questionnaire. In the end, the participants commented on how they felt about the metaphor map interface.

4.2 Experimental Results

Experimental results show that the metaphor map interface design is helpful for users to recognize information quickly. Participants took longer time (30–50 seconds) to finish two tasks that contrasted cognition. Task 6 deals with the analysis of metaphor maps at all stages where the participants spent the longest time (average 112 seconds).

Task load index: the task load index level reflects the execution results of each task. Task 1, Task 3, and Task 5 scored well on all indicators (higher than 18.0). Task 2 and task 4 scored above average (between 15.1 and 17.8), task 6 indicators scored the worst results (as in table 1). Due to the lack of visualization tools for similarity in metrication matters, participants usually have uncertain answers to task 6, which is one of the reasons why they spent more time.

The experiment also analyzed the qualitative comments in the results of the questionnaire. Users approve of the application of a metaphor map for visual analysis of emergency needs. One participant thought...
that interactive responses helped them to quickly understand the topics and proceed by clicking on the map plate that shows the details of the corresponding Weibo. One participant commented: “Color is very helpful in distinguishing the evolutionary stages in which the subject is located.” It is also indicated that the keywords superimposed on the map are helpful, reducing the burden of users switching among different contexts. One user suggested that other visualization tools be provided to assist in the analysis of the evolution of the topic, and now the timeline visualization still requires users to visually search each stage of the metaphor map plate for each topic, which increases the mental burden.

This experiment preliminarily verified the usability of the visualization system for exploring Weibo datasets, and the deep-seated experiment is still in progress. In the following research, we will explore in-depth including 1). Explore whether each question tested all the different parts of the interface and what cognitive impact the visual interface produced; 2). Explore whether the visualization method based on metaphor map has obvious advantages over other visualization methods such as the topic flow and metaphor map. Not only does the study achieve the expectation of the differentiated representation of demand theme scale and type in social media, but it also characterizes the evolutionary relationship such as the continuing, standalone, emerging, and ending of themes.

Acknowledgments: This work was supported by the National Natural Science Foundation of China (71403201), The Ministry of Education in China, Project of Humanities and Social Sciences (19YJA870009), the Natural Science Foundation of Shaanxi Province of China (2020JM-056), and Major Theoretical and Practical Problems of Social Sciences in Shaanxi Province (2020Z199).

References


https://doi.org/10.2478/dim-2020-0026
received August 15, 2020; accepted September 15, 2020.

Abstract: In this paper, we report the results of a meta-analysis of 50 publications on international Open Government Data (OGD) practices instantiated via their OGD sites or portals. Specific information about 67 individual countries'/regions' OGD sites was extracted and compared, including the levels of OGDs, the number and types of data formats, the number of datasets, and the number and types of data categories. Upon comparing the data characteristics by types and countries, the top 10 countries based on the number of data formats, datasets, and data categories were presented. Significant correlations were found among individual countries' number of data formats, datasets, and data categories. Follow-up research that examines, confirms, and traces the data processing capacity of international OGDs is currently underway.

Keywords: Open Government Data (OGD), OGD levels, data format, datasets, data categories, meta-analysis

1 Introduction

The Open Government Data (OGD) movement in the USA started in 2009 with the mandate from the Obama administration for transparency and open government (White House, 2009, 2013). Around the world, many countries started their OGD efforts around the same time, including New Zealand's open-data catalog in 2009 (Nayek, 2018), followed by the UK in 2010 (data.gov.uk, n.d.), and Kenya's national open-data site in 2011 (Nayek, 2018). As pointed out by Wang and Shepherd (2020), “OGD emerged as an important, multilateral movement among state administrations by the turn of the 21st century. It represents a paradigm shift in the way that governments shape their relationships with the public.” (p. 101405)

With the proliferation of OGD sites, the research reports and empirical studies about the OGDs around the world have significantly increased. These studies covered the characteristics of OGDs developed by a wide range of countries. However, to date, there have been very few studies aggregating the information about these OGDs to provide a landscape view of global OGD development.

In the present study, we performed a meta-analysis of the characteristics of the data processing capacities of OGDs in different countries/regions, as reported in the publications on OGDs. Specifically, we examined the level of OGDs, the number of datasets, the number and types of data formats, and the number and types of data categories. Additionally, correlational analyses were performed for the previously mentioned categories. The results of the study revealed the most common data formats and categories, as well as the leading countries in which the dataset information is available on OGDs. The meta-analysis helps to establish an understanding of the current status of OGD practices around the world, and provides potential directions for countries to make meaningful progress in their OGD development.
2 Literature Review

2.1 Definitions

There have been numerous definitions or descriptions about the characteristics of open data. According to the Open Data Handbook, “Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and share alike”. (http://opendatahandbook.org/guide/en/what-is-open-data/). To make data “open,” Ayre and Craner (2017) point out that (1) data must be “legally” open; and (2) data must be open in the “technical” sense of the word. Stephens (2011) also describes the key technical and legal factors for open data as the following: (1) permissive licensing, (2) mechanisms supporting data retrieval, and (3) the format of the data. Meanwhile, as declared by the White House (2009), “transparency, participation, and collaboration” are “the cornerstone of an open government.” According to Hellekson (2009), the Open Government Directive “requires the US agencies to, among other things, publish government information online, release high-value datasets in open formats on Data.gov, and create an Open Government website.”

Connecting open data with the open government resulted in the creation of OGDs. Xiao, Jeng, and He (2018) define OGD as follows: “Open government data (OGD) refers to any data or information generated by public bodies, such as all government levels, and are made available to the public to access, reuse, and redistribute without copyright restrictions” (p. 573). The essential aspects of OGD portals or websites, as outlined by Tang, Gregg, Hirsh, and Hall (2019), include “open and non-proprietary datasets and its accessibility by any member of a diverse public” (p. 256). Lv and Ma (2019) further point out that “the OGD concept is a working philosophy that enables citizens to have free access to public data without any restrictions. It aims to make the government more open and accountable, and it is the changing role of the government sector from an information gatekeeper into an information provider” (p. 287). The actual implementation of an OGD is accomplished by creating OGD portals or sites. In this study, we refer OGDs as those sites or portals that enable the general public to access government information or public data.

2.2 International landscape of OGD

Since the worldwide emergence of the OGD portals and sites, numerous research studies on OGD covering a wide range of international OGDs have been published. For example, Saxena (2018e) studied the major drivers and barriers in reusing the datasets under the context of the Philippines’ national OGD portal. In the same year, Saxena (2018a) investigated the OGD initiatives among six Middle Eastern countries, Cyprus, Turkey, Egypt, Iran, Lebanon, and Jordan. Research studies on OGDs either focused on a single country (e.g., Tang, et al., 2019; Wang, Shepherd, & Button, 2019; Wieczorkowski & Paweloszek, 2018) or investigated multiple countries in one study (e.g., Yi, 2019; Nayek, 2018; Lourenço, 2015). As an example of a multi-country investigation, Ham, Koo, and Lee (2019) covered 13 countries from North America, Europe, Asia, South America, and Oceania. Ham et al. (2019) focused on comparing the diverse open data readiness, implementation, and impact among the 13 countries. Examples of a single country study include Tang et al. (2019) on the US states and capital cities’ OGDs, Wang et al. (2019) on the UK OGDs, and Wang, Chen, & Richards (2018) on the Chinese province-level OGDs.

Overall, there has been very limited research cross-examining publications on OGD. Nevertheless, in 2018, Neto, Neves, Santos, Junior, and Nascimento (2018) conducted a Systematic Literature Mapping (SLM) on OGD publications to identify which sectors of public administration have a higher incidence of publication on OGD. They found that “Education, Health, and Finances as public sector areas with Open Data major publication incidence” (p. 1). Furthermore, based on an analysis of 69 publications, they discovered six main barriers to OGD, such as public engagement, culture, economics, politics, technical, and data quality. The authors pointed out that “These barriers mainly deal with the legibility and data access, two main features to datasets reuse” (p. 5).

In reviewing the countries and continents covered in recent publications on international OGDs, we found a total of 67 countries/regions with reported OGDs. Table 1 lists those countries by region.

In summary, there have been numerous empirical studies focused on the OGDs from individual countries as well as multiple countries. However, there is still a lack of synthesis or meta-analysis comparing all the reported data processing capacity of these OGDs across research studies. Although there have been a few meta-analysis studies (e.g., Lv & Ma, 2019; Neto, et al., 2018) that examined the OGDs data across publications, there has been no research exclusively investigating the reported data processing capacity of OGDs by countries in an aggregated manner. The present study intends to address this research gap by answering research questions about OGD levels, data formats, datasets, and data categories.
**3 Research Questions**

The present study attempts to answer the following research questions:

- **RQ1.** Among publications examining the international presence of OGDs, what countries and world regions are reported?

- **RQ2.** With regard to international OGDs under study, what kinds of data formats are available? What are the most common data formats?

- **RQ3.** With regard to the OGD levels and the number of datasets provided, what are the most common level of OGD reported and which countries/regions were reported to have high numbers of datasets?

---

*Two world regions, Europe Union and Africa were reported by Bello et al. (2016) and Khurshid et al. (2019).*

### Table 1
List of the Publications on International OGDs

<table>
<thead>
<tr>
<th>Region</th>
<th>Publications &amp; Countries Covered</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>South America</strong>&lt;br&gt;(n=5)</td>
<td><strong>Argentina:</strong> Bonina &amp; Eaton (2020), Rumbul (2016); <strong>Brazil:</strong> Ham et al. (2019), Machado, Martini, Viterbo, Bernardino, &amp; Barcellos (2018), Oliveira, Oliveira, Oliveira, &amp; Lóscio (2016); <strong>Chile:</strong> Gonzalez-Zapata (2018) [dissertation]; <strong>Columbia:</strong> Quarati &amp; Martino (2019); <strong>Uruguay:</strong> Bonina &amp; Eaton (2020)</td>
</tr>
<tr>
<td><strong>Europe</strong>&lt;br&gt;(n=23)</td>
<td><strong>Austria:</strong> Ham et al. (2019), Radl, Skopek, Komendera, Jäger, &amp; Mödritscher (2013); <strong>Belgium:</strong> de Juana-Espinosa &amp; Luján-Mora (2020); <strong>Bulgaria:</strong> de Juana-Espinosa &amp; Luján-Mora (2020); <strong>Croatia:</strong> de Juana-Espinosa &amp; Luján-Mora (2020); <strong>Czech Republic:</strong> de Juana-Espinosa &amp; Luján-Mora (2020); <strong>Denmark:</strong> Gomes &amp; Soares (2014); <strong>Estonia:</strong> McBride, Toots, Kalvet, &amp; Krimmer (2018); <strong>France:</strong> Ham et al. (2019), Lourenço (2015), Ruijier, Détiennne, Baker, Groff, &amp; Meijer (2020), Quarati &amp; Martino (2019); <strong>Germany:</strong> Ham et al. (2019), Wieczorkowski &amp; Paweloszek, (2018); <strong>Greece:</strong> Alexopoulos, Loukis, Moustakis, Petychakis, &amp; Charalabidis (2018); <strong>Italy:</strong> Gomes &amp; Soares (2014); <strong>Ireland:</strong> Fitzpatrick (2012), Quarati &amp; Martino (2019), Sandoval-Almazan &amp; Styriń (2018); <strong>Latvia:</strong> Quarati &amp; Martino (2019); <strong>Moldova:</strong> Ham et al. (2019); <strong>Netherlands:</strong> Ruijier, Détiennne, Baker, Groff, &amp; Meijer (2020), Safarov (2019); <strong>Norway:</strong> Gomes &amp; Soares (2014), Ham et al. (2019); <strong>Poland:</strong> Quarati &amp; Martino (2019), Wieczorkowski &amp; Paweloszek, (2018); <strong>Portugal:</strong> Gomes &amp; Soares (2014), Ham et al. (2019), Quarati &amp; Martino (2019); <strong>Russia:</strong> Koznov, Andreeva, Nikula, Maglyas, Muromtsev, &amp; Radchenko (2016), Quarati &amp; Martino (2019), Sandoval-Almazan &amp; Styriń (2018); <strong>Slovenia:</strong> Quarati &amp; Martino (2019); <strong>Spain:</strong> Gomes &amp; Soares (2014), Ham et al. (2019), Ubaldi (2019); <strong>Sweden:</strong> Safarov (2019); <strong>UK:</strong> Fitzpatrick (2012), Gomes &amp; Soares (2014), Ham et al. (2019), Khurshid et al. (2019), Lourenço (2015), Nayek (2018), Quarati &amp; Martino (2019), Safarov (2019), Sandoval-Almazan &amp; Styriń (2018), Talukder et al. (2019), Wang &amp; Shepherd (2020), Wieczorkowski &amp; Paweloszek, (2018), Yi (2018); <strong>USA:</strong> Koznov, Andreeva, Nikula, Maglyas, Muromtsev, &amp; Radchenko (2016), Quarati &amp; Martino (2019), Sandoval-Almazan &amp; Styriń (2018); <strong>Japan:</strong> de Juana-Espinosa &amp; Luján-Mora (2020), Saxena-a. (2017), Noda, Yoshida, &amp; Honda (2019); <strong>Jordan:</strong> Saxena-a. (2017); <strong>Kuwait:</strong> Saxena-f (2018); <strong>Lebanon:</strong> Saxena-a. (2017); <strong>Malaysia:</strong> Khurshid et al. (2019), Zainal, Hussin, Rahim, Nazri, &amp; Suhaime (2019); <strong>Oman:</strong> Saxena-d (2018), Saxena-f (2018); <strong>Pakistan:</strong> Khurshid et al. (2019), Saxena &amp; Muhammad (2018); <strong>Philippines:</strong> Saxena-e (2018); <strong>Qatar:</strong> Saxena-f (2018); <strong>Saudi Arabia:</strong> Saxena-f (2018); <strong>Singapore:</strong> Chan (2013), Khurshid et al. (2019), Zhenbin, Kankanathalli, Ha, &amp; Tayi (2020); <strong>South Korea:</strong> Ham et al. (2019), Quarati &amp; Martino (2019), Yi (2018); <strong>Taiwan, China:</strong> Koznov, Andreeva, Nikula, Maglyas, Muromtsev, &amp; Radchenko (2016), Quarati et al. (2019), Quarati &amp; Martino (2019), Wang et al. (2018); <strong>Turkey:</strong> Saxena-a. (2017); <strong>United Arab Emirates:</strong> Saxena-a. (2017), Pardo, &amp; Cook (2014), Saxena-f (2018)</td>
</tr>
<tr>
<td><strong>Asia</strong>&lt;br&gt;(n=23)</td>
<td><strong>Bahrain:</strong> Saxena-f (2018); <strong>Bangladesh:</strong> Talukder, Shen, Hossain Talukder, &amp; Bao (2019); <strong>China (w/HK):</strong> Wang et al. (2018); <strong>Cyprus:</strong> de Juana-Espinosa &amp; Luján-Mora (2019), Saxena-a. (2017); <strong>India:</strong> Ham et al. (2019), Nayek (2018), Quarati &amp; Martino (2019), Saxena-b. (2017); <strong>Indonesia:</strong> Jacob, Fudzee, Salamat, &amp; Rahman (2019); <strong>Iran:</strong> Saxena-a. (2017); <strong>Japan:</strong> Noda, Yoshida, &amp; Honda (2019); <strong>Jordan:</strong> Saxena-a. (2017); <strong>Kuwait:</strong> Saxena-f (2018); <strong>Lebanon:</strong> Saxena-a. (2017); <strong>Malaysia:</strong> Khurshid et al. (2019), Zainal, Hussin, Rahim, Nazri, &amp; Suhaime (2019); <strong>Oman:</strong> Saxena-d (2018), Saxena-f (2018); <strong>Pakistan:</strong> Khurshid et al. (2019), Saxena &amp; Muhammad (2018); <strong>Philippines:</strong> Saxena-e (2018); <strong>Qatar:</strong> Saxena-f (2018); <strong>Saudi Arabia:</strong> Saxena-f (2018); <strong>Singapore:</strong> Chan (2013), Khurshid et al. (2019), Zhenbin, Kankanathalli, Ha, &amp; Tayi (2020); <strong>South Korea:</strong> Ham et al. (2019), Quarati &amp; Martino (2019), Yi (2018); <strong>Taiwan, China:</strong> Koznov, Andreeva, Nikula, Maglyas, Muromtsev, &amp; Radchenko (2016), Quarati et al. (2019), Quarati &amp; Martino (2019), Wang et al. (2018); <strong>Turkey:</strong> Saxena-a. (2017); <strong>United Arab Emirates:</strong> Sayogo, Pardo, &amp; Cook (2014), Saxena-f (2018)</td>
</tr>
<tr>
<td><strong>Oceania</strong>&lt;br&gt;(n=2)</td>
<td><strong>Australia:</strong> Chatfield &amp; Reddick (2017), Ham et al. (2019), Khurshid et al. (2019), Lourenço (2015), Nayek (2018); <strong>New Zealand:</strong> Lourenço (2015), Nayek (2018)</td>
</tr>
<tr>
<td><strong>Africa</strong>&lt;br&gt;(n=11)</td>
<td><strong>Burkina Faso:</strong> Bello, Akinwande, Jolayemi, &amp; Ibrahim (2016); <strong>Egypt:</strong> Saxena-a. (2017); <strong>Ghana:</strong> Afful-Dadzie &amp; Afful-Dadzie (2017), Sayogo et al. (2014); <strong>Kenya:</strong> Afful-Dadzie &amp; Afful-Dadzie (2017), Nayek (2018), Sayogo et al. (2014); <strong>Morocco:</strong> Afful-Dadzie &amp; Anthony Afful-Dadzie (2017), Sayogo et al. (2014); <strong>Nigeria:</strong> Bello et al. (2016); <strong>Senegal:</strong> Bello et al. (2016); <strong>Sierra Leone:</strong> Afful-Dadzie &amp; Anthony Afful-Dadzie (2017); <strong>South Africa:</strong> Afful-Dadzie &amp; Anthony Afful-Dadzie (2017); <strong>Tanzania:</strong> Afful-Dadzie &amp; Anthony Afful-Dadzie (2017), Shao &amp; Saxena (2019); <strong>Tunisia:</strong> Afful-Dadzie &amp; Anthony Afful-Dadzie (2017)</td>
</tr>
</tbody>
</table>
RQ4. In terms of data categories reported on OGD sites, what are most common data categories, and how do individual countries rank in terms of the number of data categories reported on their OGD sites?

RQ5. What are the top 10 countries in terms of their number of datasets, number of data formats, and number of data categories?

RQ6. Are there statistically significant correlations between the number of data categories, number of datasets, and number of data formats by countries?

4 Methods

The present research involves a meta-analysis of international OGDs data reported in a total of 50 publications. The search of publications on OGDs using the search phrase “open government data” was performed in early September 2019 on EBSCO databases, SCOPUS, ACM Digital library, and ProQuest Dissertation, and Thesis Full Text. In November 2019, another round of searching was conducted. In June 2020, the third round of searching was performed. During three rounds of searches, publications that contained specific information on data processing capacities (e.g., number of datasets, number of data format) of individual or multi-country OGDs were selected as our study sample. The final publication samples consisted of 48 journal or conference articles, one master’s thesis, and one doctoral dissertation.

We attempted to keep our publication samples as recent as possible. That was the reason for the three rounds of diligent literature searches to update our samples. Among 50 publications, 31 (62%) were published between 2018 and 2020. A total of 44 (88%) items were published in the past five years. For each publication sample, we examined the content of the specific item, extracted and coded information regarding data formats, the level of OGDs, the number of data sets, and data categories. Table 2 lists the variables we examined in this study. A spreadsheet was used for coding, with different tabs covering the overview, and the specifics of the data format, the OGD levels, the number of data sets, and the labels of data categories mentioned in the publications. For articles covering several countries, there would be multiple rows for recording the dataset specifics of individual countries that were included. Some publications contained specifics for one area of dataset characteristics (e.g., data category labels) but not in other areas (e.g., data format). For those cases, we noted the total number of publications reported a given dataset feature. The coding was checked against the publications multiple times to ensure accuracy. After the coding was completed, we produced a summary of each aspect of the data processing capacity reported in the literature. Then we proceeded with statistical analysis about the correlations of the three data processing capacities, using the country as the unit of analysis.

5 Results

The results of the study are outlined in the following sections addressing each of our research questions.

5.1 Countries and Regions Reported

There was a total of 67 countries and two regions (Europe Union and Africa) that were included in the investigations of these publications. Out of 50 publications, 23 reported information on one country, whereas 27 reported data on multiple countries. For instance, in a study by Kurshid et al. (2019), there was OGD information about Pakistan, the USA, Europe, Australia, Taiwan (China), the UK, Malaysia, and Singapore. Another study published in 2020 by Bonina & Eaton (2020) included the capital cities of three countries: Buenos Aires (Argentina), Mexico City (Mexico), and Montevideo (Uruguay). Table 3 lists the countries by six continents as reported in the 50 publication samples. Figure 1 provides a global map identifying the 67 countries investigated. As it is shown in both Table 3 and Figure 1, each continent was well represented in the global studies on OGDs.
5.2 Data Format

Among 50 research items examined, 29 reported some information on data formats, whereas 21 items did not report details of the data format. A good number of articles reported multiple cases of data formats, for instance, Olivera et al. (2016) reported 13 cases of data formats by Brazilian states. As a result, there were 73 cases of data format reports.

5.2.1 Specific Types of Data Format Reported

Of 73 sets of reports containing data format information, a total of 41 different types of data formats were reported. Figure 2 shows the data format by the frequency of occurrence appearing in the reports. The most frequently occurring data format was CSV (n=65, 95.89%), followed by XLS/XLSX/XL/XLB (n=59, 80.82%), PDF (n=52, 71.23%) and XML (n=43, 58.90%). Other frequently occurred
5.2.2 Data Format by Countries/Regions

In terms of data format by individual countries, the literature reported the data format information for 47 countries/regions. The number one country holding the highest number of data formats in their OGD portals was the USA (n=29), with a total of 29 different types of data format. Other countries that had top number of data format included the UK (n=19), Australia (n=19), Greece (n=17), Brazil (n=13), Korea (n=11), Germany (n=11), Poland (n=10), India (n=10), Chile (n=10). Figure 3 lists the number of data formats by country.

5.3 OGD Levels and Number of Datasets

5.3.1 OGD Levels

Of 50 items we examined, 47 (94%) reported their OGDs levels. The levels of OGDs reported in the literature ranged...
from national, state/province, city, and to the world region level. The majority of the publications examined the OGD at the national government level (n=38, 81.08%), with the remainder either at the city (n=9, 19.15%), the state/province level (n=8, 17.02%) or at the world region level (n=3, 6.38%). Note that more than 80% of the literature items examined reported OGD at the national government level, whereas municipal or state/province levels were reported in less than 20% and 18% of the items, respectively. Table 4 includes the frequency information of the reported OGD levels.

### 5.3.2 Number of Datasets

Of the 50 publications, 14 items did not include information about the number of datasets. Among 36 publications containing dataset information, several items reported multiple cases, as a result, a total of 117 cases were reported in the 36 publications. Among the reported number of datasets, the number ranged from as low as 18 (India-Sikkim, as reported in Saxena [b], 2017) to as high as 450,000 (the USA, as reported in Fitzpatrick, 2012). In addition to individual countries, the existing literature reported the number of datasets in two world regions, which included Europe (avg=12,669) and Africa (avg = 3841.50). The reported OGD levels of datasets originated from either a national, a state/provincial, or city/municipal level. There are also reports (n=3) from two world regions including Africa and Europe.

When looking at the number of datasets by countries, at the national level, the USA was the country with the highest average number of datasets (avg=276951.3). The country that had the second-highest average number of
At the state or provincial level, there are a total of 22 cases containing the datasets information at the state/provincial level from nine countries, with China (avg=2636), the USA (avg=1714.45), and Australia (avg=1523.6) as the top three countries that reported the highest average number of datasets. The lowest number of datasets at the state level is Pakistan (n=36). In terms of the city/municipal level OGD datasets, ten cases reported the city level datasets information from five countries, such as the USA, Brazil, Argentina, Mexico, and Uruguay. The number of datasets ranges from as high as 16,623 to as low as 33. There are six Brazilian city OGDs reported, with an average of 2978, which is the highest number of datasets at the city level. The USA had the second-highest city-level OGD datasets (n=1714.45), followed by Argentina which had the third-highest number of city-level datasets (n=168).

5.3.2.1 Top Countries/regions by Number of Datasets

At the national level, 13 top countries had an average number of datasets of more than 10,000. The top countries that had more than 20,000 datasets were the USA (avg=276951.29), India (avg=140698.5), France (n=35663), the UK (avg=31518.56), Mexico (n=27766), Korea (avg=23699.5), Germany (n=20650), and Taiwan (China) (avg=20144.5). Figure 4 displays the number of datasets by country at the national level with ranges of datasets greater than 10,000.

5.3.2.2 Middle Range Countries by the Number of Datasets

A total of 26 countries had an average number of datasets between 9000 and 100 on their national level OGDs. The countries that had more than 2500 datasets include Ireland (n=9001), Italy (n=8857), Brazil (n=4051), New Zealand (n=3435), Slovenia (n=3389), and Singapore (avg=2935.67). Figure 5 displays the number of datasets by country at the national level with ranges of datasets between 9,000 and 100.

5.3.2.3 Lower Range Countries by the Number of Datasets

A total of 12 countries had an average number of datasets in the lower range, which is less than 100. Among the 12 countries, those that had more than 70 datasets are Sri Lanka (n=89), Estonia (n=89), Jordan (n=88), Tanzania (n=84), UAE (n=78), Morocco (avg=77), and Turkey (n=74). Figure 6 displays the number of datasets by country at the national level with ranges of datasets that are lower than 100.

5.4 Data Categories

5.4.1 Number of Data Categories

Out of 50 items that we examined, only 28 reported the number of data categories on the OGD sites. The 28 publications included a total of 79 cases of reports on data categories.

5.4.2 Data Category Labels

Of the 28 publications that included data category label information, a total of 75 different data categories were reported. Examples of data categories include Education (n=49), Health (n=46), Transportation (n=38), Finance (n=35), Environment (n=33), and Economy (n=31). Other categories include Government (n=29), Science/Technology (n=27), and more. A total of 19 data categories have only one occurrence. Examples of the categories that occurred only once were Credit Records, Domestic Products, Sport, Ocean, Stakeholder, and Gas Drilling. Figure 7 illustrates data categories reported in the literature that were mentioned equal to or above five times.
Figure 4. Top thirteen countries/regions by the number of datasets (datasets >10,000).

Figure 5. Middle range countries by the number of datasets (datasets btw 9,000 and 100).
Among the publications that reported the data categories of various countries/regions, a total of 47 countries/regions had their data categories listed. These range from the USA, Australia, Malaysia, India, and the UK, which reported to have more than 20 data categories, to Chile, Taiwan (China), Morocco, Sweden, and Ghana, which reported to have less than five data categories. Two world regions (Africa and Europe), which do not have specific countries’ information, are also included. Figure 8 shows the number of data categories reported in the literature by countries/regions in descending order.

Additionally, in terms of the frequency of data categories by country, the most common data categories are listed in Table 5. Those categories that were reported as appearing on 50% or more of OGD data include Education, Health (33, 70.21%), Transportation (28, 59.57%), Economy; Finance (26, 55.32%), and Environment (24, 51.06%).

### Table 5

**Data Categories Appearing on OGDs by the Number of Countries**

<table>
<thead>
<tr>
<th>Data Categories by Country</th>
<th>(% of Countries/Regions Containing Data Category)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education; Health</td>
<td>33 (70.21%)</td>
</tr>
<tr>
<td>Transportation</td>
<td>28 (59.57%)</td>
</tr>
<tr>
<td>Economy; Finance</td>
<td>26 (55.32%)</td>
</tr>
<tr>
<td>Environment</td>
<td>24 (51.06%)</td>
</tr>
<tr>
<td>Science/Technology; Tourism</td>
<td>22 (46.81%)</td>
</tr>
<tr>
<td>Government; Law/Justice</td>
<td>19 (40.43%)</td>
</tr>
<tr>
<td>Arts/Culture/History</td>
<td>18 (38.30%)</td>
</tr>
<tr>
<td>Society; Demographic/Population; Administration/Governance</td>
<td>17 (36.17%)</td>
</tr>
<tr>
<td>Public Safety/Security</td>
<td>16 (34.04%)</td>
</tr>
<tr>
<td>Climate; Labor/Employment &amp; Skill; Geography</td>
<td>14 (29.49%)</td>
</tr>
<tr>
<td>Agriculture</td>
<td>13 (27.66%)</td>
</tr>
<tr>
<td>Energy/Natural Resources; Budget &amp; Spending; Other</td>
<td>12 (25.53%)</td>
</tr>
<tr>
<td>Geodata/Maps; International Trade</td>
<td>10 (21.28%)</td>
</tr>
<tr>
<td>Healthcare</td>
<td>9 (19.15%)</td>
</tr>
<tr>
<td>City/State Services; Housing; Statistics</td>
<td>8 (17.02%)</td>
</tr>
<tr>
<td>Communications</td>
<td>7 (14.89%)</td>
</tr>
<tr>
<td>Business; Infrastructure</td>
<td>6 (12.77%)</td>
</tr>
<tr>
<td>Crime</td>
<td>5 (10.64%)</td>
</tr>
</tbody>
</table>

### 5.5.3 Data Category by Countries/regions

Among the publications that reported the data categories of various countries/regions, a total of 47 countries/regions had their data categories listed. These range from the USA, Australia, Malaysia, India, and the UK, which reported to have more than 20 data categories, to Chile, Taiwan (China), Morocco, Sweden, and Ghana, which reported to have less than five data categories. Two world regions (Africa and Europe), which do not have specific countries’ information, are also included. Figure 8 shows the number of data categories reported in the literature by countries/regions in descending order.

Additionally, in terms of the frequency of data categories by country, the most common data categories are listed in Table 5. Those categories that were reported as appearing on 50% or more of OGD data include Education, Health (33, 70.21%), Transportation (28, 59.57%), Economy; Finance (26, 55.32%), and Environment (24, 51.06%).

### 5.5 Top Countries by Data Format, Datasets, and Data Categories

The top 10 countries/regions include 18 individual countries from six different continents across the three data-processing areas, with the USA dominating as number 1 for all three areas, followed by Australia at numbers 2 and 3 for data format and data categories. The UK and India were number two in terms of the number of data formats, and datasets. France and Malaysia were
number three in the number of datasets and the number of data categories. Table 6 lists the top 10 countries/regions by the number of data formats, the total number of datasets, and the number of data categories available on their OGD sites as reported from the publications.

5.6 Correlations

Based on the information we gathered about the individual countries’ OGD’s data characteristics, we considered it meaningful to examine the association between the number of datasets, the number of data formats, and the number of data categories using country
and world regions as the unit of analysis. Spearman’s rank correlation was used because of the fact that the reports of various data characteristics were from different publications and they might lack consistency. Significant Spearman correlations were found between number of data categories and datasets ($r_s(33)=0.605, p=0.000$), between the number of data categories and data formats ($r_s(36)=0.719, p=0.000$), between the number of datasets and data formats ($r_s(34)=0.739, p=0.000$).

6 Discussion and Conclusions

Through performing a meta-analysis of the data processing capacity of international OGDs reported in 50 publications, the results of the study revealed specific characteristics and data patterns of international OGD sites in terms of their OGD levels, data formats, datasets, and data categories. We were not only able to aggregate from the findings of the existing literature relevant information by 67 countries/regions but also able to uncover the specific data format types and data category labels occurring on these OGD sites. In identifying top data formats to be CSV, XLS/XLSX/XL/XLB, PDF, and XML, and top data categories to be education, health, transportation, finance, environment, economy, government, and science and technology, there is a potential to develop a guideline.
or recommendation for building OGDs to include these top formats and categories as the minimum requirement for an OGD site. Figure 9 shows an example based on the top types of data formats, OGD levels, and data categories. This might serve as a baseline guideline for developers as they consider designing and revising an OGD site. Of course, a final guideline would require further thorough examinations of other data processing capacities as well as conversations with OGD sponsors and developers.

Additionally, by listing the top 10 countries/regions by their reported numbers of data formats, datasets, and data categories, it is clear that certain countries such as the USA, Australia, India, and the UK, were among the top five countries across multiple counts. There are some observed disparities of OGD sites’ data provision, even though the top 10 were from six different world regions. There are remarkable variations among world regions in terms of the number of data formats, datasets, and data categories. Through revealing the data disparities on OGDs, this study calls for the action of setting up standards for OGD data provision, which can be based on our recommendations illustrated in Figure 9. It appears that a cross-country global effort is needed to articulate these standards and to ensure that these minimum standards are met as more and more countries are developing, improving, and sustaining their OGD portals.

One of the challenges that we encountered during the data extraction from publications was that not all publications contained the specifics of the data processing capacities of individual OGDs. The reports also vary in terms of the depth and levels of detail of data formats, datasets, data category labels, and so on. We recognized these reported numbers might be incomplete or inconsistent across publications, but we wanted to obtain a landscape understanding of international OGD practices from these publications. Further follow-up research that examines the specific data details from live OGD sites, which compares these details with published literature can help us see the evolution and growth of the international OGD sites.

In the current day and age when the COVID-19 pandemic is spreading around the globe, the commitment to open access to government data, both federal and local, is increasingly crucial. The importance of OGDs containing government information, specifically about key categories, such as health, education, and economy, is higher than ever. This meta-analysis reveals that according to studies on OGDs, such efforts are shown across the world. While certain countries seem to have been leading the effort by incorporating a variety of data formats, a great number of datasets with a magnitude of data categories, others are lacking in multiple areas. There is hope that with these findings and a follow-up study that examines live sites, the best practice of optimal data processing capacity can be cultivated to support the advancement of developing OGD sites. Furthermore,
these OGD sites may be used to provide real-time live virus infection data across all government levels (e.g., federal, state, county, and city) to the stakeholders, constituents, and the general public worldwide.

References


Factors Influencing the Health Behavior During Public Health Emergency: A Case Study on Norovirus Outbreak in a University

Abstract: It is known that health belief and health literacy are closely related to health behavior. But, we do not know explicitly how health belief and health literacy interact with each other and determine health behavior change under public health emergencies (PHE). Through the integration of constructs from health belief model (perceived susceptibility, severity, benefits, barriers, and self-efficacy) and diverse dimensions of health literacy (functional, interactive and critical), a research framework is proposed to examine the underlying mechanism of health behavior change during PHE. Structural equation modeling (SEM) was used to analyze 386 questionnaire data collected from Chinese university students for the research framework. The analysis results show that (1) both health belief and health literacy have significant impacts on health behavior change during PHE. However, health belief plays a mediating role which affects the health literacy’s impact on health behavior; (2) while the increase of perceived severity of disease and self-efficacy promote the health behavior change, the effectiveness of perceived susceptibility on health behavior depends on the increase of perceived severity; and (3) the enhancement of interactive health literacy effectively promotes health behavior change, while functional and critical health literacy reduces the blind change. The results throw lights on health education services and provide references and factors in understanding and encouraging health behavior changes to relevant stakeholders including social media operators, practitioners, social service providers, and policy makers.

Keywords: public health emergency, health behavior, health belief, health literacy, norovirus

1 Introduction

Public health emergency (PHE) exerts great pressure on the healthcare system, economic development, and public’s lifestyle. Based on fighting against SARS in 2003, China has issued the Regulations on Emergency Response to Public Health Emergency, which integrates a series of behavior code for the public into the legal system. People are requested to modify their daily routine behaviors according to the codes, such as hand washing steps, room ventilation, and change in specific activities, such as wearing facemasks. However, the effect of these rules and regulations are not known especially during the outbreak of a PHE.

Past researches on the epidemic patterns of norovirus-infected diarrhea has shown that schools are considered as one of the high-risk places of PHE (Zhou et al., 2020) as students share their classrooms and dormitories in high density and due to their interactive behaviors in various activities like information searching, analyzing, and sharing. During the final days of examination in December 2019, a small outbreak of norovirus infections in a Chinese university forced students’ attention to health issues. Health behaviors including washing hands frequently, flushing water before and after going to the toilet, avoiding the eating in the student canteen were widely adopted for some time, but doubts and objections were also raised at the same time. Topics were discussed widely and repeatedly on social media and BBS about the
safety of health practices, symptoms of the infections, and precautions to be taken.

These issues related to health literacy and health belief influenced students’ decision-making process to choose between maintaining their routine health behavior and modifying their routine behavior during PHE. Past researches have shown that there are close links between health literacy, health belief, health behavior, and health status (von Wagner, Steptoe, Wolf, & Wardle, 2009). However, to the best of our knowledge, no past research has integrated both health literacy and health belief constructs during PHE into one research framework to analyze their interactions. Furthermore, not much is known about the main factors that are driving people to take actions or thinking critically, and the implications does this have on health education services. This research which was conducted during the above norovirus outbreak is primarily aimed to analyze the effects of health belief and health literacy on health behavior change, and their interaction with each other. Finally, we are concerned about what implications does all this have on health education and emergency publicity.

The rest of the paper is organized as follows: we first provide a review of the literature regarding the influencing factors of health behavior in the context of PHE. We then present the research framework and put forward our hypotheses, including instruments and data collection methods, followed by the analysis results. We conclude with a discussion of the key findings, implications, limitations, and directions for future research.

2 Related Work

2.1 Health Behavior Change

The concept of health behavior was different from disease behavior in the earlier times and it referred to the actions taken by healthy individuals to prevent and detect diseases (Kasl & Cobb, 1966). When the interdisciplinary research deepened, the connotation of health behavior extended from diseases prevention to health promotion, emphasizing not only physical actions but also emotional cognition (Gochman, 1997) and social adaptation (Berkman, Glass, Brissette, & Seeman, 2000).

Behavior is so closely related to health that it promotes the development of health education that takes cultivating and changing people’s health behavior (Lv, 2002) as its core. Many achievements have been made in the research of health behavior change during this period, which can be used not only to explain the change and maintenance of health behavior, but also to guide the implementation of health education and health behavior intervention projects (Lippke & Ziegelmann, 2008). Some models focus on a wide range of intra-individual and interpersonal influences, such as individual psychological cognition and social economic status, while others focus on specific aspects of motivation, such as the desire for self-determination (Michie, West, Campbell, Brown, & Gainforth, 2014).

Theories at the individual level are mainly based on the KABP (knowledge-attitude-belief-practice) paradigm, believing that information dissemination and behavior intervention help targeted individuals to grasp scientific health knowledge, develop reasonable health beliefs, so as to take healthier actions. Among them, the Health Belief Model (Hochbaum, 1958; Rosenstock, 1966), Theory of Planned Behavior (Ajzen, 1991), and Transtheoretical Model (Prochaska & Velicer, 1997) are widely applied in China.

2.2 Health Behavior Change in PHE

Public health emergency can be defined as sudden or instant occurrences that endanger or severely endanger human health and social safety, and require urgent response from the perspective of public health and preventive medicine. After the outbreak of SARS, research carried out in China has found that the higher the threat people perceive in PHE and the greater their understanding in relevant knowledge, the more likely they are willing to change their health behavior (Lv, Tian, Yang, & Chen, 2008). There are also some studies arguing that, in the context of PHE, the health behavior change does not follow the traditional KABP pathway, but is built upon the psychological fear and panic. Moreover, the causes of panic mainly depend on the characteristics of disease, the effectiveness of treatment, and the social responses (Li, Ling, & Zeng, 2004).

Other previous studies on health behavior focused more on healthy lifestyles and the management of chronic diseases, and there was lack of discussions on the mechanism of health behavior change in terms of PHE. Current studies have noticed the importance of PHE but only addressed the health knowledge, belief, and behavior of specific subpopulations such as students (Yu et al., 2011) and staff (Gosadi, Al-Hazmi, Fadl, Alharbi,
Factors Influencing the Health Behavior During Public Health Emergency...

It has been found that there is a close relationship between health knowledge, belief, and behavior, though the overall level of health knowledge is fairly low.

Furthermore, studies have analyzed the logic behind the selection of behavior of different people in PHE (Pang, 2019), the impact of the time when individuals first pay attention to the PHE on their behavior (Tan, Zeng, Chen, & Luo, 2020), and the information strategy at different stages (Hu, Ye, Wang, & Shi, 2004), which advanced the studies on factors that drive people to take actions during the event of PHE.

2.3 Influencing Factors of Health Behavior Change

Health literacy and health belief are identified as two important influencing factors on health behavior. According to Health Belief Model (HBM), the concept of health belief refers to the cognition, perception, and evaluation of individuals on health-related issues. The original and core constructs of HBM consisted of perceived susceptibility, perceived severity, perceived benefits, and perceived barriers, which suggested that individuals’ judgment of threat to their health together with their expectation of the targeted actions could predict the likelihood of health behavior change. Several other constructs were also added into HBM later, including cues to action (Rosenstock, 1966), health motivation, and some modifying factors (Becker, 1974), which are rarely developed and analyzed in relevant studies, while the predictive power of self-efficacy (Rosenstock, Strecher, & Becker, 1988) has been proved to be effective.

Some studies have examined the cognitive constructs in Health Belief Model (HBM) to test their linear relationship with health behavior, but it has been argued that the predictive power of some constructs may depend on the value of other constructs (Glanz, Rimer, & Viswanath, 2008). Thus, it needs to take the interaction between constructs into account, such as the perceived severity and susceptibility. Moreover, there is still a lack of discussion of the role of cues to action during this process.

As the number of people having chronic diseases growing rapidly, health education providers pay more attention on the health literacy of the public (Tappe & Galer-Unti, 2001), which refers to the knowledge and capability required for individuals to obtain, understand, evaluate, and apply health information to make reasonable judgments and decisions on health care, disease prevention, and health promotion in daily life (Sørensen et al., 2012). Based on the public’s cognition level in relevant information, health literacy has been partitioned into three dimensions: Functional Health Literacy, Interactive Health Literacy, and Critical Health Literacy (Nutbeam, 2000). After the release of the trial version of Chinese Citizens’ Health Literacy——Basic Knowledge and Skills, a series of investigations and studies have been launched on the implication of health literacy in China (Shen, Hu, Liu, Chang, & Sun, 2015; Zheng, Shi, & Cao, 2010).

Previous research has shown that health literacy is closely related to health behavior and health status (Liu, Liu, Li, & Chen, 2015), but the specific mechanism is still not clear. Besides, although theoretical pathways have been proposed, the relationship between health literacy, health belief, and health behavior has not been investigated extensively, and only few studies have integrated the diverse dimensions of health literacy together.

3 Research Framework and Methods

3.1 Research Framework and Hypotheses

This study uses structural equation model (SEM) to examine the interactions between constructs in HBM and dimensions of health literacy and to explore the relationships between health literacy, health belief, and health behavior in the context of a campus norovirus outbreak. Further, our study will provide recommendations for health education through evidence-based analysis of cues for action. The research framework is proposed as shown in Figure 1.

The dimensions of health literacy considered here are functional, interactive, and critical literacy. We list and use the four original and core constructs in HBM plus self-efficacy to measure health beliefs. The classification and detailed items of health behavior change is adapted from the preventive measures recommended in the Norovirus Outbreak Investigation and Prevention and Control Technical Guideline (2015 Edition) issued by the Chinese Center for Disease Prevention and Control (CCDPC). Besides, control variables including gender, knowledge structure (measured by department), education level (measured by grade), economic status (measured by monthly living expenses), and factors from cues to action (such as information sources and moment to start following the outbreak) are considered.

HBM has always been one of the most widely used conceptual frameworks in health behavior research.
Its effectiveness and operability have been confirmed by empirical studies in many countries (Carpenter, 2010), especially in predicting whether individuals are willing to accept health examinations (Grimley, Kato, & Grunfeld, 2020; Maseko, Huang, & Lin, 2019), and guiding the prevention and management of chronic diseases. However, some studies suggest that the prediction of HBM is a simple linear relationship under the same category, oversimplifying a complex problem (Prentice-Dunn & Rogers, 1986). In addition, some constructs may have insufficient explanatory power, such as the role of perceived severity is questioned in predicting HIV antibody detection (Zak-Place & Stern, 2004). Hence, we hypothesize:

**H1:** In the context of PHE, health belief has a significant impact on health behavior. The higher the perceived severity and susceptibility, the more obvious the benefits, the smaller the barriers, and the higher the self-efficacy, the more likely one is to change his health behavior.

One of HBM’s limitations is that it assumes that everyone has access to equal health information and services on the disease, or at least has the same literacy. Some scholars believe that health knowledge, belief, and behavior are integral parts of health literacy (Xiao et al., 2008). Some found that there is a significant path from health literacy to knowledge to self-efficacy to physical activity, and to health status (Osborn, Paasche-Orlow, Bailey, & Wolf, 2011). Some pointed out that patients with limited health literacy have less health knowledge, and perceives more barriers to take actions (Peterson, Dwyer, Mulvaney, Dietrich, & Rothman, 2007). von Wagner, et al. (2009) developed a framework to draw the possible pathways linking health literacy to health behavior, but there is only limited empirical evidence for the intermediary role of the socio-cognitive variables such as belief and attitude (Fransen, von Wagner, & Essink-Bot, 2012). These different results may be due to the fact that different dimensions of health literacy are rarely taken into consideration. Hence, we hypothesize:

**H2:** Three dimensions of health literacy have a significant impact on health behavior through the mediate effect of constructs in health belief.

**H3:** Three dimensions of health literacy can also directly affect people’s health behavior.

### 3.2 Research Design and Instruments

A questionnaire was used to collect empirical data. The items included were developed primarily on the basis of the scales used in the previous studies. The measurements of health literacy are adapted chiefly from the 14-item health literacy scale (HLS-14) (Suka et al., 2013), which values the three dimensions of health literacy in an operable manner. The health belief was measured utilizing items from the scale developed for breast self-examination.
Factors Influencing the Health Behavior During Public Health Emergency ... (Champion, 1984). Meanwhile, relevant items were added to the scale according to the topics frequently discussed on social media, BBS, and the news website of the university, which provide specific themes related to norovirus. Each item in the scale has a score range of 1–4, representing four levels namely “completely inconsistent”, “relatively inconsistent”, “relatively consistent”, and “completely consistent” respectively.

Structural equation modeling technique was used to test the research framework with AMOS, which is also suitable for models with formative constructs and relatively small samples. Based on the recommended procedure, we assessed the reliability and validity of each latent variable measurement as well as the paths between the constructs and their significance level.

4 Data Collection

Students who had undergone through the norovirus outbreak in the university were recruited to answer the questionnaire and test the research framework. Using convenience and snowball sampling, we distributed online questionnaires in social media, BBS, and WeChat groups in the university via Tencent Questionnaire platform. Further, a pilot survey was conducted and 32 students participated. Feedback and suggestions were received and we modified the questionnaire accordingly. The main survey was launched from March 22, 2020 to April 2, 2020. A total number of 411 responses were received, among which 386 valid responses were included in the analysis after the cleansing by completion time and filtering options. The gender and grade distribution of the sample are relatively uniform, as shown in Table 1.

5 Data Analysis and Findings

5.1 Reliability and Validity

The reliability of the constructs was assessed utilizing Cronbach’s Alpha in order to test the internal consistency and stability. As shown in Table 2, Cronbach’s Alpha values for most constructs in the scale are greater than 0.6, except for Perceived Benefits (0.539) and Critical Health Literacy (0.517), which are also considered acceptable as there are only fewer items measuring these constructs (Terwee et al., 2007). Therefore, the measurements indicated that they are reliable.

Confirmatory factor analysis (CFA) was performed to assess the construct validity of two scales (health literacy and health belief). Model fitness was assessed by the goodness-of fit index (GFI), comparative fit index (CFI), incremental fit index (IFI), and root mean square error of approximation (RMSEA). As shown in Table 3, values of GFI, CFI, and IFI are considered acceptable (>0.8, close to 1) (Hu & Bentler, 1999), values of RMSEA are considered good (<0.1), which indicated an acceptable fit of the three-factor health literacy and five-factor health belief model.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Demographics of the Survey Sample (N=386, person, %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Frequency</td>
</tr>
<tr>
<td>Male</td>
<td>154</td>
</tr>
<tr>
<td>Female</td>
<td>232</td>
</tr>
<tr>
<td>Dept.</td>
<td></td>
</tr>
<tr>
<td>Science</td>
<td>54</td>
</tr>
<tr>
<td>Humanity</td>
<td>61</td>
</tr>
<tr>
<td>Social Science</td>
<td>81</td>
</tr>
<tr>
<td>Economics and Management</td>
<td>48</td>
</tr>
<tr>
<td>Medicine</td>
<td>89</td>
</tr>
<tr>
<td>Interdisciplinary</td>
<td>10</td>
</tr>
</tbody>
</table>

(Champion, 1984).
5.2 Actions Taken Responding to the Outbreak of Norovirus

Behavior changes reported by students fall into four categories, including toilet habits (84%), lifestyle (74%), dietary habits (73%), and hand-washing habits (70%) in the face of the norovirus outbreak (shown in Table 4).

Among the specific actions taken to reduce the chance of infection of norovirus, the topmost four actions selected are washing hand frequently (70%), using toilet separately from students having diarrhea (65%), reducing gathering activities (56%), and ordering takeout or dining out of student canteen (49%). On the contrary, avoiding the dining peak of student canteen (21%) and wearing masks (20%) rank the lowest. Compared with the fact that ordering takeout added to the financial burden of students and avoiding the dining peak was not realistic, hand sanitizers installed in the toilets on campus and alcohol-based hand sanitizer distributed for every student could greatly reduce the hand-washing barriers. It is probably because that the smaller the barriers encountered, the higher the possibility of occurrence of change in the health behavior. Moreover, though it is still unclear about the effectiveness of some actions such as ordering takeout or wearing masks, their adoption rates are not low due to the influence of suspicion and anxiety.

5.3 Hypothesis Testing

5.3.1 Demographic Features

Previous work has showed that demographics, social psychological and social structural variables have an impact on health belief and affect health behavior ultimately (Glanz et al., 2008). Thus, analysis of variance (ANOVA) of demographic features including gender, grade, department, and living expenses was performed to comprehend the change of health behavior. Each option within the same category of action was scored separately (no=0, yes=1), and after standardization of the summed-up score of each category, the degree of health behavior change was measured. The results show that both education level (F=3.869, p<0.01) and the knowledge structure (F=3.908, p<0.01) have significant impacts on the change of health behavior. Specifically, students who are in their sophomore or junior year or majoring in Economics and Management, or Information and Engineering Science score higher than others.

In detail, educational level is found to affect health behavior indirectly through perceived susceptibility (F=3.401, p<0.01), severity (F=5.262, p<0.01), and barriers (F=3.064, p<0.01), which is consistent with previous study (Glanz et al., 2008). Students with different knowledge structure also perceived different severity (F=4.7, p<0.01) and barriers (F=2.481, p<0.05). Moreover, perceived barriers (F=7.998, p<0.01) and self-efficacy (F=7.168, p<0.01) of males are significantly higher than those of females.
5.3.2 Impact of Health Belief on Health Behavior Change

As shown in Table 2, when facing the norovirus, students clearly perceived the benefits (mean score=13.51) of observing personal hygiene and held a high self-efficacy (mean score=11.05). However, many students believed that the chances of getting infected themselves will be less than the average. As to the perceived severity, the main concern of students was their academic performance, which is closely related to their status and the fact that the incident occurred during the final exam week.

Multiple linear regression in Table 5 shows that health belief has a significant impact on change in health behavior. The signs of coefficients in model 1 are all consistent with H1, but only the coefficients of perceived severity and self-efficacy are statistically significant. The t value of the perceived susceptibility in model 2 increased greatly, indicating that there is indeed an interaction between severity and susceptibility, as confirmed in other literatures (Glanz et al., 2008; Yu et al., 2020).

The structural equation model of health behavior and health belief (RMSEA=0.069, CFI=0.832, IFI=0.834) considers the interaction factor. The result also shows that in the context of PHE, perceived severity and self-efficacy are more reliable predictors of health behavior change, with direct effect coefficients of 0.43 and 0.31, respectively. The increase of perceived susceptibility (total effect coefficient = 0.21) can also raise the possibility to take actions by increasing perceived severity. The effect of perceived benefits and barriers on health behavior is not statistically significant.

5.3.3 Impact of Health Literacy on Health Belief and Health Behavior Change

As shown in Table 2, although students in the university have a high level of critical (mean score=13.31) and interactive (mean score=16.36) health literacy, they still cannot make health decisions independently. Their self-rated functional health literacy level is rather low, which may be due to the higher self-expectations.

Multiple linear regression of the constructs in HBM on three dimensions of health literacy was performed to explore the relationship between health literacy and health belief. It is found that the improvement of functional health literacy reduced perceived severity (β=-0.290), susceptibility (β=-0.241), and barriers (β=-0.238); the improvement of interactive health literacy increased perceived severity (β=0.165), benefit (β=0.140), and self-efficacy (β=0.204); and the improvement of critical health literacy reduced perceived severity (β=0.373) and barriers (β=0.249). This indicated that functional and critical health literacy helped people process health information more calmly and rationally, eliminating the impact of emotions such as panic and fear, and thus reducing the perceived threats to health. And interactive health literacy helped people to collect more comprehensive information and communicate with friends, effectively enhancing the perceived benefits and self-efficacy.

We established a structural equation model (RMSEA=0.055, CFI=0.826, IFI=0.828) based on the statistically significant impact parameters mentioned in the last paragraph. And further, Model 3 in Table 5 indicates

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Actions Taken Against Outbreak of Norovirus</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Whether to change the dietary habits (Yes=1)</strong></td>
<td><strong>Mean</strong></td>
</tr>
<tr>
<td>Eating less cold or uncooked food</td>
<td>0.73</td>
</tr>
<tr>
<td>Ordering takeout or dining out</td>
<td>0.42</td>
</tr>
<tr>
<td>Avoiding the dining peak of student canteen</td>
<td>0.49</td>
</tr>
<tr>
<td>Changing other eating habits</td>
<td>0.21</td>
</tr>
<tr>
<td><strong>Whether to change the toilet habits (Yes=1)</strong></td>
<td><strong>Mean</strong></td>
</tr>
<tr>
<td>Using toilet separately from students with diarrhea</td>
<td>0.03</td>
</tr>
<tr>
<td>Reducing the use of toilets in public places</td>
<td>0.39</td>
</tr>
<tr>
<td>Increasing the frequency of flushing</td>
<td>0.21</td>
</tr>
<tr>
<td>Changing other toilet habits</td>
<td>0.01</td>
</tr>
</tbody>
</table>
that health literacy can hardly explain any variance in health behavior independently, thus rejects H3. The SMC of health behavior is 0.26, indicating that health belief and health literacy can explain 26% of the variance in health behavior. It is found that all dimensions of health literacy have an indirect significant effect on health behavior change via health belief, supporting H2. The improvement of functional (total effect coefficient=-0.32) and critical (total effect coefficient=-0.38) health literacy will reduce the change of health behavior, while the improvement of interactive health literacy (total effect coefficient=0.32) will increase it. Therefore, the revised research framework is shown as in Figure 2.

5.4 Students’ Information Sources and Cues to Action

As shown in Table 6, nearly 70% of the respondents have at least three information sources while less than 10% of them only rely on a single source, and the distribution of information sources is listed in Table 6. Personal social networks (73%) and school information platforms such as BBS (56%) have played an important role during the norovirus outbreak. However, the information on these platforms tends to be fragmented and diverse and it is difficult to differentiate the true information from false information (Tian & Duan, 2017), which may cause panic among the student community easily. There are only a few students who only rely on counselors, dormitory directors, and News Website or university TV. These sources are considered to be highly credible but cannot achieve instant interactions, which are used mainly as a supplementary.

The anchoring effect indicates that people’s health belief may be related to the initial information they are exposed to. Many respondents (63.2%) noticed the norovirus outbreak after the discussion on the food safety of student canteen, which could have created an impression that norovirus spread from the canteen, resulting in the higher proportion of ordering takeout or dining out. Other students had the knowledge of this incident after each official respond to the disputes or doubts of students. Moments that drive the students to follow the norovirus outbreak is shown in Table 7. This suggests that debates on social media raised the perceived risk and severity, and ultimately forced people to take actions. In addition, the infection of oneself or surrounding people generated greater motivation to change because it increased the perceived susceptibility.

There have been some studies indicating that emotions such as fear are effective predictors of health behavior (Champion et al., 2004), and are positively correlated with perceived severity, susceptibility, and
barriers (Leung, Wong, & Chan, 2014). Moreover, as the norovirus outbreak occurred during the final exam week, it is not hard to imagine that pressure and anxiety would have a greater impact on students. Thereafter, we examined the impact of emotional factors on health behavior. We selected two emotion-related items in the HBM scale and used their scores to do correlation analysis with constructs in HBM. The result shown in Table 8 suggests that emotional factors are closely related to cognitive factors.

### 6 Conclusion and Implications

#### 6.1 Conclusion

Traditional health education theory emphasizes the unity of knowledge, belief, and behavior, believing that the dissemination and teaching of health knowledge help the targeted population to build reasonable belief on health, and finally promote the change of health behavior. We investigated the influencing factors of college students’
Health behavior change during the norovirus outbreak (an exemplar of PHE) by considering health literacy and clarified the relationship between health literacy, health belief, and health behavior in the context of public health emergency, and analyzed the factors driving them to take actions.

We found that both health belief and health literacy have significant impacts on health behavior change during norovirus outbreak. It is interesting that health literacy does not directly affect health behavior change, but depends on health belief's mediating role. Further, the severity, susceptibility, and barriers perceived by students will be reduced by increasing the functional and critical health literacy, and therefore reduce the blind health behavior change. On the contrary, the enhancement of interactive health literacy effectively promoted health behavior change as it increased the perceived benefits and self-efficacy of students.

Health belief is positively associated with health behavior change, supporting H1. Specifically, the increases of perceived severity of disease and self-efficacy can promote the possibility of health behavior change, but the effectiveness of perceived susceptibility on health behavior changing depends on the increase of perceived severity.

The effect of health literacy depends on health belief’s mediating role to affects health behavior changes. Thus, H2 is supported. Generally, the improvement of functional and critical health literacy will reduce the change of health behavior, while the improvement of interactive health literacy can promote it. Contrarily, health literacy is not directly associated with health behavior change, thus H3 is not supported.

Cues to action have been considered as an important factor for the success of taking healthier actions. Personal social networks and school information platforms played

Table 7
Moments to Start Following the Norovirus Outbreak

<table>
<thead>
<tr>
<th>Key moment</th>
<th>Frequency</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>December 27, some students with diarrhea discussed the food safety issue via BBS. University Hospital posted a notice confirming them as norovirus infection later.</td>
<td>244</td>
<td>63.2</td>
</tr>
<tr>
<td>December 28, students questioned the decision to quarantine norovirus-infected students in another university district on BBS. Later, University Hospital posted a notice stating that sufficient wards had been opened.</td>
<td>70</td>
<td>18.1</td>
</tr>
<tr>
<td>December 29-30, Professor Li accepted an interview. Website, WeChat accounts and counselors began to publicize relevant knowledge systematically.</td>
<td>37</td>
<td>9.6</td>
</tr>
<tr>
<td>January 2, students discussed the suspected closure of a campus building on the Tree Hole. Later, relevant college posted a statement to explain that it was relevant departments cleaning up the floor.</td>
<td>22</td>
<td>5.7</td>
</tr>
<tr>
<td>January 5, Tree Hole began to top a health reminder about norovirus.</td>
<td>9</td>
<td>2.3</td>
</tr>
<tr>
<td>Others _ Infected oneself</td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>Total</td>
<td>383</td>
<td>99.2</td>
</tr>
</tbody>
</table>

Table 8
Correlation Analysis of Emotional Factors and Constructs in HBM

<table>
<thead>
<tr>
<th>Perceived Severity</th>
<th>Perceived Susceptibility</th>
<th>Perceived Benefits</th>
<th>Perceived Barriers</th>
<th>Self-Efficacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.795**</td>
<td>0.482**</td>
<td>0.019</td>
<td>0.284**</td>
<td>0.130*</td>
</tr>
</tbody>
</table>

On thinking of Norovirus, I feel scared or sick

| If I am infected with Norovirus, my perception and emotion of myself will change |
|---------------------------------|---------------------------------|
| 0.747**                         | 0.450**                         |

Notes: * p<0.05, ** p<0.01, two-tailed tests.
an important role during the norovirus outbreak. However, the information on these platforms are fragmented and diverse, and critical health literacy is needed to differentiate the true information from false information.

### 6.2 Implications

This study reported the significant impact of health belief and health literacy on health behavior change and contributes to new insights in understanding health behavior changing in PHE. Therefore, we present the following theoretical contributions.

1. First, when compared to previous research, this study measured health behavior change in the context of a real norovirus outbreak as an exemplar of PHE, and this enriches prior research studies from a historical and practical point of view.

2. Second, this study advances the theoretical progress in understanding health behavior change in PHE from the integrated perspectives of health literacy and health belief. Prior studies focused on studying health belief and literacy separately, but our study integrated them in predicting health behavior change, and explained how health belief mediates the impact of health literacy on health behavior change and helps to predict health behavior change together.

3. Third, this study contributes to the improvement of the Health Belief Model by clarifying the interaction between its constructs such as the perceived severity and susceptibility, rather than treating it as a linear relationship.

4. Finally, this study provides further evidence that information sources selection activity plays a critical role in affecting health behavior change. It must be noted that this study was conducted among Chinese college students utilizing university BBS, which is a university platform with a restricted environment for information seeking with more privacy settings for controlling the flow information. The other important point is that the information on these platforms tends to be fragmented and diverse, and critical health literacy is needed to differentiate the true information from false information.

The findings of this study also provide several implications for stakeholders wishing to understand health behavior change. Social media operators, such as BBS, need to be aware of college students’ preference while selecting the information sources and responding to PHE outbreak. Therefore, operators can understand users’ behavior more accurately. The practitioners, based on the characteristics and properties of college students, could push related content or some attractive topics to increase information interaction among them. Furthermore, social service providers can provide a communication platform that can build independent communication circles by facilitating users in the classifying and differentiation of the shared object and target. Moreover, strategies are needed to enhance the health information education to build the students’ belief by literacy or knowledge acquired in order to encourage the targeted users to take healthier actions during PHE.

### 6.3 Limitation and Future work

There are several limitations in this study. “Cues to action” was considered as an influencing factor in the research, however, it was not considered as a structural variable. Though the key moments relevant to health behavior change were explored in our study, their impact still needs further exploration. Though the ranking of actions taken by college students during norovirus outbreak were explained in our study, the reasons for students choosing certain actions are not clear, which may need further in-depth exploration.

### Acknowledgement

This work is funded by Beijing Social Science Fund, PR China (18XCB007).

### References


Shijuan Li, Qian Jiang, Pengyi Zhang


questionnaires. *Journal of Clinical Epidemiology, 60*(1), 34–42. doi:10.1016/j.jclinepi.2006.03.012


Examining User Perception and Usage of Voice Search

Ning Sa*, Xiaojun (Jenny) Yuan

https://doi.org/10.2478/dim-2020-0046
received August 15, 2020; accepted September 15, 2020.

Abstract: With the development of mobile technologies, voice search is becoming increasingly important in our daily lives. By investigating the general usage of voice search and user perception about voice search systems, this research aims to understand users’ voice search behavior. We are particularly interested in how users perform voice search, their topics of interest, and their preference toward voice search. We elicit users’ opinions by asking them to fill out an online survey. Results indicated that participants liked voice search because it was convenient. However, voice search was used much less frequently than keyboard search. The success rate of voice search was low, and the participants usually gave up voice search or switched to keyboard search. They tended to perform voice search when they were driving or walking. Moreover, the participants mainly used voice search for simple tasks on mobile devices. The main reasons why participants disliked voice search are attributed to the system mistakes and the fact that they were unable to modify the queries.

Keywords: voice search, mobile behavior, user perception

1 Introduction

About a decade ago, voice search was “not a well-known technology” (Crestani & Du, 2006). Nowadays, the rapid development of technologies has made voice search a part of people's daily lives. Major search engine companies (e.g., Google and Microsoft) have embedded voice search features on desktop devices and have launched voice search products on mobile devices. With the development of automatic speech recognition (ASR) technology, the word error rate in speech-to-text transcription has dropped dramatically in spoken language systems and applications (Black et al., 2011). The improvement of ASR has provided users with the option of speaking their queries when searching on the Internet.

Most of the previous studies in the field of voice search either used search engine logs or asked participants to perform predefined tasks in a laboratory experimental environment to examine various research questions. To our best knowledge, few studies have investigated how users perform voice search in their daily lives.

In this study, an online survey was distributed to gather information about the general usage of voice search and to find out user perception of current voice search systems. In particular, we were interested in the following research questions (RQs):

RQ1: How do users perform voice search?
RQ2: How do users perceive voice search?
RQ3: What are the issues that users face when using voice search?

We believe recognition error is an issue to be addressed in developing ASR technology. Partially modifying the query could be an effective way to correct recognition errors. However, with current voice search systems, users are not given the chance to partially modify the current query. Our research assumes that the lack of modification ability is perceived by users as a problem in voice search. These assumptions were included in the survey questions and will be discussed in later sections.

2 Previous Work

Research on users’ voice search behavior has mainly focused on the comparison of textual input and voice input. Existing studies are based on either search engine
log data or controlled experiments (Crestani & Du, 2006; Schalkwyk et al., 2010; Kamvar & Baluja, 2007; Begany, Sa, & Yuan, 2016). Studies on search engine log analysis reflect real-life usage conditions but have been restricted by their ability to discover the reasons behind the findings. In controlled experiments, researchers are able to communicate with the participants. But the study results of such experiments cannot be easily generalized to a different experimental setting because of the limited size of the data set and the small pools of participants.

Using Google log data collected from BlackBerry phone users, Schalkwyk et al. (2010) investigated the topic distribution of spoken queries and compared these with written queries. Spoken queries were found to be more likely about local service and less likely about sensitive subjects or interaction-intensive tasks. Spoken queries were also reported to be shorter and more likely to ask “Wh” or “how” questions (Schalkwyk et al., 2010). Based on the same data set used by Schalkwyk et al. (2010), Kamvar and Baluja (2007) researched on the cases when the users would like to speak a query instead of typing one. They obtained similar results as Schalkwyk et al. (2010) on topic distribution and query length. They also found that users having a compressed keyboard were more likely to issue spoken queries than users having a full standard keyboard. In addition, users also tended to put forth a query by speaking when they could get quick answers without having to get further detailed information.

A survey (Google, 2014) about Google voice search indicated that 55% of the teens between the ages of 13 and 18 years used voice searches on mobile phones every day. Direction-related search was found to be one of the most frequently performed voice search among teens and adults. The study also included questions about the context of voice search. It reported that voice search was often used when users were doing activities including cooking or watching television. More teens than adults performed voice search when they were accompanied by someone else. As to why they searched with voice, most of the teens responded that it was “just for fun”.

In a laboratory experiment, Crestani and Du (2006) compared written queries and voice queries collected from desktop devices in terms of query length, time, part of speech, and retrieval effectiveness measured by recall and precision. Spoken queries were found to be longer than written queries and contained more stop words. No significant difference was found regarding the time taken to generate the voice query and the written query. The search effectiveness was not found to be significantly different, either. However, because the participants were allowed to generate only one query for each task, no reformulation process was involved. In an experimental study, voice search was reported to generate longer queries and involve less query reformulation (Yuan, Beklin, & Sa, 2013). In their follow-up work, by analyzing the interview data of the experiment, Begany et al. (2016) identified the factors affecting user perception of voice input and textual input (keyboard and mouse). It was found that participants liked the textual input mainly because of their familiarity with the keyboard/mouse and because of the faster/easier query input with keyboard/mouse. However, the participants thought that the spoken interface is fun and novel.

When comparing the length of voice queries to the length of textual queries, results were not consistent (Crestani & Du, 2006; Schalkwyk et al., 2010; Kamvar & Baluja, 2007; Yuan et al., 2013). Crestani and Du (2006) and Yuan et al. (2013) found voice queries to be longer, while Schalkwyk et al. (2010) and Kamvar and Baluja (2010) concluded that textual queries were longer. When explaining why the voice queries were shorter, Schalkwyk et al. (2010) mentioned that “users may avoid longer queries because they are harder to ‘buffer’ prior to speaking”. To generate longer voice queries, participants would request longer time to formulate the queries before speaking it out. In real-life searches, users may need to get the answers quickly, so they tend to avoid long queries, while in an experimental environment, the users were allowed enough time to formulate the queries. The interview data of the two experimental studies seem to support this argument. Crestani and Du (2006) reported that “the cognitive load on a participant to speak out their thoughts was also high. Some of them commented that they had to formulate their queries in their heads before speaking aloud with no mistakes.” One participant in a study (Begany et al., 2016) also said that “I thought a lot more about how to frame what I was going to ask when I spoke (to it).”

Though the average voice query was longer in the two experimental studies (Crestani & Du, 2006; Yuan, Belkin, & Sa, 2013), some participants issued short voice queries. It may be that some users felt comfortable with textual input and awkward when speaking to the machine, so they just “spoke the written queries” (Crestani & Du, 2006). Alternately, some users reported that they did not believe “the system could understand everything they said”, so they kept the spoken query short (Begany et al., 2016). It can be noticed that these reasons were not reported in the log-based studies.

The search engine log-based studies reported that the voice queries contained more “Wh” (e.g., “what”, “which”, “why”, “when”, and “where”) and “how” questions. The
two experimental studies found more stop words in voice queries than in written queries. Regarding why the users preferred voice queries over textual queries, the survey study (Google, 2014) and the study of Begany et al. (2016) agreed on the “fun” and “novel” factors. The log-based studies reported about the “keyboard constriction” factor, which was not reported in the experimental studies. However, some participants in the study of Begany et al. (2016) mentioned that voice queries were able “to free up your hands”, which is similar to the “click free” factor in the study of Schalkwyk et al. (2010) and the “multitasking” and “efficient” features in the Google survey (2014).

To conclude, most of the user voice search behavior studies were based either on search engine log data or on experimental studies. Google (2014) conducted a survey study on Google voice search. However, the survey population and the complete survey results were not available. In this study, a survey was used to elicit information about the general usage of voice search and to find out user perceptions about current voice search systems. The results reveal the problems of the existing voice search systems and point out possible ways of improvement. The task types and topics collected could be used in the design of future voice search experiments.

3 Methodology

The online survey is composed of five parts: demographic information, search background and search experience, electronic device usage, voice system usage, and voice search. Each of the first four parts contains 3–5 questions, which collected general information from the users. The perceived ease of use and perceived usability of the current voice search system were obtained from the following questions in the fifth part:

• Under what circumstances would people like to perform a voice search over a textual search? On what devices? In what kind of environment? Alone or with other people?
• How often do people perform a successful voice search?
• What do people like or dislike about voice search? What features do they want to add to the current voice search system?

Finally, people were asked to describe the most recent/impressive voice search they performed. They could tell about what they had searched for and whether they had succeeded or failed. This could, therefore, provide examples of voice searches as well as real-life voice search topics. Most of the questions were developed based on the findings of our previous studies (Yuan et al., 2013; Begany et al., 2016) and the relevant literature (Kamvar & Baluja, 2010; Google, 2014).

Google forms were used to design the questionnaire and collect the responses. The complete survey can be viewed online at: http://goo.gl/forms/6ZVR78wp8a. The link was distributed to the listservs of the students and faculty/staff of a university in the United States and to people outside the university by using public listservs.

4 Results

Sixty-four complete responses were collected. The descriptive statistical results of the survey are presented below.

4.1 General Information

Among the 64 participants, 38 (59.38%) were female, 28 (43.75%) were between 20 and 29 years of age, 46 (71.88%) were Caucasian, and 28 (43.75%) held a master’s degree. The details of the demographic information are shown in Table 1 below. Accordingly, 46 (71.88%) participants gave their job title, which included professor, engineer, graduate assistant, and so on; 34 (53.13%) participants indicated that they were students and the majors covered Library and Information Science, Public Health, History, Physics, and so on.

In the search experience section, the participants answered questions about how often they performed a Web search, their favorite search engine, and whether or not they used query suggestions. All the 64 participants indicated that they visited search engines/Websites every day. Google was used most often by 58 (90.63%) participants. The other search engines used were Bing, Baidu, Google Scholar, and Portal Capes.

Regarding the electronic devices that the participants owned and how often they used the devices to do a Web search, out of the 64 participants, 61 (95.31%) had laptops, 57 (89.06%) had smartphones, 43 (67.19%) had desktops, and 36 (56.25%) had tablets. In addition, 22 (34.38%) participants indicated that they had all the four kinds of devices listed above. The other devices owned include iPod, workstation, dumb cell with text, and smart television. Compared with 63 (98.43%) participants searching on desktop/laptop every day, searching on a
smartphone or a tablet was performed every day by 46 (71.88%) participants and two or three times a week by 13 (20.31%) participants. Three (4.69%) participants had never searched on a tablet or a smartphone. However, none of them owned a smartphone.

When asked about the usage of voice assistant systems (such as Siri or Google Voice), 41 (64.06%) participants reported having used such a system. Thus, 31 (75.61%) of them had used Siri, 22 (53.66%) had used Google Voice, and 3 (7.32%) had used a Microsoft Voice product. Figure 1 shows that participants used voice systems on smartphones the most, on tablets to a lesser extent, and on Android TV the least. Internet search was performed most by participants (87.8%), followed by generating text messages (43.9%) and opening an application (39.2%).

### 4.2 Voice Search

Twenty-five (39.06%) of the 64 participants had never used voice search. When asked why they had not performed a voice search, some participants said that they typed their search only because “I’m very comfortable typing in my searches” and “I prefer to type-it is faster and easier.” Some had language concerns that the system could not recognize their queries because of their accents. One participant wrote “I also believe that it won’t ‘get’ what I am asking for, since I am not a native English speaker and perhaps won’t understand my accent.” Some did not want to disturb the people around them. Quite a few participants mentioned that there was no need to do a voice search, as one of them said “I never felt the need to.”

The participants who performed a voice search were then asked how often they performed a voice search and how often they had succeeded in the search. Table 2 presents the answer distributions accordingly. It can be observed that although 39 participants said that they had performed a voice search, only three (7.69%) of them did a voice search every day and almost half of them seldom searched with voice. As mentioned above, all 64 participants performed Web search every day, and so it seems to us that voice search is much less popular than keyboard search. The success rate of voice search was found to be low. According to Table 2, only 1 (2.56%) of the 39 participants succeeded in every voice search; 17 (43.59%) said most of the voice searches were successful and 18 (46.15%) indicated that only some of the searches were successful. In later sections, the difficulties that users encountered during voice search will be explored.

The next several questions were about how voice searches were performed. Thus, 32 (82.05%) out of 39 participants said that they had used smartphones to do a voice search. 11 (28.21%) participants had used tablets, 5 (12.52%) had used laptops, and only 2 (5.13%) had conducted voice search on desktops. Similar to a keyboard search, Google (34 out of 39: 87.18%) was also the most-used search engine in voice searches. Table 3 displays where and when the participants had performed a voice search. We found that the participants performed a voice search when they were staying in a room (22 out
of 39; 56.41%), as well as when they were moving (19 [48.72%] when driving and 15 [38.46%] when walking). One participant said that he/she had done a voice search when cooking; 26 (66.67%) participants had spoken a search when they were by themselves and 11 (28.21%) had searched when there were acquaintances around. Only one (2.56%) participant said that the voice search was conducted when there were strangers around.

Voice search topics are displayed in Figure 2. The topics are selected from the top-level categories on DMOZ (http://www.dmoztools.net/). Reference was found to be the most popular topic. Leisure, news, recreation, and shopping were also frequently searched. The more serious topics, such as society, health, regional, news, and business, were less likely to be spoken out aloud.

In the section of user perception about voice search, the reasons why the participants performed a voice search and what they liked and disliked about voice search were explored. The results are displayed in Table 4. The top three popular reasons why the participants performed a voice search are “your hands are occupied by something else and it is not convenient to type” (24 participants: 61.54%), “for fun” (14 participants; 35.90%), and “you don’t want to type” (13 participants: 33.33%). Unavailable keyboard and a difficult-to-use keyboard are not very popular reasons and were selected by 7 (17.95%) and 4 (10.26%) participants accordingly.

Over half of the participants who had performed voice search (22 participants: 56.41%), found voice search convenient, 11 (28.21%) found it fun, and eight (20.51%) found it fast; four (10.26%) participants indicated that they actually did not like voice search. The majority of the participants (87.18%) complained about the system mistakes. More than 40% participants found “cannot modify the query” a drawback of voice search.

Table 2
The Frequency of Voice Search, Successful Voice Search, and Those who Gave Up

<table>
<thead>
<tr>
<th>Options</th>
<th>No. of answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>How often do you perform a voice search?</td>
<td></td>
</tr>
<tr>
<td>Every day</td>
<td>3 (7.69%)</td>
</tr>
<tr>
<td>Two or three times a week</td>
<td>10 (25.64%)</td>
</tr>
<tr>
<td>Two or three times a month</td>
<td>7 (17.95%)</td>
</tr>
<tr>
<td>Seldom</td>
<td>19 (48.72%)</td>
</tr>
<tr>
<td>How often have you been successful in a voice search?</td>
<td></td>
</tr>
<tr>
<td>Every time</td>
<td>1 (2.56%)</td>
</tr>
<tr>
<td>For most of the voice searches</td>
<td>17 (43.59%)</td>
</tr>
<tr>
<td>For some of the voice searches</td>
<td>18 (46.15%)</td>
</tr>
<tr>
<td>Never</td>
<td>3 (7.69%)</td>
</tr>
<tr>
<td>How often do you give up voice search or switch to typing?</td>
<td></td>
</tr>
<tr>
<td>Every time</td>
<td>2 (5.13%)</td>
</tr>
<tr>
<td>For most of the voice searches</td>
<td>14 (35.90%)</td>
</tr>
<tr>
<td>For some of the voice searches</td>
<td>22 (56.41%)</td>
</tr>
<tr>
<td>Never</td>
<td>1 (2.56%)</td>
</tr>
</tbody>
</table>

Table 3
Where and When Voice Search Was Performed

<table>
<thead>
<tr>
<th>Options</th>
<th>No. of answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Where was a voice search performed?</td>
<td></td>
</tr>
<tr>
<td>Staying in a room</td>
<td>22 (56.41%)</td>
</tr>
<tr>
<td>Driving</td>
<td>19 (48.72%)</td>
</tr>
<tr>
<td>Walking</td>
<td>15 (38.46%)</td>
</tr>
<tr>
<td>Traveling using public transportations</td>
<td>1 (2.56%)</td>
</tr>
<tr>
<td>Other</td>
<td>4 (10.26%)</td>
</tr>
<tr>
<td>When was a voice search performed?</td>
<td></td>
</tr>
<tr>
<td>When you are alone</td>
<td>26 (66.67%)</td>
</tr>
<tr>
<td>When there is acquaintance around</td>
<td>11 (28.21%)</td>
</tr>
<tr>
<td>When there are strangers around</td>
<td>1 (2.56%)</td>
</tr>
<tr>
<td>You did not pay attention to who are around</td>
<td>8 (20.51%)</td>
</tr>
</tbody>
</table>
Pronunciation was a concern of 12 (30.77%) participants, and nine (23.08%) participants felt voice search was slow. Finally, the participants were asked to provide two real voice-search examples, one successful and the other failed. The successful searches could be divided into two major types: getting addresses and getting facts. The participants had successfully received the address of a restaurant, an auto zone, or a business. Factual information included weather, recipes, movie time, football game time, a song, business hours, and so on. Some examples are listed below.

- “These are usually facts – like who was the star of a movie or what year did something happen.”
- “Looking up recipes while cooking.”
- “I asked Siri to find a phone number to a specific business and then dialed the number so I was connected.”
- “What is the address of a restaurant (but a restaurant that isn’t part of a chain)?”

When giving examples of unsuccessful voice searches, some participants provided general answers, such as “when the query is longer” and “When talking to the phone when others are around”. Some had given specific examples, as listed below. “Searching for a specific technical term like spoilation (should be spoliation here) is difficult”

“Searches for specific persons in history almost always lead to responses that relate to words sounding just like or similar to the name (e.g., names that are also occupations such as Farmer, Taylor, Archer, Baker; names that sound like or similar to other nouns or verbs such as Hare, Bacon, Call, Hardin, Curry, Sprain). I now use the keyboard for such searches.”

“Searched for historical information about John Muir”

“What day is national pug day?”

“I was also trying to do it as a phrase search, which meant using the keyboard to put the quotation marks in. I was driving at the time and eventually I just parked and used the keyboard.”

“I wanted to know about the Notre Dame football game score. It searched for something like Noder Dame and then it gave me the history of football there instead of the recent info I wanted.”

### Table 4

**Users’ Perception about the Current Voice Search Systems**

<table>
<thead>
<tr>
<th>Options</th>
<th>No. of answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Why voice search?</td>
<td></td>
</tr>
<tr>
<td>Your hands are occupied by something else and it is not convenient to type</td>
<td>24 (61.54%)</td>
</tr>
<tr>
<td>For fun</td>
<td>14 (35.90%)</td>
</tr>
<tr>
<td>You don’t want to type</td>
<td>13 (33.33%)</td>
</tr>
<tr>
<td>The keyboard is not available</td>
<td>7 (17.95%)</td>
</tr>
<tr>
<td>The keyboard is not easy to use</td>
<td>4 (10.26%)</td>
</tr>
<tr>
<td>Like voice search</td>
<td></td>
</tr>
<tr>
<td>It is convenient</td>
<td>22 (56.41%)</td>
</tr>
<tr>
<td>It is fun</td>
<td>11 (28.21%)</td>
</tr>
<tr>
<td>It is fast</td>
<td>8 (20.51%)</td>
</tr>
<tr>
<td>I prefer talking to typing</td>
<td>3 (7.69%)</td>
</tr>
<tr>
<td>Other (I don’t like it)</td>
<td>4 (10.26%)</td>
</tr>
<tr>
<td>Dislike voice search</td>
<td></td>
</tr>
<tr>
<td>The system mistakes</td>
<td>34 (87.18%)</td>
</tr>
<tr>
<td>Cannot modify the query</td>
<td>16 (41.03%)</td>
</tr>
<tr>
<td>Not sure about some of the pronunciations</td>
<td>12 (30.77%)</td>
</tr>
<tr>
<td>It is slow</td>
<td>9 (23.08%)</td>
</tr>
<tr>
<td>Cannot copy/paste</td>
<td>6 (15.38%)</td>
</tr>
<tr>
<td>Other</td>
<td>2 (5.13%)</td>
</tr>
</tbody>
</table>

5 Discussion

In this study, the general usage of voice search and user perception about voice search systems was investigated. It is worth noting that among the 64 participants, 25 had never performed a voice search. Out of the 39 participants who had carried out voice search before, only three (7.69%) claimed that they used voice search every day. Given that all 64 participants used a search engine every day, it seems to us that voice search is much less popular than keyboard search. In the Google survey (2014), 55% of US teens (13–18 years of age) were reported to use voice search every day. Since all the participants in our study were at least 20 years old, and Google (2014) did not give the statistics of how often adults used voice search, we feel that the results of Google (2014) may not be able to be generalized to our study.

We found that there were several main reasons why voice search was not as widely used as keyboard search. Many participants who had never performed voice search found there was no need to speak out about their search. As one put it, “I never felt the need to”. It seems that they would always perform a search when typing was available. One participant said, “I usually think about searching as typing and I usually search for information when I can use my hands”. When the option of voice search was given, they still typed their queries because they got used to...
typing. Typing was faster and easier, and typing was so naturally connected with search that some "forgot that voice searching was an option". In addition, the current voice search systems have some drawbacks, which do not make the systems as user-friendly as keyboard systems. The participants disliked voice search because the system made errors, the query could not be edited, they were uncertain about some pronunciations, and so on. As a result, the majority of the participants (36 out of 39; 92.31%) gave up voice search or switched to typing for some or most of the time. Two participants reported that they gave up/switched to typing for every voice search. According to Shokouhi, Jones, Ozertem, Raghunathan, and Diaz (2014), in all the mobile query reformulation samples, 13.3% was voice query followed by voice query, while 2.9% was voice query followed by text query. Providing multimodal functionalities seems a good solution to solve the problems of voice search systems. However, based on an early study of Halverson Horn, Karat, and Karat (1999), users will not switch input methods until they achieve certain “spiral depth”, which refers to “the number of times a subject continues to re-dictate the same word, despite incorrect recognition” (Oviatt & VanGent, 1996). It seemed that when the users finally switch to typing from speaking, they simply give up voice input.

The top reason why the participants liked voice search is that “it is convenient”. In the section about how voice search was performed, 22 (56.41%) of the participants with voice search experiences recalled that they had performed a voice search when staying in a room, while 24 (61.54%) had performed voice search when they were driving or walking. When asked why they used voice search, 24 (61.54%) participants selected, “your hands are occupied by something else and it is not convenient to type”. It is not difficult to understand why the participants found voice search convenient when they were moving or when they could not type. Correspondingly, voice search was found more frequently on mobile devices, including smartphones (32 participants: 82.05%) and tablets (11 participants: 28.21%), than on laptops (5 participants: 12.82%) or desktops (2 participants: 5.13%). Google survey (2014) also reported that “76% of all Americans think voice search is great for multitasking”. When the users’ hands are occupied, the multimodal approach might not be a proper solution. As one participant said in his example search, he could not do a phrase search by speaking and finally he had to stop driving and use the keyboard.

Voice search topics were also investigated in the survey, and real-life examples were collected. The popular voice search topics were found to be reference (maps, education, libraries¼), leisure (movies, televisions, music¼), news (media, newspaper, weather¼), recreation (travel, dining out, outdoors¼), and shopping (clothing, gifts¼). Our findings were similar to those of Schalkwyk et al. (2010) that the top voice search topics were reported to be business of a specific place, food and drink, shopping, and traveling. Later in the survey, many of the participants gave voice search examples of finding addresses and the hours of local businesses or restaurants, checking movie times or football game times, looking for the latest news, and so on. All these are factual information retrieval (IR) tasks. According to Kim (2008), factual tasks collect facts and are close ended; interpretive tasks and exploratory tasks are open ended and more complicated. The latter two types of tasks have been frequently used in previous IR experiments. However, it may be the case that users mostly use voice search for factual information.

6 Conclusion and Future Work

This study explored users’ voice search behavior in general. An online survey was carried out to investigate the usage and user perception of voice search systems.

The results indicated that users liked to use voice search, but the ASR technology needs improvement. Voice search systems can better facilitate users by enhancing the voice query process with the ability of partial modification. With traditional IR systems, users can easily modify a query by using a keyboard and a mouse. However, in voice search, it is hard for the users to perform modifications. A good speech system should be able to allow users to make any type of modifications on any part of the query without any intervention. It would be interesting to further investigate how to establish such a system and then compare the effectiveness of the system with the traditional voice search system.

Acknowledgments: We thank all the volunteers of the study.

References

American Society for Information Science and Technology, 57(7), 881–890. doi: 10.1002/asi.20350


**Abstract:** Contextual factors have been found to be an important factor in information searching behaviors, however, little attention has been paid to the influence of contextual factors on users’ query typing behaviors. This study aims to explore the influence of two different contextual factors (with or without time constraint and two kinds of task type) on users’ query typing behaviors. We recruited 40 college students and conducted a user experiment, in which each participant completed two types of search tasks (Fact Finding and Information Understanding) in two different time conditions. The results show that time constraint encourages users to increase their query typing speed. Furthermore, the task type affects query length and rate of keystroke errors.

**Keywords:** query typing behaviors, time constraint, task type

**1 Introduction**

Research in interactive information retrieval has found that contextual factors could influence users’ search behaviors. Task-related features are often examined in empirical studies, e.g., task difficulty, task type, or task complexity (Byström & Järvelin, 1995; Gwizdka & Spence, 2006; Kim, 2014). Time constraint has also been explored recently in multiple studies (Crescenzi, Kelly, & Azzopardi, 2015; Liu, Liu, Gedeon, Zhao, Wei & Yang, 2019; Liu & Wei, 2016). However, these studies often examined users’ page browsing or selection behaviors, either on content pages, or on search result pages. Not many studies have examined users’ query typing behaviors and how such query typing behaviors would be influenced by contextual factors (Ong, Järvelin, Sanderson, & Scholer, 2018). Current technology allows users to issue queries to search systems via multiple methods, such as importing images, oral speaking, or traditional query typing in the search box. In this study, we focus on the traditional query typing behaviors in search box and examine how task type and time constraint would influence searchers’ query typing behaviors.

Examining query typing behaviors helps us to better understand how users might react in different situations and what kind of system support they need during the search process. This study was therefore designed to examine the effects of time constraint and task type on users’ query typing behaviors. Specific research questions are as follows:

**RQ1: What is the impact of time constraint on query typing behaviors?**
1) How would time constraint influence users’ query typing behaviors?
2) How would time constraint influence users’ query typing behaviors under different task types?

**RQ2: What is the impact of task type on query typing behaviors?**
1) How would task type influence users’ query typing behaviors?
2) How would task type influence users’ query typing behaviors under different time conditions?
2 Related Work

Research has found that perceived time pressure and task difficulty are negatively related to searchers’ satisfaction with their search strategies (Crescenzi, Capra, & Arguello, 2013). Previous analysis has showed that searchers often employed different search strategies at different search stages, depending on their time conditions. When there was no time constraint, users tended to employ economic-style search strategy at the beginning of search, but when time constraint was imposed, users became more selective and cautious in examining the search results (Liu & Wei, 2016). Time constraint also significantly influenced searchers’ pre-search confidence, evaluation of search performance, knowledge acquisition, and aftersearch affective states, and time constraint have different influence in two types of tasks (Liu, Zhang, Jiang, Yang, & Zhao, 2015). It is also found that time constraint and usefulness had interaction effects on the first dwell time, while usefulness had positive relationship with total dwell time. Knowing time constraint helps predict document usefulness from dwell time (Liu, Liu, & Yan, 2018). In the current study, we mainly focused on the effects of task type and time constraint on users’ query typing behaviors.

Studies that had examined users’ query typing behaviors mainly focused on the effect of search interface or certain functions. For example, Kamvar and Baluja (2008) asked participants to conduct search using nine-key-buttons interface on mobile devices, and providing two types of search interface, with or without query suggestions. Their results demonstrated that users who were using the interface with query suggestions could use significantly fewer keystrokes, so that to reduce searchers’ working load and to increase their search engagement. Even though using this interface would lead to an increase in their total number of keystrokes, users still prefer such interface. Ong, Járvelin, Sanderson and Scholer (2018) examined users’ typing speed and considered it as part of search efforts into Search Economic Theory (SET) model. They found searchers with faster typing speed would issue much shorter queries, but they often issued more queries, and they tended to use shorter time to evaluate search results. Google launched Google Instant Search in 2010, which was aimed to help users see the search results faster. With Google Instant, as searchers begin to type queries into the Google search box, they would see the search results of the suggested queries before finishing typing. However, Google decided to remove this feature in 2017 (Whitwam, 2017). This also reminds us that users’ query typing behaviors are important and we need more investigation to understand how users are typing queries and how various contextual factors could influence such behaviors. Better understanding of query typing behaviors could help us design more appropriate and effective search tools to help searchers find what they need on the Internet.

3 Method

To investigate the effects of time constraint and task type on user’s query typing behaviors, a user study was carried out.

3.1 Participants

Forty undergraduate students (20 males and 20 females) aged from 18 to 23 with a mean age of 20.25 were recruited via Bulletin Board System (BBS) of Peking University. They were evenly distributed in all grades: 5 freshmen, 11 sophomores, 15 juniors, and 9 seniors. A QWERTY keyboard was provided for the experiment, and each participant was asked to complete four assigned search tasks using Windows 10 system on a desktop computer provided by our lab. After they completed the experiment, they would be compensated with an 8 GB USB flash drive for participation.

3.2 Procedure

Data were collected as part of a larger user experiment where individual searching behavior were collected (Liu, Liu, Gedeon, Zhao, Wei & Yang, 2019). First, participants were requested to sign a consent form and filled out a background questionnaire subsequently. Next, they were provided a brief introduction to the experiment and were then given a training task to get familiar with the search systems and our experimental settings. Finally participants were asked to conduct searches for four tasks. After they read the task description and before they conducted search for each task, a pre-search questionnaire about their expected task difficulty and topic familiarity was given; during searching, participants were asked to respond to each task by typing or copying/pasting useful information for their tasks into a notebook file; after completing the task (or time is up in time constraint conditions), they need to finish a post-search questionnaire to evaluate their own task performance. The interactions between searchers and the computer were recorded using Morae Recorder 3.0.
3.3 Time Constraint

There were two types of time condition in our study: with time constraint (TC) and without time constraint (NTC). In TC conditions, searchers were required to complete a single task in 5 minutes. Based on prior studies (Topi, Valacich, & Hoffer, 2005; Weenig & Maarleveld, 2002), pilot tests determine the limit. To force a severe time constraint, the time constraint was set to 50% of the average time spent by participants in our pilot studies without time constraint. While in NTC conditions, there was no time limit, and searchers could stop searching once they thought enough information has been collected to complete the task.

To investigate the two independent variables (time constraint and task type), we conducted a 2×2 factorial within-subject design. Each participant was required to complete two types of search task under TC and NTC conditions. Meanwhile, a 2×2 Graeco–Latin Square design was used to systematically balance the order of time conditions and search tasks (Kirk, 2013).

3.4 Search Tasks

According to the first two cognitive process dimensions (remember and understand) of Taxonomy of Learning (Anderson & Bloom, 2001), we designed two types of search tasks: Fact Finding (remember: retrieving, recognizing, and recalling relevant knowledge from long-term memory) and Information Understanding (understand: constructing meaning from oral, written, and graphic messages through interpreting, exemplifying, classifying, summarizing, inferring, comparing, and explaining). There were two tasks with different topics for each type of search task, and each task was constructed in a simulated work task situation (Pia, 2003). The specific description of these four tasks is shown as below:

**Fact Finding 1 (FF1):** You heard that India has very interesting wedding traditions, and now you want to search for the following aspects of Indian Weddings: Wedding dresses, painted hands, and the type of food served.

**Fact Finding 2 (FF2):** One of your friends said he was bitten by a rove beetle and felt very itchy, and the wound festered after scratching. You were quite worried about this type of beetles. You want to search what is a rove beetle? Is it poisonous? What should you do if you see a rove beetle? If bitten by a rove beetle, how should we treat it?

**Information Understanding 1 (IU1):** Your nephew is considering trying out for a football team. Most of your relatives are supportive of the idea, but you think this sport is dangerous and are worried about the potential health risks. Specifically, what are some long-term health risks faced by teen football players?

**Information Understanding 2 (IU2):** Doric column is a distinctive architectural form in ancient Greek architecture. Please search for information about the general characteristics and representative examples of Doric columns, and whether Doric columns have had any influence on Chinese architecture. If so, what are some representative examples?

3.5 Dependent Variables

Users’ query typing behaviors are the main research objects in our study, and we considered the following four variables to measure users’ query typing behaviors:

**Typing Speed:** We first considered users’ typing speed, which is calculated by the length of query in bytes, divided by the interval time between the first and last keystrokes, as shown in the following formula. In Chinese, each character occupies two bytes.

\[
\text{Typing speed} = \frac{\text{length of query in bytes}}{\text{interval between the first and last keystrokes}}
\]

**Number of Keystroke Errors (Num_keyerrors):** When typing Chinese characters, users usually need to type letter keys and then use number keys, Enter and/or Space keys to select the right characters. Therefore, in this study, we counted the number of keystrokes that are not letter keys, number keys, nor Enter/Space keys that appeared during query typing process, e.g., Backspace or Delete (DEL) keys, and recorded it as the number of keystroke errors (Num_keyerrors).

**Rate of Keystroke Errors (Rate_keyerrors):** This measure is calculated as the number of keystroke errors divided by the total number of keystrokes in this query. The rate of keystroke errors can reflect the adjustment frequency and incorrect keystroke frequency of users in query typing.

\[
\text{Rate of keystroke errors} = \frac{\text{Num_keyerrors}}{\text{the total number of keystrokes in this query}}
\]

**Query Length:** We used the Rwordseg package in R to segment Chinese words, and simply segmented English word by spaces. The number of segmented query words in each query is the query length.
Examination of Effects of Time Constraint and Task Type on Users’ Query Typing Behaviors

4 Results

4.1 The Overall Characteristics of Users’ Query Typing Behaviors

A total of 160 search sessions were collected in this experiment. Data of nine sessions were lost due to uncontrollable reasons. Thus, this study analyzed the data of the remaining 151 search sessions, among which 890 queries were issued by participants. Table 1 shows the basic characteristics of the four variables.

4.2 The Effects of Time Constraint on Query Typing Behaviors

4.2.1 The Overall Impact of Time Constraint on Query Typing Behaviors

The Mann-Whitney $U$ test results (Table 2) reveal significant effects of time constraint on query typing speed ($p = 0.011$) and query length ($p = 0.001$), indicating users had significantly faster query typing speed in TC than in NTC conditions, and they tended to construct longer query in TC than in NTC conditions. The average number of keystroke errors and rate of keystroke errors in TC conditions were a bit higher than those in NTC conditions, however, their differences were not significant.

4.2.2 Effects of Time Constraint in FF Type of Tasks

In our study, 365 queries were issued in Fact Finding tasks. As shown in Table 3, query length in TC conditions was much longer than that in NTC groups ($p = 0.006$) in Fact Finding tasks. Even though the typing speed and number of keystroke errors in TC conditions were slightly higher than those in NTC conditions, the difference was not significant.

4.2.3 Effects of Time Constraint in IU Type of Tasks

Users issued 525 queries in Information Understanding tasks in total. Table 4 shows a significant effect of time constraint on typing speed ($p = 0.008$) and query length ($p = 0.014$) in Information Understanding tasks. Consistent with the results in Section 4.2.1, searchers with time constraint conducted significant more bytes per minute and longer queries than without time constraint.

4.3 The Effects of Task Type on Query Typing Behaviors

4.3.1 The Overall Impact of Task Type on Query Typing Behaviors

In this study, 365 queries were issued in Fact Finding tasks, and 525 in Information Understanding tasks. In general, searchers tended to construct significantly
shorter query and have a higher number of key errors in Fact Finding tasks, as shown in Table 5. It also showed that typing speed and number of keystroke errors in Information Understanding tasks were higher than those in Fact Finding tasks, but the difference between the two types of tasks was not statistically significant.

4.3.2 Effects of Task Type in TC Conditions

It’s clear from Table 6 that searchers constructed shorter query in Fact Finding tasks than in Information Understanding tasks ($p < 0.001$), when searching with time constraint. Though no significant effect was found in the other three variables, the results indicate the searchers would have higher typing speed and lower keystroke error and error rate in Information Understanding tasks.

4.3.3 Effects of Task Type in NTC Conditions

There were 512 queries issued without time constraint. The results of Mann-Whitney $U$ tests reveal significant effects of task type on query length and rate of keystroke errors in NTC conditions (Table 7). When searching without time constraint, searchers with Fact Finding tasks had much higher error rate ($p = 0.015$) and constructed shorter query ($p = 0.000$) than in Information Understanding tasks. The detail for each of the variables were presented in Table 7.

---

**Table 3**

Nonparametric Test Results of Query Typing Behaviors in FF Type of Tasks under Different Time Conditions

<table>
<thead>
<tr>
<th></th>
<th>Mean (SD)</th>
<th>Mann–Whitney $U$ test $Z (p)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TC ($N=165$)</td>
<td>NTC ($N=200$)</td>
</tr>
<tr>
<td>Typing Speed (in bytes)</td>
<td>356.99 (532.70)</td>
<td>267.72 (379.77)</td>
</tr>
<tr>
<td>Num_keyerrors</td>
<td>3.81 (5.64)</td>
<td>3.29 (3.66)</td>
</tr>
<tr>
<td>Rate_keyerrors</td>
<td>0.20 (0.19)</td>
<td>0.21 (0.20)</td>
</tr>
<tr>
<td>Query Length</td>
<td>3.16 (1.59)</td>
<td>2.81 (1.46)</td>
</tr>
</tbody>
</table>

**Table 4**

Nonparametric Test Results of Query Typing Behaviors in IU Type of Tasks under Different Time Conditions

<table>
<thead>
<tr>
<th></th>
<th>Mean (SD)</th>
<th>Mann–Whitney $U$ test $Z (p)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TC ($N=213$)</td>
<td>NTC ($N=312$)</td>
</tr>
<tr>
<td>Typing Speed (in bytes)</td>
<td>487.83 (899.11)</td>
<td>301.42 (461.19)</td>
</tr>
<tr>
<td>Num_keyerrors</td>
<td>3.79 (5.00)</td>
<td>3.86 (5.32)</td>
</tr>
<tr>
<td>Rate_keyerrors</td>
<td>0.19 (0.20)</td>
<td>0.17 (0.21)</td>
</tr>
<tr>
<td>Query Length</td>
<td>4.66 (2.09)</td>
<td>4.19 (2.12)</td>
</tr>
</tbody>
</table>

**Table 5**

Nonparametric Test Results of Query Typing Behaviors with Different Task Types

<table>
<thead>
<tr>
<th></th>
<th>Mean (SD)</th>
<th>Mann–Whitney $U$ test $Z (p)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FF ($N=365$)</td>
<td>IU ($N=525$)</td>
</tr>
<tr>
<td>Typing Speed (in bytes)</td>
<td>308.08 (456.82)</td>
<td>377.05 (679.48)</td>
</tr>
<tr>
<td>Num_keyerrors</td>
<td>3.53 (4.66)</td>
<td>3.83 (5.19)</td>
</tr>
<tr>
<td>Rate_keyerrors</td>
<td>0.21 (0.20)</td>
<td>0.18 (0.21)</td>
</tr>
<tr>
<td>Query Length</td>
<td>2.97 (1.531)</td>
<td>4.38 (2.118)</td>
</tr>
</tbody>
</table>
5 Discussion

5.1 The Impact of Time Constraint on Users’ Query Typing Behaviors

Our findings suggest that time constraint had significant effects on users’ query typing behaviors, especially the typing speed and query length. Searchers with time constraint tended to be more efficient in typing queries and constructed longer queries. Usually, long queries indicate more specific queries semantically, so the findings may indicate that users tended to issue more specific queries when there was time constraint, aiming to find the specific answer to the question more quickly. This is only a possible explanation and further work is needed to examine the content of queries and their reformulations during search process.

When considering the FF and IU tasks separately, the results were generally consistent with the overall situation. However, there are several points to highlight for each type of tasks. For example, time constraint was only found to have significant effects on typing speed in IU type of tasks but not significant in FF type of tasks. The result might be explained by the fact that users often issued longer queries in IU tasks, and they often reformulated the queries quickly with minor changes in previous queries, which led to much faster typing speed in IU tasks.

5.2 The Impact of Task Type on Users’ Query Typing Behaviors

With respect to task type, the main significant effects were on the rate of keystroke errors and query length. Specially, participants had a much higher error rate and shorter queries in FF tasks than in IU tasks. In regards to the query length, it’s possible that compared with FF tasks, IU tasks are more complex, so searchers need to construct more specific queries to meet their information needs, resulting in longer queries. And in terms of the rate of keystroke errors, we may anticipate more errors in keystrokes in IU tasks since the task is more complex; however, the results demonstrate that users were more likely to make mistakes in cognitively easier tasks.

Furthermore, it is found that when there was no time constraint, searchers made significantly lower rate of keystroke errors in IU tasks than in FF tasks. When users were searching for IU type of tasks, their rate of keystroke errors was lower in NTC than in TC, even though such difference was not significant. Thus, users did make higher rate of mistakes in query typing in TC than in NTC for IU type of tasks, so the time constraints mainly had effect in IU tasks. However, the rate of keystroke errors in FF tasks did not change significantly in TC and NTC. Further analysis is needed to examine whether the pattern would be consistent when searching for other more complex tasks.
6 Conclusion

Our research explored the influence of time constraint and task type on users’ query typing behaviors in a user experiment. The main findings are as follows: First, both time constraint and task type had significant influence on query length. Specially, when searching with time constraint or when they were searching for Information Understanding tasks, searchers tended to issue longer queries. Second, time constraint mainly influenced the typing speed. Searchers often had faster typing speed when there was time constraint than when there was no time constraint, and such difference was also significant in Information Understanding tasks. Third, task type influenced the rate of keystroke errors during query typing. Compared to searchers with Information Understanding tasks, searchers with Fact Finding tasks had higher error rate, and the difference was particularly significant in NTC conditions.

The purpose of this study is to examine the effects of time constraint and task type on users’ query typing behaviors, so as to fill the gap in relevant research and form a more comprehensive and in-depth understanding of users’ query typing behaviors. Better understanding of users’ query typing behaviors may shed lights on the optimization of query autocompletion or query suggestion features for search systems or indicating users’ search contexts for further contextualization of their search process. Although the findings of the current study have answered our research questions to some extent, there are some limitations in this study: first, as a laboratory experiment, the data may not be exactly the same as the real life; second, college students may not be sufficiently representative of other user sample; furthermore, only four variables were considered in this study, which may not be enough to fully reveal users’ query typing behaviors. Further investigations on users’ query typing behaviors are needed, e.g., we will consider individual differences (e.g., pre-search familiarity) and their effects on users’ query typing behaviors, or other variables of query typing behavior, e.g., taking query reformulation behaviors into account in the future, and to examine the relationship between query typing behaviors and their task performance. More studies are also needed to explore the effects of contextual factors on other types of query issuing methods, e.g., image input or oral speaking of queries.

Acknowledgements: This research was funded by National Science Foundation of China (NSFC, project number 71303015).

References

Bentao Zou, Yuefen Wang*

A Comparison Study of Measures to Quantify the Evolution of Prolific Research Teams

https://doi.org/10.2478/dim-2020-0028
received August 15, 2020; accepted September 15, 2020.

Abstract: Scientific research teams play an increasingly significant role in scientific activities. To better understand the dynamic evolution process of research teams, we explored measures that quantify the evolution of prolific research teams. We collected our data from the Web of Science in the field of artificial intelligence, and applied the label propagation algorithm to identify research teams in the co-authorship network. The Top 1‰ prolific teams were selected as our research object, whose node stability and two types of edge stabilities were measured. The results show that prolific teams are much more stable during the evolution process, in terms of both member and membership stability. The measure of stability has varying degrees of impact on teams with different sizes, and small-sized teams get considerably different stability results by different measures.

Keywords: prolific research teams, evolution analysis, comparison study, artificial intelligence

1 Introduction

Teamwork has been the intrinsic characteristic of scientific activity since the very dawn of science, and it is still the characteristic of research today (Perianes-Rodríguez, Olmeda-Gómez, & Moya-Anegón, 2010). The increasing interdisciplinarity and complexity of research problems, combined with the diverse research environment, partly owing to internationalization, call for more and more scientists to collaborate in research teams (Verbree, Horlings, Groenewegen, Van der Weijden, & van den Besselaar, 2015). It has been proved that teamwork plays a crucial role in advancing scientific research, wherein (1) teams dominate over solo scientists in the production of high-impact, highly cited science (Jones, Wuchty & Uzzi, 2008) and (2) teams with consolidated memberships result in an enhanced capacity to foster collaboration in funded research projects, and increase the potential to publish articles in journals covered in the Journal Citation Reports (Martín-Sempere, Garzón-García, & Rey-Rocha, 2008; Rey-Rocha, Garzón-García, & Martín-Sempere, 2006). Furthermore, funding agencies, conference organizations, and academic research institutions across the world have recognized the increasing importance of teamwork, which has also given rise to the “Science of Team Science” (SciTS), an emerging field focused on the examination of processes by which scientific teams organize, communicate, and conduct research (Börner et al., 2010; Stokols, Hall, Taylor, & Moser, 2008).

At its most general, due to constant changes of members and memberships, the formation and development of scientific teams is a process characterized by dynamic evolution (Palla, Barabási, & Vicsek, 2007). The related work can be grouped into three categories: (1) event-based research, which identifies a series of events and transitions (birth, split, dissolve, merge, etc.) to characterize the behavioral patterns of individuals and teams (Asur, Parthasarathy, & Ucar, 2009; Palla, et al., 2007; Takaffoli, Fagnan, Sangi, & Zaiane, 2011); (2) network topology-based research, which measures the topology of the collaboration network at a given moment. The time series analysis of these quantities uncovers the evolution of teams (Barabási, Jeong, Néda, Ravasz, Schubert & Vicsek, 2002); and (3) topic associated-based research, which combines semantic information (mainly

*Corresponding author: Yuefen Wang, Information Management and Information System, School of Economics and Management, Nanjing University of Science and Technology, Nanjing, China, Email: yuefen163@163.com
Bentao Zou, Management Science and Engineering, School of Economics and Management, Nanjing University of Science and Technology, Nanjing, China
the topic of individuals or teams) to implement the evolution analysis of teams (Backstrom, Huttenlocher, Kleinberg, & Lan, 2006; Hopcroft, Khan, Kulis, & Selman, 2004; Osborne, Scavo, & Motta, 2014).

Team stability is one of the most important and significant characteristics during the dynamic evolution process. An earlier study on authors’ collaboration network revealed that large research teams persist longer if they are able to keep a lower stability of teams’ composition, and small teams favor unchanged team members if they purchase longer life spans (Palla, et al., 2007). Cugmas, Ferligoj, and Kronegger (2016) obtained a core co-authorship structure and measured the stability of these cores with the Modified Adjusted Rand Index, by which they found that there were no differences in the average core stability between the natural and technical sciences and the social sciences and humanities. Unlike the above researches, Takaffoli et al. (2011) focused on the behaviors of individuals and proposed a stability metric measuring the tendency of an individual to communicate with the same nodes over the observation time. To avoid small perturbations in the dataset that cause significant changes to the detected clusters (communities), Hopcroft et al. (2004) defined natural communities, including subsets of clusters that remain stable after multiple clustering runs.

Among most of the aforementioned studies, researches were conducted at the macrolevel, which considers almost every team detected and pursues insights about patterns of communication that broadly exist in teams. Price’s Law (Nicholls, 1988) clarifies the relationship between the literature and the number of authors in a field, by which we learned to select prolific teams in artificial intelligence as a significant research object. The stability is highly correlated to the evolution process and the scientific process of teams, whereas little work compares different ways of measuring the stability. For example, the works of Palla, Barabási and Vicsek (2007) and Takaffoli, Fagnan, Sangi and Zaíane (2011) did not cover the stability of memberships in a team. In this paper, on one hand, we try to examine both the author and co-authorship stability of scientific research teams, and on the other hand, track the changes in the structures of research teams in the evolution process. Thus, through study of research teams’ stability measures, the further goal of illustrating the rules and characteristics for them could be reached, which would also be helpful to support the development of scientific and technology.

2 Methods

2.1 Data

We collected our data from the Web of Science core collection database, covering 421,148 papers published between 2009 and 2018 in the field of artificial intelligence. The search strategy was set as “WC = Computer Science, Artificial Intelligence,” and all works about this collection process finished within January 2019.

There were 537,271 authors in our selected data before carrying out any data cleaning process. For author name disambiguation, both co-author and institution information were used. First, authors sharing the same name were grouped into different groups by institutions. Here, we assume that if two authors have the same name (including family name and first name) in the same institution group (and in the same subject field), then these two are identified as one author entity. Considering that authors may have multiple affiliations (authors work at multiple institutions at the same time, or authors leave one institution and move to another), we then classify two authors as the same if the two at different institutions share at least one co-author. This process will be iterated until no intersection is found between every two institutions (for the same name). Every disambiguated author name ends with an extra number (e.g., “Thomas Hanks_1”, “Thomas Hanks_2”), and finally we get 656,668 unambiguous authors.

2.2 Research Teams’ Detection

The overall co-authorship network is constructed via Pajek, and the research teams are detected by a semi-synchronous label propagation algorithm, which combines the advantages of both synchronous and asynchronous models (Cordasco & Gargano, 2010). Initially, each vertex is assigned a unique label that determines which community it belongs to. Subsequently, an iterative process is performed to group vertices into communities. At each step, each vertex updates its label based on the most frequent label among its neighbors. The process iterates until no new label changes occur, and vertices that share the same label are recognized as a community (research team). The algorithm has been implemented in the Python NetworkX package, and we call the corresponding function and detect 140,037 teams. Figure 1 shows the count of teams that vary from sizes. The blue line describes the number of teams with regard
Bentao Zou, Yuefen Wang

Figure 1. The distribution of team sizes. (Teams with single members are equal to single authors; we choose to keep those parts of data for the sake of further comparisons.)

To different sizes, while the gray line corresponds to the cumulative percentage against team size. From Figure 1, we can see that most of the teams (92%) are under the size of 10 members, among which 2-, 3-, and 4-member teams are the Top 3 with the greatest number of teams.

We select the Top 1‰ (140) prolific teams with the greatest number of papers as our research object, which covers 41,068 papers, accounting for 9.75% of all data. Note that 3,802 papers among them are duplicated, indicating interactions between prolific teams. Table 1 shows examples of our selected prolific teams; every team has a unique team code and basic information is integrated (e.g., team size and number of publications). Figure 2 shows the relationship between prolific teams and all teams (except the largest team with 2,536 authors and 9,465 papers).

To capture the evolution of high prolific teams, for a given team, we partition the papers into 10 time slices corresponding to the publication year ranges from 2009 to 2018. We then construct a small collaboration subnetwork for each time window (if exists), and the team’s evolution path is depicted by connecting all subnetworks.

2.3 Stability Measures

In this paper, we examine the stability of both authors and co-authorships (represented by node stability and edge stability in collaboration network, respectively), which is in contrast with most of the previous works relating to this topic with the focus on authors. Equation (1) quantifies

Table 1

<table>
<thead>
<tr>
<th>Rank</th>
<th>Team code</th>
<th>Size</th>
<th>No. of papers</th>
<th>Top 3 authors (no. of papers by full-count)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>86</td>
<td>450</td>
<td>1,577</td>
<td>Veloso, Manuela(83), De Raedt, Luc(63), O’Sullivan, Barry(61)</td>
</tr>
<tr>
<td>2</td>
<td>89</td>
<td>308</td>
<td>1,309</td>
<td>Van Gool, Luc_2(211), Huang, Thomas S.(102), Zuo, Wangmeng(100)</td>
</tr>
<tr>
<td>3</td>
<td>130</td>
<td>248</td>
<td>886</td>
<td>Cao, Jinde_1(219), Huang, Tingwen(189), Zeng, Zhigang_1(135)</td>
</tr>
<tr>
<td>4</td>
<td>108</td>
<td>408</td>
<td>882</td>
<td>Li, Haizhou(118), Xu, Bo_1(86), Tan, Kay Chen_1(84)</td>
</tr>
<tr>
<td>5</td>
<td>120</td>
<td>416</td>
<td>860</td>
<td>Jiao, Licheng(254), Yao, Xin_4(177), Gong, Maoguo(92)</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>136</td>
<td>121</td>
<td>59</td>
<td>184</td>
<td>Pietikainen, Matti(92), Zhao, Guoying(82), Wang, Su-Jing(27)</td>
</tr>
<tr>
<td>137</td>
<td>134</td>
<td>54</td>
<td>184</td>
<td>Coello Coello, Carlos A.(88), Tanaka, Kiyoshi(39), Aguirre, Hernan(26)</td>
</tr>
<tr>
<td>138</td>
<td>24</td>
<td>52</td>
<td>183</td>
<td>Ji, Qiang_1(123), Wang, Shangfei(60), Chen, Jixu(19)</td>
</tr>
<tr>
<td>139</td>
<td>17</td>
<td>95</td>
<td>183</td>
<td>Konar, Amit(118), Nagar, Atulya K.(70), Rakshit, Pratyusha(34)</td>
</tr>
<tr>
<td>140</td>
<td>90</td>
<td>126</td>
<td>182</td>
<td>Chang, Chin-Chen_1(80), Cai, Yiqiao(18), Huang, Yongfeng_3(18)</td>
</tr>
</tbody>
</table>
the relative overlap between two states of the same team (Palla et al., 2007):

\[ S_n(t) = \frac{|N(t_0) \cap N(t_0 + \Delta t)|}{|N(t_0) \cup N(t_0 + \Delta t)|} \]  

(1)

where \(|N(t_0) \cap N(t_0 + \Delta t)|\) is the number of common nodes in \(N(t_0)\) and \(N(t_0 + \Delta t)\), \(|N(t_0) \cup N(t_0 + \Delta t)|\) is the union of \(N(t_0)\) and \(N(t_0 + \Delta t)\), \(t_0\) denotes the first occurred time window of a team, and \(\Delta t\) denotes the length of time gap, \(\Delta t=1\) (year) in this work.

Similarly, let \(E(t_0)\) represent co-authorships at a given time step; replace \(N(t_0)\) with \(E(t_0)\) to form equation (2) for calculating edges stability in two collaboration subnetworks as

\[ S_e(t) = \frac{|E(t_0) \cap E(t_0 + \Delta t)|}{|E(t_0) \cup E(t_0 + \Delta t)|} \]  

(2)

There is another aspect to the edge stability, which is a transformation of equation (2) under a broad time span, comparing edges at the current time slice with edges from all previous slices. In equation (3), denotes the union of edges from \(t_0\) to \(t_i\) as

\[ S_{ec}(t) = \left\{ \begin{array}{ll} S_e(t), & i=0 \\
\frac{\|U_{t_0}^{t_0+\Delta t} E(t) \cap E(t_0)\|}{\|U_{t_0}^{t_0+\Delta t} E(t)\|}, & i>0 \end{array} \right. \]  

(3)

Next, we use \(\delta_j\) as stationarity to measure the average stability during the dynamic evolution process as

\[ \delta_j = \frac{\sum_{t_{max}^{-1}}^{t_{max}} \delta_j(t)}{t_{max} - t_0}, \quad j \in \{n, e, ec\} \]  

(4)

where \(t_{max}\) denotes the last time step before the extinction of a team, and \(\delta_n, \delta_e, \delta_{ec}\) correspond to nodes stationarity, adjacent edges stationarity, and broad edges stationarity, respectively.

### 3 Stabilities of Prolific Teams

#### 3.1 Stability of Authors

Basically, individuals’ activity of joining and leaving result in constant changes in teams. In view of the stability of nodes in collaboration networks, we can quantify the evolution process of authors by equation (1). Figure 3 (Left) reflects the comparison of average node stability in three groups at different time steps (prolific teams, overall, and overall teams that have stationarity larger than 0). Prolific teams far outweigh the others, showing that members in high productive teams are much more stable than those in teams that published fewer papers. Note that the average stationarity of each group leads to a different result, where the overall teams (stationarity>0) lead and prolific teams are the second (\(\delta_e\)). Further analysis of team size and
duration provide reasonable explanations. Almost all prolific teams exist for the entire 10-year period and with the average size far larger than the other two team groups, whereas teams with stationarity over 0 are averaged at 6 members and half of them last no longer than 6 time steps; the details are shown in Table 2. That is to say, small-sized teams exist for a short time and with little change, and in turn gain larger average node stationarity.

For comparisons of distributions among prolific teams, as illustrated by the right graph in Figure 3, it is clear that most of them have a similar dense distribution with the median value at around 0.2. The most dispersed time slices are $t_1$, $t_3$, and $t_9$, indicating that some teams vary a lot at these times in terms of node stability. Correspondingly, there are no significant differences as to the time window at which the team achieves the maximum node stability ($t_1$ and $t_3$ slightly lead). For the minimum node stability, 38% (53 in 140) prolific teams gain the least node stability at $t_9$, as shown in Figure 4.

### 3.2 Stability of Co-authorships

The changes of members naturally affect teams’ evolution, but memberships are an important part of this process as well. When authors in a team remain unchanged, it does not mean that the relationships (co-authorships) among them are kept unchanged. New relationships emerge and existing ones break, forming dynamic evolution. In this paper, we examine two types of co-authorship (edge) stabilities, the adjacent stability of edges (by equation (2)) and the broad stability of edges (by equation (3)).

#### 3.2.1 Adjacent Stability of Edges

To simply compare the stability of co-authorships and authors, the adjacent stability of edges is used to reveal the ratio of common edges in two adjacent time windows. Figure 5 clearly demonstrates that the average adjacent stability of edges is far less than that of nodes. Not only does the value of prolific teams decrease a lot, but also that of the overall team level and teams with edge stationarity larger than 0. The distribution depicted by the boxplot reveals that even the largest edge stability is less than 0.2, which is the average level of node stabilities. Nevertheless, the consensus they reach is that stabilities of prolific teams distribute evenly and densely most of the time, and even the maximum/minimum edge stability is the same as that of nodes, as revealed by Figure 5 (Right). Intuitively, measuring the adjacent edge stability or node stability

![Figure 4. The probability of yielding the max/min node stability of each time step.](image)

shows that the results are similar when considering the probability of yielding the maximum or minimum value of stability at various time steps.

#### 3.2.2 Broad Stability of Edges

This type of stability prolongs the time window of comparing common edges. For a given team, it considers edges in the current time and all edges that occurred in the past. Consider the example in Table 3: $a–f$ are authors in a team, $(a, c)$ represent the edge (co-authorship) between authors $a$ and $c$, and we get $S_i(t_i) = 0$ (i=1,2), $\delta_a = 0$ and, $S_{ac}(t_4) = 0$, $S_{ac}(t_5) = 0.5$, $\delta_{ac} = 0.25$. Obviously, the broad stability of edges is closer to the real evolution process in this situation. With this in mind, Figure 7 plots the result of this measure.

The result derived from the measure of broad stability of edges differs considerably from that of node stability or adjacent edge stability. As Figure 7 (left) shows, edge stability of the prolific teams nearly plummeted over time windows, in sharp contrast with that of overall teams. Also, the boxplot illustrates a much more skewed distribution of prolific teams, sparse at the first several

<table>
<thead>
<tr>
<th>Groups</th>
<th>Pro teams</th>
<th>Overall</th>
<th>Overall (stationarity&gt;0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg size</td>
<td>119.3286</td>
<td>4.6892</td>
<td>9.8832</td>
</tr>
<tr>
<td>Median size</td>
<td>101</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Avg duration</td>
<td>9.9929</td>
<td>2.5975</td>
<td>5.9228</td>
</tr>
<tr>
<td>Median duration</td>
<td>10</td>
<td>1</td>
<td>6</td>
</tr>
</tbody>
</table>

![Figure 4. The probability of yielding the max/min node stability of each time step.](image)
time steps and slightly centralized at the end. From this perspective, memberships in prolific teams are less likely to be changed with the passage of time, but the probability would be near to that of overall teams having edge stationarity over 0 at the last two time steps. In Figure 8, the line represents the probability of yielding the minimum edge stability of every time step keeping the same as that in Figure 6, but the other line describing the probability of gaining the maximum edge stability experienced continuous decreases, which also verifies the above conclusion.

### 4 Discussion

We predicted that different measures for quantifying the evolution process of scientific teams would lead to different results. However, our analyses on different stability measures of prolific research teams varied slightly. We have to note that the average stability is significantly affected by both the size and duration of research teams. Table 2 indicates that prolific teams have significantly larger sizes and longer durations. Furthermore, to avoid the effect from size and duration, Figure 9 shows the comparison of different stationarity measures with prolific teams. Node stationarity largely outweighs edge stationarity, but it is the sparser one in the three measures in the same time. Most of the prolific teams may get a small stationarity, while discrepancies among them are relatively slight. In other words, edge stationarity can efficiently highlight the commonality between prolific teams.

It is worth mentioning that we absolutely cannot draw a conclusion like that members keep more stable than memberships in prolific teams. To confirm this, we randomly select 140 teams (equal to the size of prolific

---

**Figure 5.** Left. The average adjacent edge stability of teams at different time steps. Right. The distribution of edge stabilities of prolific teams.

**Figure 6.** The probability of yielding the max/min adjacent edge stability of each time step.

**Table 3**

*An Example Shows the Function of Broad Stability of Edges*

<table>
<thead>
<tr>
<th>Edges</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012–2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a, c)</td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>(e, f)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a, b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a, d)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
teams) as samples from three data groups: Group A published no more than 50 papers, Group B published 50–100 papers, and Group C with 100–150 papers, respectively. Figure 10 reveals that the phenomenon aforementioned is common to all teams, which has nothing to do with whether the team is prolific or not. In other words, the node stability measure results in a larger value of stability compared with the measure of edge stability.

To make further comparison of the above measures, we draw a graph to describe the number of teams (only for those with stationarity larger than 0) against the team size. As Figure 11 shows, the meaningful point is that different stability measures actually have no difference over large-sized teams, which also supports the above analysis of prolific teams. Specifically, if only teams with a positive value of stationarity are considered, some small-sized teams would be filtered out. That is to say, when your research object covers teams with size less than about 20, your research samples are dependent on the measure you used.

Table 4 shows detailed intersections of three measures. The node stability measure yields the largest number of teams (32,773) with stationarity over 0, followed by the broad edge stability measure (31,901). All teams measured by the adjacent stability measure are included in both the other measures, while more than 80% (26,061) teams measured by the broad edge stability involved in that of node stability.

![Figure 7. Left. The average broad edge stability of teams at different time steps. Right. The distribution of broad stabilities of prolific teams.](image)

![Figure 8. The probability of yielding the max/min broad edge stability over time.](image)

![Figure 9. Comparison of different stationarity measures with prolific teams.](image)
5 Conclusion

There are researchers from various fields engaged in the study of scientific research teams (Azaouzi, Rhouma, & Ben Romdhane, 2019), and numerous findings derived from different perspectives. This paper suggests a closer concentration on (1) prolific teams, as sometimes the “part” is as important as the “whole” and (2) measures about the stability of research teams, as different measures may affect the research object selection for a deeper analysis.

In this paper, we first used a label propagation algorithm to identify scientific teams in the field of artificial intelligence, and then partitioned the data into 10 time slices corresponding to the evolution process of teams. For quantifying the evolution process, we designed two types of measures to calculate the membership stability in scientific teams. Combined with the member stability, we applied three types of stability measures to value the stationarity of the teams detected, especially to prolific teams. The results show that stability of prolific teams significantly outweigh the others, in terms of both author stability and co-authorship stability. For the given 10 time slices, prolific teams are more likely to reach the least stability at the last slice, and have no significant difference to yield the maximum or minimum stability at other slices. In the discussion, we further analyzed the dissimilarity caused by different measures, and concluded that measures largely affect the quantifying of small-sized teams, and have little influence on teams with a large number of members. Researchers should proceed their studies with caution when filtering out data under the criterion of stationarity. Future works will focus on illustrating the relationship between scientific performance and the evolution process of teams.

Acknowledgement: The authors wish to acknowledge the assistance of the National Social Science Fund Major Project (No. 16ZDA224).
References


Exploring Significant Characteristics and Models for Classification of Structure Function of Academic Documents

Bowen Ma, Chengzhi Zhang*, Yuzhuo Wang

Abstract: With the increasing abundance of literature resources, how to acquire knowledge elements efficiently and accurately is the key to achieving accurate literature retrieval and utilization of available literature resources. The identification of the structure function of academic documents is a fundamental work to meet the above requirements. In this study, the proceedings of the Association for Computational Linguistics (ACL) conferences are used as the primitive corpus, and the training corpus of chapter category is obtained by manual annotation. Based on the chapter titles and the in-chapter texts, traditional machine learning and deep learning models are both used for classifier training. Our results show that the title of a chapter is more beneficial to the identification of the structure function of academic documents than the in-chapter texts. The highest $F_1$ value in our experiments is 0.9249, which is obtained on the traditional logistic regression (LR) and support vector machine (SVM) models (slightly higher than on the convolutional neural network [CNN]). And through the experiment of adding other chapter characteristics based on the traditional model, we find that combining the relative position of chapters can effectively improve the classification performance. Finally, this study compares the results of experimental groups with different methods, analyzes the misclassification of the structure function of academic documents, and points out the main direction to improve the classification performance in the future.

Keywords: structure function of academic documents, text classification, characteristic selection, machine learning, deep learning

1 Introduction

With the rapid development of science and technology, document-based resources are gradually being enriched. In the field of academic research, the output of scientific literature also shows an explosive growth (Xia, Wang, Bekele, & Liu, 2017). Researchers urgently need to extract knowledge elements, such as knowledge entities, knowledge relations, and so on, from massive full-text literature so as to improve the accuracy and efficiency of literature retrieval and optimize the organization and classification of knowledge resources in literature. However, knowledge elements often play different roles in different structure functions of academic documents, and in order to acquire knowledge elements more accurately, it is necessary to grasp the angle of expression and importance of literature information transmitted by knowledge elements. Therefore, the automatic identification of the structure function of academic documents is surely the basis of full and accurate extraction of knowledge elements and can further meet the requirements of the mining and utilization of massive literature. In addition, the structure function of academic documents can be introduced into bibliometrics (Hu, Chen, & Liu, 2013), as well as in citation evaluation and other studies as position characteristics (Zhu, Turney, Lemire,
This study focuses on the identification of the structure function of chapter granularity, i.e., to classify the chapters in academic papers based on their structure functions.

The core contribution of this article is as follows:

(1) The first contribution of this study is adopting a variety of classification models, including the traditional machine learning models and the deep learning models. Meanwhile, in terms of the adoption of different characteristics in the experiment, we not only fully compare the lexical characteristics of in-chapter texts and chapter title texts but also further excavate the nonsemantic potential characteristics that are probably associated with the categories of the structure function of chapters by the way of characteristics combination. Through multiple groups of comparative experiments, the effective characteristics of and the better ones among the classification models are screened out, which lays a foundation for further improving the overall classification performance through multiple characteristic fusion schemes in the next step.

(2) Another contribution of this study is the large-scale tagging corpus containing >4100 academic documents built by manual annotation in our study, which also provides a guarantee for the reliability of this study.

### 2 Related Works

Identifying structure functions of the chapters in documents is regarded as a classification task, and various classification models and methods have been proposed in existing work. Table 1 shows some major classification models that classify the structure function according to a certain scientific logic.

We can see that different classification systems are suitable for different tagging granularities, which also determine the different levels of research significance.

In the research of automatic recognition of multigranularity structure function, the methods used can be divided into two main categories:

(1) Identification based on rule-based methods

The core of the rule-based methods is to find the characteristics that are closely related to the corresponding categories of structure function. The popular rules include heuristic words or phrases, syntactic rules, and patterns based on combination of features. Kim, Le, and Thoma (2000) achieved the automatic indexing of titles, authors, abstracts, and other structural information in papers through the layout characteristics of key information in the documents and optical character recognition (OCR) technology. Gupta and Manning (2011) extracted detailed structural information in the titles and abstracts of literature by means of pattern matching, syntactic analysis, and the construction of seed rule words.

The rule-based methods can achieve good results in specific scenarios, but with the expansion of corpus and refinement of classification systems, the rule-based methods consumes more manpower, and its actual performance is difficult to guarantee.

(2) Identification based on learning-based methods

Some researchers regard the classification of structure function as a text classification issue. Teufel and Moens (2002) used their proposed argumentative zoning (AZ) model to conduct automatic classification research. They adopted Naive Bayes (NB) to train the corpus containing 80 computer linguistics papers and added many other characteristics, including position and length of the sentence, term frequency–inverse document frequency (TF-IDF) weight, syntactic characteristics, references,
formulas, and so on. Kambiz et al. (2018) used support vector machine (SVM), logistic regression (LR), and NB classifiers to identify from the perspective of sentence granularity; they further added the voting mechanism to further improve the classification performance. Other researchers use the idea of sequence labeling to excavate the positional rules of different structure function parts. Lu, Huang and Cheng (2014) adopted the conditional random field (CRF) model and various features to identify the structure function of chapters based on the chapter titles and finally achieved the F1 value of 89.6%.

With the rapid development of deep learning, the application of neural network to automatically extract text features provides a new thought for the identification of structure function. Wang, Lu, Liu and Cheng (2019) used long-short-term-memory (LSTM) and convolutional neural network (CNN) models to conduct training at three levels, including the chapter title, the in-chapter text, and the chapter paragraph, respectively, and optimized the model by introducing the voting mechanism, obtaining the highest F1 value of 84%.

In general, various traditional machine learning methods, new ideas of deep learning methods, and various auxiliary characteristics in current studies are worth being referred to. However, there is lack of a large-scale manually tagged corpus of structure function in the chapter level for academic documents, and few research works have compared traditional models with deep learning models.

So, in this paper, we adopt the above two kinds of methods to train the classifier and explore a variety of characteristics or combination of characteristics to find a better classification model.

3 Methodology

3.1 Framework

As shown in Figure 1, this study took the proceedings of the Association for Computational Linguistics (ACL) conferences over the period 1989–2015 in extensible markup language (XML) and portable document format (PDF) formats as the original corpus and then imported the corpus into the online annotating platform (Ma, Wang, & Zhang, 2020). The structure functions of valid chapter samples were annotated by some annotators, and finally, we obtained the tagging results with high consistency. The core characteristics used in our experiment have two parts, namely, in-chapter texts and chapter titles, so that we can compare the importance of the two kinds of text characteristics. We designed the experiment adopting traditional machine learning models by constructing eigenvectors of chapters and attempting to combine additional characteristics to optimize the performance. And we also set the comparative experiment using deep learning methods to automatically obtain features and classify the chapters.
3.2 Data Annotation

This study selected the corpus of ACL conferences from 1989 to 2015 and downloaded (1) original data in XML format from the ACL Anthology Reference Corpus (Bird et al., 2008) and the (2) corresponding PDF files from the ACL Anthology website, yielding a total of 4190 articles. Then, the XML files were parsed and uploaded to the annotation platform with the corresponding PDF files for manual annotation (two persons in a group to annotate a part of the corpus).

We invited 12 annotators (divided into six groups) and two checkers to annotate the structure function of chapters. Six categories of structure function are shown in Table 2. Before the formal annotation, we required everyone to learn the tagging standard, and finish prelabeling work. The checkers reviewed the doubtful tagging items given by annotators to improve the accuracy of annotated results.

Table 2
The Description of the Structure Function

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>The research background, problems, purposes, and so on</td>
</tr>
<tr>
<td>Related works</td>
<td>A summary of the relevant work</td>
</tr>
<tr>
<td>Method</td>
<td>Description of research methods</td>
</tr>
<tr>
<td>Evaluation &amp; results</td>
<td>Experimental setup, process, evaluation method, results, and discussion</td>
</tr>
<tr>
<td>Conclusion</td>
<td>The summary of the research and the prospects for future work</td>
</tr>
<tr>
<td>Other</td>
<td>Other chapter categories</td>
</tr>
</tbody>
</table>

Table 3
Statistical Analysis and Kappa Value

<table>
<thead>
<tr>
<th>Group ID</th>
<th>Article Number</th>
<th>Chapter Number</th>
<th>Kappa Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>657</td>
<td>3874</td>
<td>0.859</td>
</tr>
<tr>
<td>2</td>
<td>669</td>
<td>3998</td>
<td>0.816</td>
</tr>
<tr>
<td>3</td>
<td>656</td>
<td>3879</td>
<td>0.820</td>
</tr>
<tr>
<td>4</td>
<td>657</td>
<td>3835</td>
<td>0.726</td>
</tr>
<tr>
<td>5</td>
<td>693</td>
<td>4110</td>
<td>0.778</td>
</tr>
<tr>
<td>6</td>
<td>660</td>
<td>3896</td>
<td>0.810</td>
</tr>
<tr>
<td>Overall</td>
<td>3992</td>
<td>23592</td>
<td>0.801</td>
</tr>
</tbody>
</table>

3.3 The Construction of Feature Vector of Traditional Model

Generally, the text feature vectors are constructed manually in traditional learning methods. First of all, the data preprocessing reduces the dimension of the lexical features, and in terms of feature selection, we refer to the work by Yang and Pedersen (1997), which found that...
information gain (IG) and chi-square test (CHI) are the most effective methods in the experiment they designed; further, this paper finds that through preliminary experiment CHI has better performance in the current task, so we adopt CHI to conduct the feature selection, and finally 5357 features of words are obtained. And then, TF-IDF is used as the feature weight to construct the text vector of each chapter. Finally, a chapter is represented by a vector of 5357 dimensions. There are fewer lexical features in chapter titles, so we got 4967 features of words without feature selection.

Meanwhile, this paper also attempts to merge in some nonsemantic characteristics to better identify the structure function of chapters with reference to previous works (Guo et al., 2010; Liakata, Saha, Dobnik, Batchelor, & Rebholz-Schuhmann, 2012), which focus on the identification of the structure function of sentence granularity. We find that the nonsemantic characteristics mainly fall into three categories: Location, History, and Citation. Location refers to the positional relationship between the text fragment to be classified and the full text of the article or the specific parts that the study focuses on, so as to obtain location information. History feature usually refers to the category of previous text in the same article, which conveys a kind of sequence information. Citation represents the number of references in the text to be classified.

In this paper, Location and Citation are selected as additional features and introduced to the classification system. History is excluded because the characteristic in the test set is based on the predicted results, which may cause the accumulation of errors and interfere with the evaluation of overall classification performance, so it is not adopted in our experiment. We further incorporate some potential characteristics mentioned by Ahmed & Afzal (2020), namely, the number of references, the number of figures, the number of tables, and so on, which are easy to be ignored. In order to simplify the number of additional characteristics, this paper views the sum of tables and figures as another characteristic introduced into the experimental exploration.

It is necessary to illustrate that we also take into account the relevant characteristics of chapter length, mainly including the number of paragraphs and sentences in the chapter. However, because the corpus used in this paper has no level of paragraph, and because in the preliminary experiment, it is found that the characteristic of sentence number has little effect on the classification model, this kind of characteristic is not added in our experiment. The final additional characteristics used in this paper are described below in detail.

### (1) Relative position of chapter (Loc)
Relative position is a characteristic with potential value to the classification of structure function, and the general interval of the chapter category can be roughly judged. In this study, the sequence number of chapters in the paper (counting from “1”)/the total number of chapters in the paper (except chapters with “other” label) is taken as the value of relative position of chapters.

#### (2) The number of citations (Cite)
There are some differences in the number of citations in different chapter categories. For example, the number of citations in the Introduction and Related Work chapters may be relatively large, while the number of citations in other chapters may be relatively small. Therefore, the number of citations (counting by articles) is of positive significance to the classification of structure function.

#### (3) Number of figures and tables (F&T)
Figures and tables are commonly used in paper writing. For empirical studies, the number of figures and tables may be relatively concentrated in the experimental chapter. Therefore, this study counts the sum of the numbers of different figures and tables in each chapter as another auxiliary classification characteristic.

The specific characteristic introduction approach aims to divide the additional characteristic value into different intervals and then generate a set of random numbers for each interval, finally concatenate to the original feature vector, as shown in Equation (1). Thus, \( x_i \) is the vector representation for chapter \( i \).

\[
x_i = [x_i^{\text{Lexical item}}; x_i^{\text{Additional features}}]
\]  

#### 3.4 Classification Model Selection

### 3.4.1 Traditional Machine Learning Models.

Four common and classic machine learning models were selected in this paper: NB based on multinomial distribution, LR, K-Nearest Neighbor (KNN) model, and SVM. Because our research belongs to the multiclassification task, the LR and SVM models should adopt the adaptive scheme of multiple classification, such as “One vs Rest” or “One vs One”. The specific setting is determined by the preliminary experiment.

### 3.4.2 Deep Learning Model.

We selected bidirectional LSTM (Bi-LSTM), which is a time-series based model and suitable for feature extraction of
longer text, to extract the features in chapter text. This paper also drew on the model of Hierarchical Attention Networks (HAN) proposed by Yang Z., Yang D., Dyer, He, Smola, and Hovy (2016) to compare the model of word-level coding through Bi-LSTM.

This study also set up a contrast experiment based on the chapter titles. In order to extract the features of this short text sequence, we used the CNN model proposed by Kim (2014) to achieve feature encoding and classification.

4 Experiments and Analysis of Results

4.1 Experimental Setup

We designed the four experiments based on traditional models and the deep learning model, as well as two core characteristics containing in-chapter texts and chapter titles. The experimental setup in the traditional model is as follows:

**Experiment A:** Four classifiers – NB, LR, KNN, and SVM – were used in the experiment. And the feature vectors were built based on in-chapter texts. Then, we also introduced three other chapter characteristics to assess whether the performance of the classifiers can be improved.

**Experiment B:** The feature vectors were built based on chapter titles, and the preprocessing aimed at just removing the digital characters without any feature selection. The four classifiers described in Experiment A were also used for training.

The evaluation indexes were calculated by five-fold cross-validation. The optimal model parameter settings were selected through preexperiments, as shown in Table 4.

The experimental setup in the deep learning model is as follows:

**Experiment C:** Bi-LSTM, Hierarchical Bi-LSTM (sentence level), and Hierarchical Bi-LSTM with attention models were used to extract the features of in-chapter texts and, then, the softmax layer was used for classification.

**Experiment D:** Since the chapter title is short and lacks the structure of complete sentences, this study decided to utilize the CNN model to extract feature vectors and then complete the classification through the softmax layer.

In the experiment, the overall data are divided into training set, test set, and validation set in a ratio of 8:1:1. The specific parameter settings are shown in Table 5.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning rate</td>
<td>0.01</td>
</tr>
<tr>
<td>Batch size</td>
<td>128</td>
</tr>
<tr>
<td>Dropout</td>
<td>0.5</td>
</tr>
<tr>
<td>Bi-LSTM hidden size</td>
<td>100</td>
</tr>
<tr>
<td>CNN filter number</td>
<td>50</td>
</tr>
<tr>
<td>CNN filter height</td>
<td>1, 2, 3</td>
</tr>
</tbody>
</table>

4.2 Evaluation Method

In this paper, the precision (P), recall (R), and F1-score (F1) values are used to evaluate the classification performance of each model. To measure the performance of the overall classification system, we used the macro averages of precision, recall, and F1-score values as the evaluation indexes.

4.3 Analysis of Results

(1) Results from traditional models

The results of the two experimental groups from the traditional models are shown in Table 6. Experiment A is based on in-chapter texts, and Experiment B is based on the chapter titles.

In Experiment A, LR and SVM models show better results, and their performance indexes are basically the same, while KNN performs poorly. After further integrating the characteristics of relative position, the number of citations, and the number of figures and tables, all models are greatly improved, among which NB and LR are more obvious, and the F1 values of LR and SVM are further improved to 0.84. From Experiment B, we find that the results only based on the chapter titles are significantly better than those based on in-chapter texts; the F1 values
of all classifiers are >0.9, and the LR and SVM models also performed better.

In order to further analyze the impact of three additional chapter characteristics on the classification effect, this study took the LR model as an example to analyze the impact of different characteristic combinations on the results, as shown in Table 7:

It can be seen that the relative position characteristic of chapters plays a significant role in the improvement of classification performance. On only adding the loc characteristic, the result is close to the one obtained by adding all three characteristics, while addition of cite and f&t characteristics has limited improvement in classification performance.

(2) Results from deep learning model
The results of the two groups of experiments based on deep learning are shown in Table 8. Experiment C is based on in-chapter texts, and Experiment D is based on chapter titles.

Compared with the nonhierarchical network, the hierarchical network in Experiment C has a better effect. However, after the addition of attention mechanism, the effect of the model is reduced, and the F1 value is close to that using nonhierarchical network. It is speculated that after highlighting the importance of certain features according to the weight, it weakens the grasp of the global features of text sequence, thus leading to the deterioration of classification performance. In Experiment D, the significance of chapter titles for classification is further proved, and the F1 value reached 0.9217, which is equivalent to the best result from the traditional model.

5 Discussion
In this section, we first conduct a comprehensive analysis of the experimental results through the specific evaluation indexes of each chapter to explore effective models and key classification characteristics. Then, the emphasis on future improvement is analyzed by confounding matrix.

5.1 Key Characteristics and Model Analysis
We further made a comparative analysis of the experiments based on the in-chapter texts or titles, traditional models, or deep learning models. The training results with the best performance in Experiments A, B, C and D were listed, and the evaluation indexes of each chapter category were compared. The results are shown in Table 9.
It can be seen from the table that training using chapter titles can greatly improve the classification performance in both the traditional model and the deep learning model. In particular, the identification performance of introduction and related work is improved more obviously.

By comparing the optimal results of Experiments A and C, we find that the classification performance of most chapters is improved except for the related work, which indicates that the improvement after adding the loc characteristic for the recognition performance of related work is relatively limited, while the deep learning model is better for the extraction of features in this chapter. And we can analyze that the deep learning model has a stronger ability to extract text feature only when considering the features of in-chapter texts. On comparing the optimal results of Experiments B and D, we find that the classification performance of the two methods is equivalent. The recognition based on related work and evaluation & result is obviously better in the traditional model than in the CNN model, while the recognition based on conclusion is worse. Thus, the feature extraction of chapter titles based on the CNN model has potential for further research.

5.2 Chapter Misclassification Analysis

Furthermore, the confusion matrix is used to analyze the optimal training results in all experiments, as shown in Table 10.

It is obvious that the chapters of method and evaluation & result are mutually misclassified. Many evaluation & result chapters are divided into method, which indicates that the model based on chapter titles has certain limitations in the recognition of these two categories of chapters; therefore, it is necessary to further improve the classification performance by combining the characteristics of in-chapter texts and exploit more effective information.

This paper argues that the main idea to solve this problem is to add additional feature information for the differentiation of these two categories of chapters. From the perspective of traditional artificial feature engineering, the number of high-frequency words and adjectives (comparative degree) can be considered, and the N-gram model can be further used to expand the features and improve the identification capability of these two categories of chapters. And from the perspective of deep learning, it is essential to incorporate edge information between different chapters, which could be interpreted as additional information derived from the
conversion of chapter category in the article. The specific approach to practice the idea is the fusion of information of the previous chapter(s), current chapter, and next chapter(s), to make the classification system obtain the boundary information, which can also be understood as a kind of sequence information.

6 Conclusion

Taking large-scale academic documents in the field of computational linguistics as the corpus, this paper explores the automatic classification of structure function of chapter granularity in academic documents. The annotated corpus with high consistency was obtained through manual annotation assisted by the self-built online annotating platform. The traditional machine learning model and deep learning method are used respectively to train the classifier. The results show that the classification performance based on chapter titles is better than that based on in-chapter texts. In the traditional model, the LR and SVM methods perform better, and the F1 value of 0.9249 is the highest obtained by training based on chapter titles; we also find that the relative position of chapters is the key characteristic to improve the classification performance in the training based on the in-chapter texts.

In the deep learning model, the experimental results based on the in-chapter texts are better than those in the traditional model without additional characteristics and are equivalent to the results with additional characteristics, which shows that the deep learning model is superior to the manual construction in the extraction of information in the in-chapter texts. The experiment based on chapter titles got the highest F1 value of 0.9217 through the CNN model.

In the future, this paper will retrieve more effective characteristics to optimize the ability to distinguish between method and evaluation & result. Inspired by previous experiments, the specific attempt would be to combine the characteristics of in-chapter texts on the basis of the characteristics of chapter titles. And we can also introduce the concept of relative position to the deep learning model and view the boundary information between the chapters as the locational sequence information, so as to improve the overall classification ability of chapter categories of structure function.

Acknowledgments: This work was supported by the National Natural Science Foundation of China (grant No.72074113). We thank Song Y, Xie Y, Hao W, Wang R, Ding R, Li Z, Li X, Liu L, Qian Y, Meng M, and Hu S for annotating the corpus. We also thank Zhao Y for checking the corpus, whose earnest and diligent efforts are the foundation for the research.

References


Qin Qin, Qing Ke*, Jia Tina Du, Yushan Xie

How Users’ Gaze Behavior Is Related to Their Quality Evaluation of a Health Website Based on HONcode Principles?

https://doi.org/10.2478/dim-2020-0045
received August 15, 2020; accepted September 15, 2020.

Abstract: While the health website is an easily accessible source for patients to use when seeking health information, the quality of online health information has been a critical issue that concerns all stakeholders in healthcare. The aim of this research was to examine the relationship between users’ evaluation of the health website quality and their gaze behavior on the web pages. Eye tracking and a self-report questionnaire based on the HONcode principles were used to address the objective. We found that (1) the evaluations of authority, privacy, financial disclosure, and advertising policy are positively correlated with the fixation count and total fixation duration toward corresponding page components, while the evaluations of complementarity and attribution are negatively correlated with the fixation count and total fixation duration to corresponding page components; and (2) the fixation count and total fixation duration toward health information sources are negatively related to the evaluation of health website quality, while the fixation count and total fixation duration to site owner are positively related to the quality evaluation. Users’ attention to page components is closely related to the evaluation of principles, and also has a certain impact on the overall quality evaluation of a health website. Based on the findings, our research may serve to improve the health website design and be a foundation to develop an automatic evaluation approach of the health website quality.

Keywords: health website, online health information, quality evaluation, eye tracking, HONcode

1 Introduction

Patients nowadays use health-related websites as important information sources; however, some issues such as its fast and uncontrollable growth, heterogeneity, and lack of publishing control have focused the attention on the quality of these websites (Moreno, Morales del Castillo, Porcel, & Herrera, 2010). The quality assessment of online health information becomes especially important due to the fact that a low-quality health website might be threat to patients’ health and life safety (Daraz et al., 2019). The debate about how to evaluate the quality remains open and we can find different viewpoints and approaches on this issue, such as the evaluation of the substance and formality of information content, as well as the design of technological platforms (Zhang, Sun, & Xie, 2015).

Professional evaluation instruments for health information are generally used by researchers and medical experts, whereas evaluating the quality of a health website remains a major challenge for information searchers. In addition, the popularity of health websites is often affected by users’ evaluations of information quality, and thus it is necessary to know how users evaluate the quality of health websites in the online environment (Yi, Yoon, Davis, & Lee, 2013). Few evidences could be found in the current literature to identify how people pay attention to different page components in their quality evaluation process, as most studies have used self-report methods or memory scores to evaluate the websites. Eye-tracking technology is...
a growing field used to detect eye movements and analyze the human information processing characteristics. It can be used to visualize users’ evaluations of the information quality by illuminating the decision-making through the examination of eye movement patterns (Mele & Federici, 2012). In the current study, we collected both participants’ self-report and their eye-movement data to make a correlation analysis. The aim of the research was to detect the relationship between users’ attention allocation to page components and their evaluations of the quality of health websites. Specifically, we investigated the following research questions:

RQ1: What is the relationship between users’ attention allocation to page components and their evaluations of the corresponding quality principles (e.g., the relationship between attention to qualifications of doctors and the evaluation of authority)?

RQ2: Which page components that users pay attention to have impacts on the evaluation of the health website quality?

2 Literature Review

2.1 Principles, Instruments, and Methods of the Quality Evaluation of Health Website

The quality of health information concerns every stakeholder involved in healthcare, particularly since the web has made access to this information unprecedentedly easily. A major focus of the research was to examine and synthesize the principles used for the evaluation of website quality (Wyatt, 1997; Post & Mainous, 2010; Fahy, Hardikar, Fox, & Mackay, 2014). Silberg, Lundberg, and Musacchio (1997) identified four sets of objective principles for judging the quality of website content by referring to the principle applicable to the print world: authorship, attribution, disclosure, and currency. Eysenbach, Powell, Kuss, and Sa (2002) grouped the quality principle into five categories: technical, design, readability, accuracy, and completeness. Based on the elements being evaluated in previous research, the quality principle can be grouped into two major categories: (1) principle evaluating the information content and (2) principle evaluating the design of website (Zhang, Sun, & Xie, 2015). Focusing on the content and design of health websites, some quality principles have been applied in a large number of studies. For example, researchers attempted to evaluate the content of health information from the principle such as accuracy (Buhi et al., 2010; Foster, Florea, Fahrenbruch, Blackwood, & Rea, 2017), completeness (Dutta-Bergman, 2004; Alwhaibi et al., 2019), currency (Williams, 2003), credibility (Avery, 2010; Ma & Atkin, 2017), and readability (Chi, Jabbour, & Aaronson, 2017; Lam, Roter, & Cohen, 2013). Website design was usually evaluated by principles of accessibility (Rezniczek et al., 2015), esthetics (Weymann, Härter, & Dirmaier, 2014), and interactivity (Klila et al., 2013).

Some proprietary instruments were developed to unify and standardize the quality of health websites, for example, DISCERN (Azer, Alghofaili, Alsultan, & Alrumaih, 2018; San Giorgi, de Groot, & Dikkers, 2017), HONcode (Ahmed, Sullivan, Schneider, & McCormy, 2012; Patton, George, & Hollowell, 2014), LIDA (Schreuders et al., 2017), JAMA (Bizzi, Ghezzi & Paudyal, 2017), and readability formulas (Hansberry et al., 2017; Sax, Razak, Shetty, Cohen, & Levi, 2019). A great deal of previous research has focused on the assessment rules and approaches, and suggest that self-report questionnaires are the most prevalent method of evaluation (Maramba, Chatterjee & Newman, 2019). However, questionnaires could not provide reliable and objective evidences because of the human memory limitation and subjectivity. To mitigate the effects of bias and ambiguity in questionnaires, several researchers have tried to integrate eye-tracking and questionnaire analysis in their studies, and the effectiveness of this methodology is demonstrated (Suzuki, Shirahada, Kosaka, & Maki, 2012; Mahardika, Wibirama, Ferdiana, & Kusumawardani, 2018).

2.2 Impacts of Website Features on the Quality Evaluation of Health Websites

The features of health websites are important factors in determining patients’ intention to accept the information (Mathialagan et al., 2018). In reviewing researches about the evaluation principles of health websites, information content principles such as accuracy, completeness, and readability of health information have been influenced by the features of disease information, treatment methods, medication instructions, and information sources in a great deal of previous research (Kräenbring et al., 2014; CandelarioVazquez, Jackson, & Reilly, 2017; Phillips, Lam, & Palmisano, 2014; Stinson et al., 2011). Researchers have attempted to explore the perceived information credibility from the features of health information contents, sources, qualifications of doctors or authors, websites contact information, and site sponsors (Avery, 2010; Freeman & Spyridakis, 2009; Liao & Fu, 2014; Gao, Tian, & Tu, 2015). Availability of hyperlinks and websites contact
information is used to evaluate the accessibility of health information in several studies (Skovrlj, Silvestre, Ibeh, Abbatematteo, & Mocco, 2015; Scott & Errett, 2018). As for the quality principles of website design, esthetics is mainly reflected in the interface design such as site layout, color schema, and images; the design of advertising affects the evaluation of esthetics in these ways (Moshagen, Musch, & Göritz, 2009). It is now well established that evaluation of the interactivity is based on the way of health information exchange, for example, Q&A communities, online consultations, and personalized service (Lustria, 2007; Kim, 2011; Tsai, 2011). The above brief review suggests that users’ evaluation of health websites is mainly based on the interface components of a website.

2.3 Application of Eye-Tracking Technology in the Quality Evaluation of Health Websites

Eyesight is an important aspect of cognition, and thus the data of eye movement could be valuable for detecting human perception. Eye-tracking technology makes it possible to gather qualitative and quantitative data from participants. In recent years, eye-tracking technology has attracted attention in the field of human–computer interaction, and gradually can be applied to researches on user information behavior. The evaluation of website design is a research focus of eye-tracking technology applied to evaluation of health website quality. For example, several studies have used eye-tracking techniques to examine how users interact with the web interface and put forward some suggestions for improving the interface design (Rules & Xie, 2011; Harris, Sillence, & Briggs, 2009). Besides, some researchers (Koybasi & Cagiltay, 2012; Cho et al., 2019; Schaarup et al., 2015) have used eye-tracking technology to evaluate the usability of online system, which increased the accuracy and completeness in the usability tests. However, in a systematic review, Maramba et al. (2019) have identified, explored, and summarized the methods used in the usability testing of eHealth applications, and suggested that automated methods such as eye tracking have not gained traction in evaluating health apps. In addition to the evaluation of website design, recent studies try to explore the influences of individual characteristics on users’ judgment on the quality of health information content, such as age (Inthiran & Macredie, 2018), memory capacity, and emotion (Chang & Gwizdka, 2018).

The present research focuses on the quality evaluation of a health website; eye movement evidence was collected while users visited the website, and the HONcode instrument was used to obtain their quality evaluation data. In this way, we attempt to explore the relationship between users’ attention allocation to webpage components and their evaluation of health website quality.

3 Method

3.1 Participants

Participants were recruited using WeChat, QQ, and e-mails in a university and finally eye-movement data from 31 participants met the accurate rate requirement. The total sample of 31 (below 27 years of age) consisted of 9 males and 22 females; they were 26 undergraduates and 5 postgraduates. All participants had a good educational background (above bachelor degree) which ensured that they would have the ability to judge the information quality of health websites. Most of the participants had a high demand for health information (n = 28, 90%) but seldom visited health websites (n = 30, 97%). Their knowledge about health websites were at low levels, which could avoid the perception bias caused by their familiarity with the websites.

3.2 Measurements

3.2.1 Quality Evaluation

In this study, quality evaluation was measured using principles of HONcode (the Code of Conduct for Medical/Health Websites). HONcode is a United Nations approved agency that accredits websites to publish health-related information. It is one of the most commonly used instruments to evaluate the quality of health websites and contains eight principles (Table 1): authority, complementarity, privacy, attribution, justifiability, transparency, financial disclosure, and advertising policy. Based on the principles of HONcode, a 5-point Likert scale questionnaire was designed and served as a tool of quality evaluation; the score from 1 to 5 indicates an increasing value of each principle, and the scores of eight principles were added to reflect the participants’ whole evaluation of the website quality.
3.2.2 Attention

We used two metrics measuring participants’ gaze behavior: fixation count (FC) and total fixation duration (TFD). FC explains the number of fixations within a specific area of interest (AOI) or AOs group, while TFD presents the length of the fixation in seconds within a particular A0I or AOs group. FC and TFD represent how many times and how long participants spend on the AOI or AOs group. In the context of information behavior research, these two metrics have been frequently used to examine the attentional effort spent by users on specific areas of web pages (Lorigo et al., 2008). Each measure was calculated as mean values from the eye-tracking software (count numbers in FC and seconds in TFD).

3.3 Data Collection

We defined eight A0I groups matching with the eight principles of HONcode. The division of AOI groups is based on the meaning of each principle and the existing research experience. Medical advice provided and hosted on the websites will only be given by medically trained and qualified professionals while the qualifications of doctors represent the authority of health information (Al-Mandhary, Al-Zakwani, & Afifi, 2007). The function of online consultations reflects the complementarity of the website and has an impact on the relationship between patients and doctors in reality (Xiang & Stanley, 2017). Information about the identity of the patient or visitor is regarded as privacy and is respected by the website. In order to meet the need of attribution, the health information contained on this site will be supported by clear references to source data and, where possible, have specific HTML links to that data (Boyer, Selby, Scherrer, & Appel, 1998). For justifiability, any claims relating to the benefits of a specific treatment, commercial product, or service will be supported by appropriate evidence (Neumark, Flum, Lopez-Quintero, & Shtarkshall, 2012). The website contact information should be transparent for visitors to seek further information or support and the owner of the website should be disclosed. If advertising is a source of funding, it will be presented to viewers in a manner and context that facilitates differentiation between it and the original material created by the site (Choi, 2013). To sum up, Table 2 lists the number of A0Is, page components, and snapshots in each A0Is group.

In our experiment, quality evaluation data were collected from a self-report questionnaire and the eye-movement data were collected using the Tobii Pro X3-120 eye tracker connected to a Dell optiplex 7020 desktop. A total of eight AOI groups were created on 658 webpages visited by all participants (with 20 invalid webpages eliminated).

3.4 Procedure

The experimental scenario was built on a very famous Chinese health website (www.120ask.com); this website has typical elements, such as a searching box, navigation menus, advertisements, and informative contents on Q&A pages. Before the formal experiment, participants got familiar with the experimental environment, understood
the experimental process, and signed the informed consent. In order to obtain the desirable eye-movement data, a research assistant helped participants to calibrate both their eyes, including guiding participants to visualize their position and adjust their sitting distance and position. After the calibration, participants then started the experiment by clicking “Start Recording” and the page automatically jumped to the Chinese health website. Each participant freely visited the webpages and the whole browsing time for each participant was limited to 10 min. After completing the eye-tracking experiment, participants were asked to fill in questionnaires including personal basic information (gender, age, education, health website visiting frequency and health information needs), and their evaluation of the health website quality.

4 Result

4.1 Correlations between Eye Movement and Evaluation of Quality Principles

Correlations between either FC or TFD in each AOIs group and the values of principles are shown in Table 3, which summarizes the direction, strength, and significance of correlation coefficient for these measures. The evaluations of authority (0.790**), privacy (0.636**), financial disclosure (0.447*), and advertising policy (0.873**) were found to be positively correlated with the FC to the corresponding page components, the evaluations of complementarity (−0.713**), and attribution (−0.470**) are negatively correlated with the FC to corresponding page components, while the correlations between the

<table>
<thead>
<tr>
<th>AOIs group</th>
<th>Number of AOIs</th>
<th>Page components</th>
<th>Snapshot</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authority</td>
<td>805</td>
<td>Qualifications of doctors</td>
<td></td>
</tr>
<tr>
<td>Complementarity</td>
<td>192</td>
<td>Online consultations</td>
<td></td>
</tr>
<tr>
<td>Privacy</td>
<td>495</td>
<td>Users’ information</td>
<td></td>
</tr>
<tr>
<td>Attribution</td>
<td>34</td>
<td>Health information sources</td>
<td></td>
</tr>
<tr>
<td>Justifiability</td>
<td>213</td>
<td>Treatment methods and medication guidances</td>
<td></td>
</tr>
<tr>
<td>Transparency</td>
<td>233</td>
<td>Website contact channels</td>
<td></td>
</tr>
<tr>
<td>Financial disclosure</td>
<td>233</td>
<td>Site owner</td>
<td></td>
</tr>
<tr>
<td>Advertising policy</td>
<td>1381</td>
<td>Advertisements</td>
<td></td>
</tr>
</tbody>
</table>
evaluations of either justifiability or transparency and the FC to corresponding page components are not significant. The strength of the correlations gradually weakens in the order of advertising policy (0.873), authority (0.790), complementarity (−0.713), privacy (0.636), attribution (−0.470), and financial disclosure (0.447). The same result can be found on the metric of the TFD.

The results showed that more the fixation count and longer the fixation duration participants spent on the qualifications of doctors, users’ information, site owner, and advertisements, the higher they scored on the values of authority, privacy, financial disclosure, and advertising policy. However, the more attention they paid on online consultations and health information sources, the lower they scored on the values of complementarity and attribution. The correlation is particularly strong on the advertising policy among all the page components. Participants’ evaluations of transparency and justifiability are not affected by their attention to website contact channels, treatment methods, and medication guidances.

### 4.2 Linear Relationship between Eye movement and Evaluation of Website Quality

Is there a linear relationship between either participants’ FC or TFD to page components and the total score of quality evaluation? If so, which page components and how does the attention allocation to them influence the quality evaluation of health website? We answered these questions by multiple linear regression; the regression results between either FC or TFD to page components and quality evaluation are shown in Tables 4 and 5. The normal distribution of residual has passed the test and no multicollinearity was found in the regression model (all tolerance value > 0.2, VIF < 5). We also considered effects of several control variables like gender, age, education, health website visiting frequency, and health information needs.

The results presented in Tables 4 and 5 show that no significant difference exists among individuals with respect to gender, age, education, browsing frequency, and demand intensity of health information in terms of quality evaluation. When individual characteristic variables were controlled, only attention toward site owner and health information sources has significant impacts on the quality evaluation. Participants’ FC and TFD to health information sources are negatively related to their quality evaluation (−0.346\*, −0.352\*), while their FC and TFD to site owner (0.389\*, 0.376\*) have significant positive impacts. Compared with health information sources, the attention to site owner has a slightly greater influence on the evaluation. The findings show that the more attention participants allocate to the health information sources, the lower their quality evaluation; the more carefully participants observe the site owner, the better they feel about the website quality.

### 5 Discussion

#### 5.1 Relationship between Attention and Evaluation of Principles

The evaluations of authority, privacy, financial disclosure, and advertising policy were found to be positively correlated with the attention to the corresponding page components (i.e., qualifications of doctors, users’ information, site owner, and advertisements). This finding is consistent with Huebner et al. (2013), who reported that the qualification of the authors, privacy policy, and disclosure of funding are key issues for online health information; cancer patients trust the authority, privacy, and financial disclosure when they focus on these features on web pages. However, our findings about privacy and advertising policy do not support the view of Al-Jefri, Evans, Uchyigit, and Ghézzi (2018), who found that privacy and advertising policies subsumed to a broad “ethical” dimension were ranked to be of low importance by the participants.

The evaluations of complementarity and attribution were found to be negatively correlated with the attention
to the online consultations and health information sources. The result is somewhat counterintuitive and little evidence can be found in the previous research, so we try to explain it in practice. By replay and observing participants’ browsing behavior, we found that if users wanted to communicate with doctors, they must click on the online consultation windows, and then the current page jumps to a new page, followed by a message which reminds that “the consultation results were not used as any diagnostic basis.” Therefore, we guess that if users pay much attention to the online consultations, they may have experienced this process and seen the message which influenced their judgment on the complementarity of the website. As for attribution, the sources of most health information including the causes, symptoms, and treatments of diseases are absent on the website. Users may spend a long time looking for the sources among the masses of information, and when they failed, lower evaluation of attribution arose.

This study did not detect any evidence for significant correlations between users’ evaluation of either transparency or justifiability and attention to the website contact channels, treatment methods, and medication guidances. In accordance with the present results, previous studies have demonstrated that it really needs solid medical knowledge to judge whether the introduction of drugs and treatments is reasonable, which is difficult for non-professionals to distinguish (Suárez-Perdomo, Byrne, & Rodrigo, 2018; Rahmatizadeh & Valizadeh-Haghi, 2018). Therefore, users can only rely on their experience to judge the justifiability of health information, which has nothing to do with the attention allocation. As for transparency, a possible explanation for this may be that only one of the users has taken notice of the website contact details; others’ evaluation basis for the transparency principle cannot be verified.

Table 4
Regression Results of FC to Page Components and Quality Evaluation

<table>
<thead>
<tr>
<th>Control variable</th>
<th>Beta</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>−0.201</td>
<td>0.192</td>
</tr>
<tr>
<td>Age</td>
<td>0.160</td>
<td>0.274</td>
</tr>
<tr>
<td>Education</td>
<td>−0.213</td>
<td>0.218</td>
</tr>
<tr>
<td>Health website visiting frequency</td>
<td>−0.088</td>
<td>0.645</td>
</tr>
<tr>
<td>Health information needs</td>
<td>0.261</td>
<td>0.163</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Independent variables (FC)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Qualifications of doctors</td>
<td>0.123</td>
<td>0.525</td>
</tr>
<tr>
<td>Online consultations</td>
<td>−0.194</td>
<td>0.451</td>
</tr>
<tr>
<td>Users’ information</td>
<td>0.188</td>
<td>0.299</td>
</tr>
<tr>
<td>Health information sources</td>
<td>−0.346</td>
<td>0.016</td>
</tr>
<tr>
<td>Treatment methods and medication guidances</td>
<td>0.007</td>
<td>0.967</td>
</tr>
<tr>
<td>Website contact channels</td>
<td>0.049</td>
<td>0.731</td>
</tr>
<tr>
<td>Site owner</td>
<td>0.389*</td>
<td>0.019</td>
</tr>
<tr>
<td>Advertisements</td>
<td>0.387</td>
<td>0.149</td>
</tr>
<tr>
<td>ANOVA</td>
<td>0.003</td>
<td></td>
</tr>
<tr>
<td>R² adj</td>
<td>0.585</td>
<td></td>
</tr>
</tbody>
</table>

Table 5
Regression Results of TFD to Page Components and Quality Evaluation

<table>
<thead>
<tr>
<th>Control variable</th>
<th>Beta</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>−0.094</td>
<td>0.592</td>
</tr>
<tr>
<td>Age</td>
<td>0.229</td>
<td>0.128</td>
</tr>
<tr>
<td>Education</td>
<td>−0.220</td>
<td>0.209</td>
</tr>
<tr>
<td>Health website visiting frequency</td>
<td>−0.176</td>
<td>0.403</td>
</tr>
<tr>
<td>Health information needs</td>
<td>0.309</td>
<td>0.109</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Independent variables (TFD)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Qualifications of doctors</td>
<td>0.223</td>
<td>0.359</td>
</tr>
<tr>
<td>Online consultations</td>
<td>−0.098</td>
<td>0.651</td>
</tr>
<tr>
<td>Users’ information</td>
<td>0.065</td>
<td>0.789</td>
</tr>
<tr>
<td>Health information sources</td>
<td>−0.352</td>
<td>0.024</td>
</tr>
<tr>
<td>Treatment methods and medication guidances</td>
<td>−0.074</td>
<td>0.622</td>
</tr>
<tr>
<td>Website contact channels</td>
<td>0.076</td>
<td>0.591</td>
</tr>
<tr>
<td>Site owner</td>
<td>0.376*</td>
<td>0.042</td>
</tr>
<tr>
<td>Advertisements</td>
<td>0.344</td>
<td>0.145</td>
</tr>
<tr>
<td>ANOVA</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td>R² adj</td>
<td>0.544</td>
<td></td>
</tr>
</tbody>
</table>
5.2 Relationship between Attention and the Whole Quality Evaluation

Users’ attention to information sources and site owner has impacts on the quality evaluation of a health website. The attention to the health information sources was found to be negatively related to the quality evaluation, though the attention to the site owner is positively related to the evaluation. Namely, users are likely to believe that the health information is incomplete when they pay attention to the sources of information. Since most web pages do not provide accurate or adequate origins of the health information, it affects user’s evaluation. This finding is consistent with Bates, Romina, and Ahmed (2007), who regard completeness as an important indicator for users’ perception of health information quality, and the information source is an essential element for complete health information. Compared with other content features, the presentation of site owner information is simpler and clearer, but it is not easy to attract users because of its location at the bottom of the page and the small font size. Therefore, users need to observe the information carefully to realize the information, which lead to high-quality evaluation. This is also correspondent with our earlier observations, which showed that users’ attention to the financial disclosure information of orthopedic and trauma surgical website helped to improve the quality evaluation, even if half of the websites were unwilling to disclose it (Rechenberg, Josten, Grüner, & Klima, 2013).

6 Implications

We focused on the relationship between users’ attention allocation to webpage components and quality evaluation of a health website in this study. The importance of page components for the quality evaluation was emphasized, and a novel research method was verified embodied in the combination of eye tracking and questionnaire.

Our research results provide some suggestions for improving user satisfaction to the website. As the results show in correlation analysis, the evaluations of authority, privacy, financial disclosure, and advertising policy are positively correlated with the attention to the qualifications of doctors, users’ information, site owner, and advertisements. In order to improve the quality of these principles, the health website designer can highlight the text by changing the font size and color on these page components and enlarge the component areas to attract the attention. On the contrary, evaluations of complementarity and attribution are negatively correlated with the attention to the online consultations and health information sources. Therefore, the model of the online consultation windows and the content of health information should be simplified, while these areas should be appropriately reduced to decrease users’ attention allocation.

Especially, the regression results show that by observing the site owner carefully, users have good-quality evaluation, while attention to the information sources causes an opposite effect. In addition to the improvement measures mentioned earlier, we believe that the sources of health information should be presented, and the information of site owners should be enlarged and displayed at the top of the page instead of at the bottom. Making decisions based on low-quality health information may lead to harmful consequences; nevertheless, evaluating the quality of health website has been a major challenge for online health users. In addition to the improvement of interface design, our findings may provide a new thought to evaluate the health website quality automatically based on users’ attention to site owners and information sources.

7 Conclusion and Limitations

In this paper, we revealed the relationship between users’ gaze behavior on health website and their evaluation of health website quality by analyzing the eye movement and questionnaire data. One of the major findings was that some correlations between evaluation of quality principles and attention to corresponding page components are positive (e.g., evaluation of authority and attention to qualifications of doctors), some correlations are negative (e.g., evaluation of complementarity and attention to online consultations), while others are not significant (e.g., evaluation of transparency and attention to website contact channels). Another finding that emerged from this study was that users’ attention to the health information sources is negatively related to their quality evaluation, while the attention to the site owner is positively related to the evaluation. Our research verified a novel method for understanding the process of users’ quality evaluation of health website; the findings are helpful to improve the design of the health website, and provided a new idea for developing automatic evaluation methods of health website quality from the perspective of attention allocation.
As we interpret the findings of our study, its limitations must be acknowledged. We conduct our study on a single health website, and this may restrict the generalizability of our findings. Besides, users’ internal mental processes during the experiment are uncertain, and we are unable to know their inner thoughts when evaluating the health website quality. Further research might involve more health websites as well as combining the think-aloud method.

Acknowledgments: The work was partly supported by the National Social Science Fund of China (Grant No. 20BTQ076) and the Social Science Fund of Jiangsu Province (Grant No. 20TQA003).

References


Exploring Public Response to COVID-19 on Weibo with LDA Topic Modeling and Sentiment Analysis

Abstract: It is necessary and important to understand public responses to crises, including disease outbreaks. Traditionally, surveys have played an essential role in collecting public opinion, while nowadays, with the increasing popularity of social media, mining social media data serves as another popular tool in opinion mining research. To understand the public response to COVID-19 on Weibo, this research collects 719,570 Weibo posts through a web crawler and analyzes the data with text mining techniques, including Latent Dirichlet Allocation (LDA) topic modeling and sentiment analysis. It is found that, in response to the COVID-19 outbreak, people learn about COVID-19, show their support for frontline warriors, encourage each other spiritually, and, in terms of taking preventive measures, express concerns about economic and life restoration, and so on. Analysis of sentiments and semantic networks further reveals that country media, as well as influential individuals and “self-media,” together contribute to the information spread of positive sentiment.

Keywords: COVID-19, Weibo, web crawling, LDA, sentiment analysis

1 Introduction

In recent decades, we have encountered several disease outbreaks such as SARS in 200 and MERS in 2012. Nowadays, we are facing another enemy: COVID-19. As of July 26, 2020, more than 16 million COVID-19 confirmed cases have been reported around the world (JHU COVID-19 Resource Center, 2020). With the first confirmed case officially reported in Wuhan, China, in late December 2019, the global COVID-19 outbreak has brought great damage to the world’s normal operations for over half a year and has been, unfortunately, aggravating, as of July 2020.

During and after the disease outbreak, public opinion is commonly collected as it is useful in many ways such as improving communications in terms of public concerns, crisis management, health knowledge promotion, and so on between governments and the public (Holmes, Henrich, Hancock, & Lestou, 2009; Mollema et al., 2015). Traditionally, surveys have played an important role in gathering public opinion, and undoubtedly, they have also been applied to disease outbreak research about public opinion. Table 1 summarizes some popularly investigated themes in survey-based research of the COVID-19 outbreak.

While surveys play an irreplaceable role, there is another way of collecting public opinion about disease outbreaks: the increasing popularity of social media and the development of text analysis techniques have given rise to mining social media data. Unlike surveys that can mainly collect only a limited volume of data and may not be able to reflect the real thoughts of the public because of some influential factors during the survey process, such as the environmental distraction, subject’s psychological pressure, and so on (Hridoy, Ekram, Islam, Ahmed, & Rahman, 2015), social media data, which is the collection of short texts posted online to express one’s feelings at any time, are generated on a large scale. Thus, the efficient use of social media data can contribute largely to the research about public opinion.

Reported as the first storm center affected by COVID-19, China had undergone an incredibly hard time at the very beginning, encountering difficulties such as city lockdowns, lack of medical supplies, lack of hospital resources, and so on. But luckily, by July 2020, China had also made satisfactory achievement by getting the disease outbreak under control with aggressive approaches (Campbell, 2020; Clinch, 2020). Weibo, the biggest social
media platform in mainland China, serves the functions of information sharing and communications in the country. Data collected from Weibo can be analyzed to understand the Chinese people’s reaction to the disease outbreak and the characteristics of semantic networks, which lead to the research questions (RQs) of this research.

RQ1: What topics can be detected from Weibo posts on COVID-19?

RQ2: How do sentiments change over time, and what are the characteristics of different semantic networks?

The rest of this article is organized as follows. The second part reviews some related work in the context of this research, after which methodologies used to conduct this research are introduced, and the results are presented and discussed. Finally, conclusions are made, together with the contribution of this work to the research context and suggestions on further research directions.

2 Related Work

2.1 Disease Outbreaks on Social Media

Table 2 records some exemplary research on the disease outbreak on social media. It can be observed that social media data contribute to several types of research on disease outbreaks, including public opinion mining, sentiment analysis, semantic network analysis, disease outbreak detection, and so on; the methods applied to social media data analysis vary from manual coding to machine learning techniques; Twitter is a worldwide platform for collection and analysis of social media data, while for social media data collection from residents in mainland China, Weibo is a more popular choice.

2.2 Topic Modeling Using LDA

Latent Dirichlet Allocation (LDA) assumes that a document is generated based on a certain number of topics, and each word in the document is randomly selected from its corresponding topic vocabulary (Blei, Ng, & Jordan, 2003; Gruber, Weiss, & Rosen-Zvi, 2009). It is an excellent probabilistic model that performs well in topic modeling and has been widely applied in research (Hu et al., 2017). For example, Barua, Thomas, and Hassan (2012) utilized LDA to discover topics and topic trends from a popular Q&A website in the programming field and found that the developer community discussed a wide range of topics and discussions of different topics are interconnected. Similarly, Hu et al. (2017) studied email corpora with the LDA model.

LDA has been applied to extract topics from various kinds of corpora, including but not limited to microblogs. For example, Huang, Yang, Mahmood, and Wang (2012) applied LDA with web usage data; Xu, Zhang, and Yi (2018) and Lim and Buntine (2014) studied tweets with LDA modeling. Therefore, it is logical to follow that exploring topics from Weibo datasets with LDA topic modeling should also harvest satisfactory results.

2.3 Sentiment Analysis

Sentiment of text data mainly refers to the emotions hidden within the text. Sentiment analysis has been widely applied to a large volume of opinion mining research, including product review analysis, public response to the stock market, and so on, as valuable information can be discovered if emotions in texts are well analyzed (Bakshi, Kaur, Kaur, & Kaur, 2016). In general, sentiments are classified into three categories: positive, neutral, and
negative, and there are mainly two ways of sentiment tagging (Ray & Chakrabarti, 2017).

Lexicon-based sentiment tagging analyzes the words in a sentence and harvests the overall score by adding up the scores of each word, for which sentiment dictionaries are used (Bhonde, Bhagwat, Ingulkar, & Pande, 2015). This method was widely applied in the sentiment tagging of social media posts. For example, Ray and Chakrabarti (2017) used the R language to tag Twitter posts for product reviews with lexicon vocabularies and completed sentiment analysis at three levels: document, sentence, and aspect; Pérez-Pérez, Pérez-Rodríguez, Fdez-Riverola, and Lourenço (2019) used a lexicon-based tagger to analyze tweets’ sentiments under each topic discovered in the Human Bowel Disease community.

Machine learning approaches to sentiment tagging are also gaining popularity. For example, Chen and Sokolova (2018) adopted an unsupervised approach to cluster sentiments of clinical discharge summaries with word embeddings generated from Word2Vec and Doc2Vec models and compared the results. Salathé and Khandelwal (2011) compared three machine learning techniques in terms of sentiment classification performance and adopted both naive Bayes and maximum entropy algorithms for the supervised classification of vaccination sentiments.

### 3 Methodologies

#### 3.1 Methodological Framework

Figure 1 shows the methodological framework of this work. This research starts with data collection of Weibo posts with a user-simulation-like web crawler. Posts are then collected and processed to remove text noises and stop words, after which text segmentation is also performed. Then topic modeling is conducted using the LDA model with the cleaned dataset. To answer the second research question, lexicon-based sentiment analysis is conducted on data, including the number of posts, network information, and so on, to reveal sentiment trends and discover the characteristics of semantic networks.

#### 3.2 Data Collection

This work examines the public opinion related to COVID-19 on Weibo, for which theme-related posts must be collected. Both web crawlers and APIs are nowadays widely used technologies to collect data from social media, including Weibo (Liu & Hu, 2019; Liu, Wu, Wang, & Li, 2014). Although Weibo API, being the official gateway...
to collect Weibo data, is easy to use, there are many constraints such as only providing a limited number of posts and so on (Zeng, Zheng, Chen, & Yu, 2014). To avoid such inconvenience, this work uses selenium to develop a simulation-based web crawler using Python, so as to satisfy the data collection task. The web crawler simulates human logins and searches Weibo posts based on given keywords. The HTML of webpages is then collected and parsed to get the posted Weibo content and relevant information such as username and so on.

To fulfill the research purposes, this work uses six keywords selected from Hu, Huang, Chen, and Mao (2020), as listed in Table 3. Hot posts on each day over the period from January 1 to June 30, 2020, are collected.

### 3.3 Data Preprocessing

#### 3.3.1 Removing Noises

Text noises are generally removed in the text mining research, which benefits the experimental outcomes (Celardo, Iezzi, & Vichi, 2016). In the context of this research, text noises include emoji codes, punctuation marks, symbols, non-Chinese words, and so on. To remove text noises, the “re” regular expression package in Python is adopted to remove all non-Chinese components in the dataset.

#### 3.3.2 Stop Word List

In Chinese text, there are many “meaningless” words like “我” (I/me), “你” (you), “了” (have done something), and so on, which are normally removed in information retrieval and text mining tasks so as to improve the experimental outcomes (Zou, Wang, Deng, & Han, 2006). There are also some Chinese stop word lists built by university NLP labs and companies. To construct a stop word list, this work integrates three public stop word lists (Baidu stop word, SCU stop word, and HIT stop word) that are widely used (Xie et al., 2019), together with some domain stop words such as “转发微博” (repost) that appear frequently in most of the posts collected.

#### 3.3.3 Chinese Text Segmentation

Unlike English text, the Chinese text needs to be segmented for analysis tasks. In terms of Chinese text segmentation tools, the “jieba” package in Python is widely used and has many advantages, such as adding customized words (Day & Lee, 2016; Peng, Liou, Chang, & Lee, 2015). In this research, the “jieba” package is adopted to perform the Chinese text segmentation task.

### 3.4 Topic Modeling

As mentioned above, LDA is a powerful tool in topic modeling. LDA can extract a given number of topics from a corpus that contains a certain number of documents. This research applies LDA to extract a certain number of topics from the cleaned dataset with the Python package “gensim.” In terms of the determination of the number of topics, both perplexity and coherence scores are taken into consideration. While the former measures how well the model is generated from the corpus (the lower the better), the latter measures the sentence similarity of each topic in the dataset (the higher the better) (Blei et al., 2003; Xie, Qin, & Zhu, 2018). After the optimal model is determined,
The preliminary findings can be further confirmed with public opinions about disease outbreak (Jalloh et al., 2017; which are in line with some previous research findings on global pandemic, hospital staff, vaccination, and so on, might cover a wide range of topics, including local disease vocabulary and (2) discussions of COVID-19 on Weibo posts, blanks, and “NA” that come probably from parsing observed: (1) undoubtedly, term frequencies of each word (including negation words that might reverse the punctuation marks; (2) each sentence is then segmented calculating the sentiment score of a post mainly contains sentiment tagging, descriptive statistics of the results (3) the overall score of the post is calculated based on the scores of each sentence, which are calculated from the scores of each word.

3.5.2 Statistics and Semantic Network Analysis

After sentiment tagging, descriptive statistics of the results are described. Time series analysis is then performed to discover interesting phenomena from the number of positive and negative posts over time. To take a closer look at the sentiments of public opinion, positive and negative semantic networks are constructed to identify important roles in and the characteristics of respective networks, for which network visualization and statistical analysis are performed.

4 Results and Discussion

4.1 Data Collected

Figure 2 is a screenshot of some collected raw data. In total, 719,570 posts are collected over the period from January 1 to June 30, 2020, based on the given keywords. After data cleaning is performed, which includes removing duplicate posts, blanks, and “NA” that come probably from parsing errors, only 374,225 posts remain.

The dataset is then further processed with Chinese text segmentation and removing stop words, after which it is ready for text mining analysis. Figure 3 visualizes the top 500 frequent terms with a word cloud, and Table 4 records the top 50 frequent words in the dataset, from which some interesting preliminary findings can be observed: (1) undoubtedly, term frequencies of each keyword selected for data collection rank top in the vocabulary and (2) discussions of COVID-19 on Weibo might cover a wide range of topics, including local disease outbreak, international relations, prevention measures, global pandemic, hospital staff, vaccination, and so on, which are in line with some previous research findings on public opinions about disease outbreak (Jalloh et al., 2017; Nickell et al., 2004; Rubin, Amlôt, Page, & Wessely, 2009). The preliminary findings can be further confirmed with the following analysis.
4.2 Topic Modeling

4.2.1 Perplexity and Coherence Scores of LDA Models

Figure 4 and 5, respectively, show the perplexity and coherence scores of models trained under different settings, namely, different numbers of topics. Overall, with an increasing number of topics set for model training, a smaller perplexity score is gained, that is, the model performs better in predicting the samples. Notably, the setting of a larger number of topics may lead to model overfitting issues, and thus, it needs to be careful in deciding the number of topics for training the model. Another measurement for deciding the number of topics is to compare the coherence scores of different models. In this case, coherence scores fluctuate with an increasing number of topics, and peaks are observed when numbers of topics for model training are set as 4, 8, 12, 15, respectively. Based on the perplexity and coherence scores, first, it can be concluded that the number of topics set for LDA training should be 12 or 15. By observing the keywords generated for each topic, the number of topics is finally determined as 12, because it is easier to code the topics from the given keywords.
4.2.2 Discovered Topics

Figure 6 is a visualization of the selected LDA model mentioned above. While the left panel shows the distribution of each topic, the word list on the right panel shows the top 30 most salient terms of the selected topic, in which the blue bar shows the overall term frequency in the dataset, and the red bar represents the estimated term frequency in the selected topic. Table 5 records the discovered topics, each with 10 representative words selected from the top 30 most salient terms.

It can be seen from Figure 6 that though there are some overlapped areas, in general, topics extracted are evenly distributed. From Table 5, it can be inferred that users on Weibo discuss a wide range of topics, among which 10 topics are related to the COVID-19 outbreak while 2 topics on “Law” and “People,” seem irrelevant to the research context. Seven topics, including “Fight the virus together,” “Knowledge,” “Assistance,” “Prevention,” “Treatment,” “Global pandemic,” and “Stay at home” are directly related to the disease outbreak, while three topics, “Economics,” “Study,” and “Celebrity and charity” could be regarded as topics derived from COVID-19, because they are either fields affected by the disease outbreak or tightly associated with it. From the results, it is interesting to know that people encourage each other in terms of fighting the disease, share disease-related knowledge such as prevention and transmission, pay attention to global disease outbreak development, and also discuss the affected life together, which are consistent with some findings of previous research (Chung, He, & Zeng, 2015; Corley, Cook, Mikler, & Singh, 2010; Lazard, Scheinfeld, Bernhardt, Wilcox, & Suran, 2015; Signorini, 2014).

4.3 Sentiment Tagging Results and Trend Analysis

After data preprocessing and text segmentation, 207,323 posts in total remained for sentiment tagging, with 79,861 positive posts and 33,049 negative posts, while the rest were all tagged as neutral posts (sentiment score is 0). Figure 7 shows the trends of the number of positive and negative posts over the data collection period. It can be observed that the number of positive posts exceed that of the negative ones over the whole period.
Figure 6. LDA model visualization

Table 5
Top 30 Most Salient Terms of Each Topic and Topic Coding Results

<table>
<thead>
<tr>
<th>Topic ID</th>
<th>Topic Label</th>
<th>10 representative words selected from the top 30 most salient words</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Fight the virus together</td>
<td>力量 (power), 加油 (add oil), 人民 (folk), 希望 (hope), 中国 (China), 抗击 (fight), 战疫 (fight the virus), 一线 (front line), 成功 (success), 努力 (work hard)</td>
</tr>
<tr>
<td>2</td>
<td>Knowledge</td>
<td>研究 (research), 专家 (expert), 原因 (reason), 科普 (popular science), 传播 (transmission), 疾病 (disease), 科学 (science), 预防 (prevention), 考试 (exam), 发现 (discover)</td>
</tr>
<tr>
<td>3</td>
<td>Assistance</td>
<td>工作 (work), 奋战 (fight), 归来 (come back), 全国 (nation), 万众一心 (united), 现场 (on site), 关注 (pay attention to), 驰援 (support), 新闻 (news), 情况 (situation)</td>
</tr>
<tr>
<td>4</td>
<td>Economics</td>
<td>全球 (globe), 经济 (economics), 影响 (influence), 基金会 (fund), 世界 (world), 国际 (internationality), 控制 (control), 合作 (cooperation), 社会 (society), 市场 (market)</td>
</tr>
<tr>
<td>5</td>
<td>Global pandemic</td>
<td>伊朗 (Iran), 病例 (case), 英国 (UK), 疫情 (disease outbreak), 新增 (new case), 日本 (Japan), 意大利 (Italy), 累计 (accumulate), 告急 (urgent), 境外 (abroad)</td>
</tr>
<tr>
<td>6</td>
<td>Prevention</td>
<td>疫情 (disease outbreak), 口罩 (face mask), 宣传 (promotion), 防护 (prevention), 做好 (do well in), 防控 (prevent), 防疫 (fight the virus), 风险 (risk), 健康 (health), 措施 (measure)</td>
</tr>
<tr>
<td>7</td>
<td>Treatment</td>
<td>隔离 (quarantine), 出院 (discharged), 无症状 (no symptoms), 医学观察 (medical observation), 密切接触 (close contact), 治愈 (cure), 发热 (fever), 重症 (severely ill), 确诊 (confirmed affection), 治疗 (treatment)</td>
</tr>
<tr>
<td>8</td>
<td>Stay at home</td>
<td>期间 (period), 生活 (life), 这次 (this time), 开心 (happy), 回家 (go back home), 在家 (at home), 喜欢 (like), 事情 (thing), 朋友 (friend), 家里 (home)</td>
</tr>
<tr>
<td>9</td>
<td>Law</td>
<td>法院 (court), 法官 (judge), 人民 (folk), 违法 (breach the law), 案件 (case), 被告 (defendant), 证据 (proof), 录音 (audio recording), 依法 (according to the law), 真相 (truth)</td>
</tr>
<tr>
<td>10</td>
<td>Study</td>
<td>学校 (school), 孩子 (children), 开学 (start school), 学生 (student), 入学 (start school), 大学 (university), 专业 (major), 家长 (parents), 作业 (homework), 高考 (college entrance examination)</td>
</tr>
<tr>
<td>11</td>
<td>Celebrity and charity</td>
<td>蔡徐坤 (Xukun Cai), 杨紫 (Zi Yang), 公益 (charity), 超话 (super topic), 粉丝 (fan), 微光 (dawn), 慈善 (charity), 肖战 (Zhan Xiao), 守护 (protect), 打败 (defeat)</td>
</tr>
<tr>
<td>12</td>
<td>People</td>
<td>直播 (live stream), 儿子 (son), 弟弟 (younger brother), 奶奶 (grandmother), 放假 (vocation), 老公 (husband), 爱豆 (idol), 价值 (value), 打榜 (boost popularity), 遇见 (meet)</td>
</tr>
</tbody>
</table>
Specifically, in terms of positive posts, three peaks can be observed, as also marked in Figure 7. The top 10 most frequent terms extracted from posts at each peak are listed in Table 6. The first peak is recorded on January 24, the day of the Chinese New Year’s Eve, and right after the lockdown of Wuhan, which took place on January 23. From the top 10 terms, it can be inferred that the posts are mostly related to wishes for the coming New Year as well as the COVID-19 outbreak in Wuhan. The second peak comes on March 3, when the central government announced the preliminary success in fighting COVID-19, for which relevant words, for example, 保护 (protect), can also be observed. The third peak comes on May 23, when the success of the phase 1 vaccine trial was announced.

Table 6
Top 10 Frequent Terms Extracted from Posts at Each Peak

<table>
<thead>
<tr>
<th>No.</th>
<th>Peak 1</th>
<th>Peak 2</th>
<th>Peak 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Top term</td>
<td>Translation</td>
<td>Freq.</td>
</tr>
<tr>
<td>1</td>
<td>希望</td>
<td>Hope</td>
<td>470</td>
</tr>
<tr>
<td>2</td>
<td>武汉</td>
<td>Wuhan</td>
<td>388</td>
</tr>
<tr>
<td>3</td>
<td>疫情</td>
<td>Disease outbreak</td>
<td>320</td>
</tr>
<tr>
<td>4</td>
<td>肺炎</td>
<td>Pneumonia</td>
<td>260</td>
</tr>
<tr>
<td>5</td>
<td>加油</td>
<td>Add oil</td>
<td>235</td>
</tr>
<tr>
<td>6</td>
<td>新型</td>
<td>Novel</td>
<td>204</td>
</tr>
<tr>
<td>7</td>
<td>平安</td>
<td>Safe and sound</td>
<td>179</td>
</tr>
<tr>
<td>8</td>
<td>冠状病毒</td>
<td>Coronavirus</td>
<td>173</td>
</tr>
<tr>
<td>9</td>
<td>病毒</td>
<td>Virus</td>
<td>157</td>
</tr>
<tr>
<td>10</td>
<td>一年</td>
<td>1 year</td>
<td>154</td>
</tr>
</tbody>
</table>
4.4 Semantic Network Analysis

Figures 8 and 9 record the visualization of semantic networks of the positive and negative sentiments, respectively. In the network visualization, the color of each modularity is different from one another, and the size of a node stands for its interaction frequency: the more interaction it has, the bigger size of node it is. In both semantic networks, it could be easily observed that country media, including 人民日報 (People’s Daily), 央視新聞 (CCTV News), 环球时报 (Global Times), and so on are leading the discussions, and basically, each of them forms a relatively independent community. As for the differences, it could be seen that there are more medium-sized nodes around each leading actor in the semantic network of positive sentiment, which means that mainstream media and influential KOL (key opinion leader), including entrepreneurs such as 乡村教师代言人-马云 (Jack Ma), 胡锡进 (Hu Xijin), self-media such as 英国那些事 (Things in the UK) and so on, play an important role in leading the information spread and discussions of positive sentiment, while in the semantic network of negative sentiment, the discussions, also led by country media, are more scattered.

To take a closer look, the reposting frequencies of each node are calculated for each semantic network and then normalized between 0 and 1 for comparison purposes. Quartiles of normalized data are shown in Figure 8.
Table 7, from which it can be seen that Q1, median, and Q3 of the reposting frequencies of the positive semantic network are all greater than those of the negative semantic network, meaning that there are more influential nodes in the semantic network of positive sentiment, which are consistent with the previous findings.

Table 7
Quartiles of Normalized Reposting Frequencies of Each Semantic Network

<table>
<thead>
<tr>
<th>Sentiment</th>
<th>Min</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>0.000</td>
<td>0.002</td>
<td>0.013</td>
<td>0.109</td>
<td>1.000</td>
</tr>
<tr>
<td>Negative</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.002</td>
<td>1.000</td>
</tr>
</tbody>
</table>

5 Conclusion and Future Work

Weibo serves as a social media platform for people in mainland China to share information and communicate with each other. With the help of 719,570 collected posts and application of the LDA model, a wide range of topics discussed in relation to COVID-19 on Weibo is discovered. In response to the COVID-19 outbreak, people gain knowledge about COVID-19, show their support for frontline warriors, encourage each other spiritually, and, in terms of taking preventive measures, express concerns about economic and life restoration, and so on. Sentiment analysis further reveals that country media are leading the discussions on Weibo in both semantic networks, while, specifically, country media, as well as influential
individuals and “self-media” together contribute to the information spread of positive sentiment, indicating that the government could better fulfill its role as crisis communicator through the utilization of such kind of media network.

Although there have been studies of public opinion on COVID-19 using surveys, scant studies focus on COVID-19 opinion mining based on social media data. With LDA’s excellent performance in topics modeling and sentiment analysis to take a closer look at people’s feelings, this work contributes to the understanding of peoples’ response to COVID-19 on Weibo and may probably serve as an example of preliminary research in the application of LDA and sentiment analysis on the COVID-19 social media dataset. Further investigation of this topic can be done in different ways. One direct way is to extend the research context, such as tracing relevant posts of each topic, analyzing peaks of negative posts, revealing the relationships between positive and negative trends, and so on. Besides, identification of topic trends, correlation analysis between the number of posts and disease development trends, and so on can also lead to meaningful findings, for which a similar approach can be seen in some previous research (Fung et al., 2013; Hu et al., 2017).

References


Barua, A., Thomas, S. W., & Hassan, A. E. (2014). What are developers so on can also lead to meaningful findings, for which a similar approach can be seen in some previous research (Fung et al., 2013; Hu et al., 2017).

References


Cross-Language Fake News Detection

Abstract: With increasing globalization, news from different countries, and even in different languages, has become readily available and has become a way for many people to learn about other cultures. As people around the world become more reliant on social media, the impact of fake news on public society also increases. However, most of the fake news detection research focuses only on English. In this work, we compared the difference between textual features of different languages (Chinese and English) and their effect on detecting fake news. We also explored the cross-language transmissibility of fake news detection models. We found that Chinese textual features in fake news are more complex compared with English textual features. Our results also illustrated that the bidirectional encoder representations from transformers (BERT) model outperformed other algorithms for within-language data sets. As for detection in cross-language data sets, our findings demonstrated that fake news monitoring across languages is potentially feasible, while models trained with data from a more inclusive language would perform better in cross-language detection.

Keywords: fake news detection, cross-language study, information detection

1 Introduction

Automatic fake news detection has become increasingly important with the rapid development of social media. Individuals may define the word “fake” differently, so each paper should clarify their definitions (Ruchansky, Seo, & Liu, 2017). In this study, fake news simply refers to fabricated news that is written to mislead readers or attract people’s attention, thus gaining political or financial benefits (Baym, 2005; Kershner, 2011; Mecacci, 2016). Cross-language research is generally considered to be the study on the impact of two or more languages to solve the problem regarding social culture or human behavior (Alghowinem, Goecke, Epps, Wagner, & Cohn, 2016; Hu & Yang, 2017).

Generally, fake news is considered one of the greatest threats to society with the rapid development of the Internet recently, since it has negative effects on democracy, journalism, and economies (Tacchini, Ballarin, Della Vedova, Moret, & de Alfaro, 2017). For instance, during the 2016 US presidential election, there was an outbreak of fake news pandemic. The impact of fake news is irreversible (Grinberg, Joseph, Friedland, Swire-Thompson, & Lazer, 2019). An example was the smear on Hillary’s health, which spread rapidly in the last 3 months before the election, causing her approval rating to drop rapidly. Although some authoritative newspapers had stated the fact, most people still doubt Hillary’s health issues. Consistent research has been carried out to address these issues. Adopting different strategies for fake news detection is one of the most fundamental research directions.

To date, most studies focus on detecting fake news in a specific language, where English is the most commonly studied language (Pérez-Rosas & Mihalcea, 2014). However, much of the information dissemination can be found not only among native English speakers but also among people from other cultures who speak a different language. It raises an important question about the applicability of existing methods for detecting fake news. To fill in the gaps in previous studies, this study aims to investigate the following two research questions (RQs).

RQ1: Which textual features are more effective in fake news detection?

Extracting features is a crucial step in building detection models. To have a good detection performance, effective textual features need to be extracted. Besides,
it will be interesting and meaningful to compare the top keyword features in two data sets from different languages. The commonality of these words can be used to raise awareness when people see news stories containing words with similar characteristics. Understanding the differences in characteristics may also help us to explain the doubts in the second RQ.

RQ2: Is it possible to transport fake news detection models in cross-language data sets?

Natural language processing (NLP) is highly related to linguistics, as all the current NLP models will convert a plain text into a numerical vector. Although the methods of conversion vary, they are all based on the underlying characteristics of the language (Raskin, 1985). This study examines the effect of linguistic backgrounds on the fake news detection problem by conducting a series of classification experiments in English and Chinese data sets. The result provides a new angle for scholars in the field of fake news detection.

In this paper, we extracted the keyword features to compare the differences between Chinese and English fake news and used the bidirectional encoder representations from transformers (BERT) model for classification. Our results show that cross-language fake news detection is possible, but the accuracy of using the Chinese model to classify English fake news is much higher than the accuracy of utilizing the English model to classify Chinese fake news. Based on the principles of the BERT model, we speculate that this effect might be caused by the fact that the features of Chinese keywords are more complex than the features of English keywords. To validate this idea, we calculated the similarity between Chinese and English keyword features, then repeated the previous experiment after removing the keyword features that were 70% similar to English in the Chinese fake news. A control group was also evaluated to eliminate the impact of reductions in keyword variables. The results showed that it was indeed the complexity of the fake news language that affected the accuracy of the cross-language model’s predictions. To summarize, in cross-language fake news detection, it may be better to use the data set in a more inclusive language (in terms of keyword features), such as Chinese, to be the training data than to use the data in a simpler language.

2 Related Work

2.1 Fake News

In recent years, social media has dramatically changed the manner of producing and disseminating information, but it has also facilitated the spread of false information. This phenomenon has caught the attention of many scholars. Research on fake news mainly focuses on its detection and impact analysis.

First, existing research has multiple definitions of fake news. Much of the contemporary media defines fake news as those rumors and false stories that look like news (Allcott & Gentzkow, 2017). Recent research has pointed out the difference between fake news and deceptions, using unique characteristics to define fake news to separate them from ordinary rumors. In this study, fake news refers to fabricated news with the purpose of misleading people. Compared with deception and rumors, the problem of fake news detection is harder as the political language is often presented in shorter and more formal statements (Wang, 2017). As for the negative effect of fake news, Chaouachi and Rached (2012) have recently observed that fake news could mislead customers’ judgment of the market by influencing their attitudes toward brands. Besides, Brewer, Young, and Morreale (2013) also studied the impact of fake news from the political perspective. They found fake news can steer public opinion in the exact opposite direction to the truth, and this adverse effect is difficult to eliminate even if the truth is clarified. Moreover, Brewer et al. (2013) argued that fake news can also subliminally influence a particular group’s perception of things and its cultural transmission.

In short, existing work has shown the adverse impact of fake news and the importance of solving this problem, which has necessitated a lot of effort to be spent on detecting fake news automatically.

2.2 Detection Models

To date, content-based and reviewer behavior-based models have dominated the academic monitoring of fake news.

The support vector machine (SVM) classification model has been one of the most effective models in previous studies. The support vector regression (SVR) model trained on Burfoot and Baldwin’s satire data set by Horne and Adali (2017) reached 71% accuracy in predicting fake news. Li, Ott, Cardie, and Hovy (2014)
collected deceptive opinions from Amazon Mechanical Turk and proposed a “gold standard” data set to build the SVR model with the best performance of 86% accuracy. It is worth noting that they invited human participants to identify fake news manually, and the results showed that the highest score was only 65%, much lower compared to the accuracy of automatic detection models. The SVR model was also applied in the study by Shojaee, Murad, Azman, Sharef, and Nadali (2013). They extracted 234 stylometric features and showed that SVM outperformed the naive Bayes (NB) model, which is another proof that SVM performs better than other traditional models. Rubin, Chen, and Conroy (2015) conducted experiments on a much larger data set with 12,800 manually labeled fake news items, also using the SVR model, but classified the fake news into five categories. They also pointed out that the latest advances in NLP might be applicable and helpful in fake news detection although he did not prove this idea with valid experiments and evidence. Ahmed, Traore, and Saad (2017) tried to combine NLP approaches with automatic fake news detection, where n-gram modeling was utilized in feature identification and fake news categorization. Moreover, SVM, linear SVM (LSVM), K-nearest neighbor (KNN), decision tree (DT), stochastic gradient descent (SGD), logistic regression (LR) were tested, and the best accuracy was improved to 90% with n-gram features and LSVM.

In general, the current study is only preliminary evidence that new NLP models developed in recent years may be useful for fake news monitoring. The BERT model, proposed by Devlin, Chang, Lee, and Toutanova (2018) in Google, has outperformed many other popular models, such as long short-term memory (LSTM) and convolutional neural network for text (TextCNN), in the text classification tasks. Therefore, to get higher accuracy, we use the BERT model to detect fake news in within- and cross-language data sets.

2.3 Cross-Language Study

Research on cross-language fake news is still at an early stage, as the cultural integration between countries and regions, especially in the context of language exchange, is a very recent phenomenon (Ahmed, et al., 2017). The definition of culture is very broad, and most researchers agree that it is very difficult to study culture exhaustively and comprehensively. At the same time, language is considered to be an important and representative part of a culture, as the contained information is a deposit of culture and history, which is why much research is currently conducted on language (Hu & Yang, 2017; Lewis & George, 2008; Prettenhofer & Stein, 2010; Savage, Merritt, Rzeszutek, & Brown, 2020).

As for the existing studies, most of them focused on deception and rumors that bear some resemblance to fake news. Lewis and George (2008) compared the deceptive behavior of groups of Koreans, Americans, and Spaniards in social media to analyze the relationship between rumors and language background. Furthermore, a study on detection of cross-language deception was presented in another paper (Pérez-Rosas & Mihalcea, 2014), where the authors used crowdsourcing to build deception classifiers, reaching 65% accuracy based on comparative experiments. All the research results suggest that both human lying behavior and outcomes are culture- and language-dependent processes, which implies the possibility of monitoring fake news from a multilanguage perspective. So far, the research on fake news detection was focused on a single language. Few scientific studies have looked at cross-lingual fake news detection, and this is a really promising direction.

3 The Data Sets

3.1 English Fake News Data Set

Developed by Ahmed, Traore, and Saad (2018) for English fake news detection tasks, this open data set consists of 17,903 fake news and 20,826 true news items. For each news item, the titles, contents, subjects, dates, and labels are extracted as attributes, and all news dates are between January and December 2017. The entire text, except the labels, is shown in English, while the labels are presented in the form of numbers “0” and “1”, where “0” stands for real news and “1” represents fake news.

All the news items were collected from real-world sources. Ahmed et al. (2018) collected fake news from articles in Kaggle, as well as from the official website of Politifact, a fact-checking organization in the USA, which aims at providing the public with the statistics on fake information. As for the real news, they were collected from some famous authoritative news websites such as Reuters.com.

3.2 Chinese Fake News Data Set

Newly built by Cao et al. (2019), this Chinese news data set contains 19,285 fake news and 19,186 real news items, and
it is publicly available in the Biendata. Only the ID, text, and labels are included in the data set, and all news dates are between November 2018 and May 2020. The involved fake news was selected from the false-news posts offered by the official Weibo Community Management Center, which is a reputable institution in fake news collection. The real ones are crawled from both Weibo and other famous authoritative news websites.

It is worth mentioning that news topics are not purposefully differentiated in this preliminary study as the textual features of fake news are the focus of this series of experiments. The two data sets are suitable for this study because they are widely used for fake news detection tasks and have been proved to be useful in previous studies (Ahmed et al., 2017; Cao et al., 2019; Shu, Sliva, Wang, Tang, & Liu, 2017). In addition, the sizes of the two data sets and the ratios of real news to fake news are roughly the same, which avoids the data imbalance problem. There are mainly three reasons to select Chinese and English languages in this study: (1) they are the two most prevalent and widely spoken languages worldwide; (2) they can represent differences between Western and Asian cultures; (3) there are sufficient labeled fake news data in these two languages (Ahmed, et al., 2018).

4 Experimental Design

In response to the first RQ, we used TextRank (Mihalcea & Tarau, 2004), a graph-based text ranking model, to extract the keyword features from both the data sets, then undertook summarization and comparison of the keyword features in the fake news from the two different languages. To answer the second RQ, we conducted a series of classification experiments with 10-fold cross-validation.

Several studies trained traditional machine learning models with these two data sets, and preliminary evidence showed that advanced models in the NLP domain have a proven impact on the classification accuracy. This study decided to use the BERT model to classify fake news as the BERT model often outperformed other algorithms in textual classification tasks. The performance was quantified by the classic measures (Accuracy, F1 score, Precision, and Recall) within the range of [0,1]. Of the four approaches, Accuracy and Recall were the two indicators of most interest in this study, as our primary goal was to detect most of the fake news. Unlike English, Chinese words are not separated by spaces and thus segmentation was the first necessary step in Chinese text processing. To solve the problem of semantic disambiguation, an off-the-shelf tool, the package of “Jieba” in Python, was used in this study, due to its superior performance reported in the literature (Sun, 2012). Baidu stopword list, which has been demonstrated to be very effective in text classification (Guan, Deng, & Wang, 2006; Sun, Esho, Liu, & Pang, 2016), was also used in this study.

In the first step, we compared the performances of within- and cross-language data sets. At the same time, we also compared within-language data set classification results of the BERT model with the results from traditional models used in previous studies to see whether the BERT model leads to an improvement. As for the classification of the within-language data set, we divided the data set into three parts for training, validating, and testing. As for the experiments on cross-language data set classification, the original test data were replaced by data from another data set, while the training data remained the same. Like most previous research, to enable the text-based experiments, we translated the training data into the language of the test data using Google Translate application programming interface (API). As the size of the data set might have an impact on the translation performance, excess data were removed randomly to ensure a consistent amount of training and test data in all experiments.

In the second step, Chinese keyword features with >70% similarity to any keywords in English were removed. The similarity was also calculated by the cosine similarity between the BERT-generated word vectors. There is no standardized criterion of similarity threshold, and generally, two targets can be considered similar if they are 50%–100% similar to each other. We chose 70% in our study. Then, the first step was repeated to record the performance after removing specific features. To eliminate the possible effect of the number of keyword features on the results, a control experiment was also conducted, in which the same number of Chinese keyword features were randomly removed.

5 Results and Discussion

5.1 RQ1: Keyword Features Extraction

Table 1 shows the top 100 keyword features extracted from the Chinese fake news data set. We can see that induced redirection and dissemination is one of the characteristics of Chinese fake news. Many seductive words, such as “diffuse”, “help”, “retransmit”, and “please”, rank high in the keyword features list. Furthermore, fake news in Chinese often includes relatively vulnerable groups,
Table 1
Top 100 Keyword Features Extracted from the Chinese Fake News Data Set

<table>
<thead>
<tr>
<th>No.</th>
<th>Features</th>
<th>No.</th>
<th>Features</th>
<th>No.</th>
<th>Features</th>
<th>No.</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>抱</td>
<td>26</td>
<td>中央</td>
<td>51</td>
<td>新闻</td>
<td>76</td>
<td>机密</td>
</tr>
<tr>
<td></td>
<td>hug</td>
<td></td>
<td>ruling party</td>
<td></td>
<td>news</td>
<td></td>
<td>secret</td>
</tr>
<tr>
<td>2</td>
<td>知情者</td>
<td>27</td>
<td>提示</td>
<td>52</td>
<td>多菌灵</td>
<td>77</td>
<td>医疗</td>
</tr>
<tr>
<td></td>
<td>informant</td>
<td></td>
<td>warning</td>
<td></td>
<td>Insecticide</td>
<td></td>
<td>treatment</td>
</tr>
<tr>
<td>3</td>
<td>看到</td>
<td>28</td>
<td>妈妈</td>
<td>53</td>
<td>居民</td>
<td>78</td>
<td>设施</td>
</tr>
<tr>
<td></td>
<td>see</td>
<td></td>
<td>mom</td>
<td></td>
<td>resident</td>
<td></td>
<td>equipment</td>
</tr>
<tr>
<td>4</td>
<td>现在</td>
<td>29</td>
<td>小孩</td>
<td>54</td>
<td>抢救</td>
<td>79</td>
<td>农药</td>
</tr>
<tr>
<td></td>
<td>now</td>
<td></td>
<td>child</td>
<td></td>
<td>salvage</td>
<td></td>
<td>pesticides</td>
</tr>
<tr>
<td>5</td>
<td>走</td>
<td>30</td>
<td>速回</td>
<td>55</td>
<td>香港</td>
<td>80</td>
<td>教学楼</td>
</tr>
<tr>
<td></td>
<td>go</td>
<td></td>
<td>flashback</td>
<td></td>
<td>Hong Kong</td>
<td></td>
<td>teaching block</td>
</tr>
<tr>
<td>6</td>
<td>小女孩</td>
<td>31</td>
<td>家长</td>
<td>56</td>
<td>屋刀</td>
<td>81</td>
<td>黑</td>
</tr>
<tr>
<td></td>
<td>little girl</td>
<td></td>
<td>parents</td>
<td></td>
<td>knife</td>
<td></td>
<td>dark</td>
</tr>
<tr>
<td>7</td>
<td>兄弟姐妹</td>
<td>32</td>
<td>接力</td>
<td>57</td>
<td>闯入</td>
<td>82</td>
<td>塘沽</td>
</tr>
<tr>
<td></td>
<td>siblings</td>
<td></td>
<td>relay</td>
<td></td>
<td>trespass</td>
<td></td>
<td>pond</td>
</tr>
<tr>
<td>8</td>
<td>扩散</td>
<td>33</td>
<td>花园</td>
<td>58</td>
<td>癌症</td>
<td>83</td>
<td>地震</td>
</tr>
<tr>
<td></td>
<td>diffuse</td>
<td></td>
<td>garden</td>
<td></td>
<td>cancer</td>
<td></td>
<td>seismic</td>
</tr>
<tr>
<td>9</td>
<td>疯</td>
<td>34</td>
<td>人民</td>
<td>59</td>
<td>肿瘤</td>
<td>84</td>
<td>联系人</td>
</tr>
<tr>
<td></td>
<td>crazy</td>
<td></td>
<td>people</td>
<td></td>
<td>tumor</td>
<td></td>
<td>contact person</td>
</tr>
<tr>
<td>10</td>
<td>信息</td>
<td>35</td>
<td>没有</td>
<td>60</td>
<td>打工</td>
<td>85</td>
<td>求</td>
</tr>
<tr>
<td></td>
<td>information</td>
<td></td>
<td>no</td>
<td></td>
<td>work</td>
<td></td>
<td>please</td>
</tr>
<tr>
<td>11</td>
<td>孩子</td>
<td>36</td>
<td>雅安</td>
<td>61</td>
<td>广西</td>
<td>86</td>
<td>爱</td>
</tr>
<tr>
<td></td>
<td>kid</td>
<td></td>
<td>Yaan</td>
<td></td>
<td>Guangxi</td>
<td></td>
<td>love</td>
</tr>
<tr>
<td>12</td>
<td>帮忙</td>
<td>37</td>
<td>拐走</td>
<td>62</td>
<td>友情</td>
<td>87</td>
<td>收</td>
</tr>
<tr>
<td></td>
<td>help</td>
<td></td>
<td>abduct</td>
<td></td>
<td>friendship</td>
<td></td>
<td>have</td>
</tr>
<tr>
<td>13</td>
<td>人</td>
<td>38</td>
<td>爸爸</td>
<td>63</td>
<td>纪念日</td>
<td>88</td>
<td>现在</td>
</tr>
<tr>
<td></td>
<td>person</td>
<td></td>
<td>papa</td>
<td></td>
<td>anniversary</td>
<td></td>
<td>right now</td>
</tr>
<tr>
<td>14</td>
<td>请</td>
<td>39</td>
<td>牛奶</td>
<td>64</td>
<td>正在</td>
<td>89</td>
<td>知道</td>
</tr>
<tr>
<td></td>
<td>please</td>
<td></td>
<td>milk</td>
<td></td>
<td>doing</td>
<td></td>
<td>know</td>
</tr>
<tr>
<td>15</td>
<td>手机号码</td>
<td>40</td>
<td>转</td>
<td>65</td>
<td>去世</td>
<td>90</td>
<td>做</td>
</tr>
<tr>
<td></td>
<td>Phone</td>
<td></td>
<td>retransmit</td>
<td></td>
<td>pass away</td>
<td></td>
<td>do</td>
</tr>
<tr>
<td>16</td>
<td>肉毒</td>
<td>41</td>
<td>儿童</td>
<td>66</td>
<td>果粒</td>
<td>91</td>
<td>药</td>
</tr>
<tr>
<td></td>
<td>venomous</td>
<td></td>
<td>children</td>
<td></td>
<td>fruit grain</td>
<td></td>
<td>drug</td>
</tr>
<tr>
<td>17</td>
<td>牛奶</td>
<td>42</td>
<td>小男孩</td>
<td>67</td>
<td>邵逸夫</td>
<td>92</td>
<td>中药</td>
</tr>
<tr>
<td></td>
<td>milk</td>
<td></td>
<td>little boy</td>
<td></td>
<td>Shao yifu</td>
<td></td>
<td>herbal</td>
</tr>
<tr>
<td>18</td>
<td>朋友</td>
<td>43</td>
<td>发生</td>
<td>68</td>
<td>爆炸</td>
<td>93</td>
<td>平南</td>
</tr>
<tr>
<td></td>
<td>friend</td>
<td></td>
<td>happen</td>
<td></td>
<td>explosion</td>
<td></td>
<td>Pingnan</td>
</tr>
</tbody>
</table>
such as “women”, “kid”, “little girl”, “little boy”, “mother”, and “children”, as characters to appear in the news. Those identity-related words make up a large proportion of the top 100 keyword features, and some of them are repeated more than once. Additionally, time (“now”, “today”, and “afternoon”) and some specific words (“crazy”, “please reply quickly”, “please do not”, “hurry”, “announcement”, and “attention”) are utilized to create a sense of urgency and oppression in Chinese fake news. Last but not the least, real-life nouns, such as a specific place name (“China”, “Yaan”, “Shuicheng City”, “Guangxi”, “Beijing”, “Zhejiang”, “Pingnan”, and “Hong Kong”), also help fake news confuse the readers.

As shown in Table 2, the top 100 keyword features were also extracted from the English data set in this study. Compared with those in the Chinese data set, the attributes of English word features are more uniform, with the vast majority of keyword features closely associated with political and social hotspots (Trump, President, Obama, Clinton, White, police, Muslim, Black, gun, woman, protesters, and violence).

In summary, some commonalities were found in both Chinese and English fake news, which implied the feasibility of generating data from different languages: (1) fake news is often combined with popular social issues that are relevant to people’s daily lives; and (2) fake news usually contains exaggerated, highly inflammatory terms to catch people’s attention and evoke emotional resonance. Apart from these traits, fake news in Chinese and English contains some unique characteristics, which may be related to the different cultural backgrounds and language habits. Differences in the attributes of Chinese and English fake news may have implications for the results of subsequent cross-language fake news detection.

5.2 RQ2: Fake News Classifiers Built for Within- and Cross-Language Data Sets

Under different combinations of training data and test data, prediction results were derived and are recorded in Table 3; the best performances are highlighted in bold font. The following initial observations can be made.

1) The BERT model worked well in predicting fake news within the two data sets. The accuracy for Chinese fake news was recorded at 0.98, while that for English reached 0.96; both were very close to the value “1”. The best performance in previous studies for the two data sets was 0.9 and 0.86, respectively. Compared with previous work, we found a significant improvement in performance using the BERT model, regardless of the language of the data set. This is also supported by Rönnqvist, Kanerva, Salakoski, and Ginter (2019), who found that the BERT...
Table 2
*Top 100 Keyword Features Extracted from the English Fake News Data Set*

<table>
<thead>
<tr>
<th>No.</th>
<th>Features</th>
<th>No.</th>
<th>Features</th>
<th>No.</th>
<th>Features</th>
<th>No.</th>
<th>Features</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Trump</td>
<td>26</td>
<td>Country</td>
<td>51</td>
<td>Money</td>
<td>76</td>
<td></td>
<td>democratic</td>
</tr>
<tr>
<td>2</td>
<td>President</td>
<td>27</td>
<td>Image</td>
<td>52</td>
<td>Federal</td>
<td>77</td>
<td></td>
<td>Racist</td>
</tr>
<tr>
<td>3</td>
<td>People</td>
<td>28</td>
<td>Russia</td>
<td>53</td>
<td>Sanders</td>
<td>78</td>
<td></td>
<td>information</td>
</tr>
<tr>
<td>4</td>
<td>Obama</td>
<td>29</td>
<td>government</td>
<td>54</td>
<td>Woman</td>
<td>79</td>
<td></td>
<td>students</td>
</tr>
<tr>
<td>5</td>
<td>Clinton</td>
<td>30</td>
<td>don</td>
<td>55</td>
<td>administration</td>
<td>80</td>
<td>Military</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Hillary</td>
<td>31</td>
<td>bill</td>
<td>56</td>
<td>City</td>
<td>81</td>
<td></td>
<td>immigration</td>
</tr>
<tr>
<td>7</td>
<td>White</td>
<td>32</td>
<td>man</td>
<td>57</td>
<td>Speech</td>
<td>82</td>
<td></td>
<td>Day</td>
</tr>
<tr>
<td>8</td>
<td>State</td>
<td>33</td>
<td>vote</td>
<td>58</td>
<td>Public</td>
<td>83</td>
<td></td>
<td>supporters</td>
</tr>
<tr>
<td>9</td>
<td>news</td>
<td>34</td>
<td>FBI</td>
<td>59</td>
<td>School</td>
<td>84</td>
<td></td>
<td>intelligence</td>
</tr>
<tr>
<td>10</td>
<td>campaign</td>
<td>35</td>
<td>law</td>
<td>60</td>
<td>Security</td>
<td>85</td>
<td>Senate</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>media</td>
<td>36</td>
<td>Americans</td>
<td>61</td>
<td>Presidential</td>
<td>86</td>
<td>War</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>video</td>
<td>37</td>
<td>World</td>
<td>62</td>
<td>Cruz</td>
<td>87</td>
<td>Department</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>police</td>
<td>38</td>
<td>Show</td>
<td>63</td>
<td>Family</td>
<td>88</td>
<td>Washington</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Republican</td>
<td>39</td>
<td>Russian</td>
<td>64</td>
<td>National</td>
<td>89</td>
<td>United</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>America</td>
<td>40</td>
<td>GoP</td>
<td>65</td>
<td>Left</td>
<td>90</td>
<td>Voters</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>American</td>
<td>41</td>
<td>CNN</td>
<td>66</td>
<td>Group</td>
<td>91</td>
<td>Congress</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Donald</td>
<td>42</td>
<td>Gun</td>
<td>67</td>
<td>Children</td>
<td>92</td>
<td>Comey</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Black</td>
<td>43</td>
<td>Fox</td>
<td>68</td>
<td>Press</td>
<td>93</td>
<td>Office</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>women</td>
<td>44</td>
<td>Told</td>
<td>69</td>
<td>Candidate</td>
<td>94</td>
<td>VE</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>election</td>
<td>45</td>
<td>Twitter</td>
<td>70</td>
<td>Investigation</td>
<td>95</td>
<td>Post</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>time</td>
<td>46</td>
<td>Court</td>
<td>71</td>
<td>Support</td>
<td>96</td>
<td>Does</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Republicans</td>
<td>47</td>
<td>year</td>
<td>72</td>
<td>Wire</td>
<td>97</td>
<td>Protesters</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>house</td>
<td>48</td>
<td>Muslim</td>
<td>73</td>
<td>Interview</td>
<td>98</td>
<td>Great</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>party</td>
<td>49</td>
<td>states</td>
<td>74</td>
<td>Story</td>
<td>99</td>
<td>Did</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>political</td>
<td>50</td>
<td>Democrats</td>
<td>75</td>
<td>Attack</td>
<td>100</td>
<td>Violence</td>
<td></td>
</tr>
</tbody>
</table>

Table 3
*Classification Performance of Within- and Cross-Language Data Sets*

<table>
<thead>
<tr>
<th>Training Data</th>
<th>Testing Data</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese</td>
<td>Chinese</td>
<td><strong>0.98</strong></td>
<td><strong>0.98</strong></td>
<td>0.97</td>
<td><strong>0.98</strong></td>
</tr>
<tr>
<td>Chinese</td>
<td>English</td>
<td>0.79</td>
<td><strong>0.82</strong></td>
<td>0.75</td>
<td>0.79</td>
</tr>
<tr>
<td>English</td>
<td>English</td>
<td><strong>0.96</strong></td>
<td>0.95</td>
<td><strong>0.98</strong></td>
<td>0.97</td>
</tr>
<tr>
<td>English</td>
<td>Chinese</td>
<td>0.55</td>
<td>0.53</td>
<td><strong>0.99</strong></td>
<td>0.69</td>
</tr>
</tbody>
</table>
model is more helpful than other popular models in language detection.

2) Transforming the fake news detection model is feasible between the two languages, as Recall was high in both within- and cross-language data set situations, which respectively deliver 0.97, 0.75, 0.98, and 0.99 performances. Compared with Precision and F1, Recall is more preferred in this study since it represents the probability of successfully identifying fake news. This means that the model in our study can detect most of the fake news. It is admissible of the model’s errors as judging real news as fake. Nonetheless, we still expect the model to perform well in terms of accuracy while maintaining a high recall rate.

3) The overall performances of the Chinese–English data set were relatively good, while the results were mediocre in the English–Chinese data set. It is surprising to see that there is a significant difference in detection accuracy between the two cross-language data sets. To find out the reasons, more experiments were conducted. Based on RQ1, we could know that the hallmark of the keyword features in Chinese fake news was increased complexity and inclusiveness, compared to the keywords in English. At the same time, the amount of training data and test data were kept consistent in both Chinese–English and English–Chinese data sets. Hence, it is reasonable to speculate whether this difference is caused by the homogeneity of English fake news. In the next experiments, we calculated the similarity between Chinese and English keyword features, then removed the Chinese keyword features that have >70% similarity to any English keyword features. Considering that changes in the number of keyword features may have an impact on performance, we also conducted a control experiment with an equal number of keywords randomly removed. The whole results are shown in Table 4.

The specific keywords here refer to Chinese keyword features with >70% similarity to any English keyword features. Chinese1 stands for Chinese data after removing specific keywords, while Chinese 2 means data after removing keywords randomly. As we can see, after removing Chinese keywords with high similarity to English, the Chinese model’s accuracy in predicting English fake news dropped precipitously from 79% to 52%, which is close to the 55% results obtained by monitoring Chinese fake news with the model trained by the English data set.

At the same time, the control group performed better compared to the experimental group, as the accuracy only declined to 63%.

These series of experiments demonstrate the feasibility of cross-language fake news detection and point out the possibility that models built from languages with more complex keyword features can monitor the fake news in “simple” languages better.

6 Conclusion and Future Work

Fake news detection has been one of the trending research directions in the field of Information Science. Research focus on fake news from different languages and cultures is still lacking in today’s increasingly globalized world. Therefore, the potential contribution of this paper is to demonstrate the feasibility of cross-language fake news monitoring, providing a new research perspective for the detection of fake news to reduce the negative impact of disinformation on people's lives and social progress.

This study addressed the task of fake news detection in within- and cross-language data sets and explored the cross-language transmissibility of fake news detection models in the text dimension. We first compared the keyword features of English and Chinese fake news and found that the keyword features of Chinese fake news are more inclusive compared to those of English fake news. Aiming to achieve an effect that spreads viciously through the community, exaggerating tensions, targeting the underprivileged, and focusing on social hotspots are the main characteristics of Chinese fake news. It is reflected in the use of specific words, such as “help” and “please not”, and the use of vulnerable communities, such as “women”, “children”, and “students”, as the subjects of the news. As for the English fake news, the keyword features are simpler and more unified, which are closely related to social hotspots. “Trump”, “Obama”, “black” and other hotspots accounted for a very high weight in the keyword features.

Next, we built a BERT model to detect fake news in within- and cross-language data sets. The results
demonstrated the possibility of cross-language monitoring of fake news in English and Chinese, as Recall is relatively high in both cross-language data sets. Furthermore, the accuracy, with a range of 97%–99% for the within-language data set models, indicates the potential effectiveness of the advanced NLP models. Last, we found that in cross-language fake news detection, it is better to use the data set in a more inclusive language (in terms of keyword features), such as Chinese, to be the training data than to use the data in a unified language.

Admittedly, this study did not evaluate all the languages from different cultures; thus, we plan on extending this work by testing more languages that are used by large populations, such as Japanese, French, Spanish, and so on. Such evaluations will provide future evidence toward cross-language fake news detection. In our future work, we would like to propose a cross-language fake news detection model to monitor fake news in mixed cultures based on the findings of this study. Furthermore, it will also be interesting to explore the relationship between fake news topics and the feasibility of cross-language detection.

References


detection on social media: A data mining perspective. *ACM
SIGKDD explorations newsletter, 19*(1), 22-36.

Sun, H., Esho, O., Liu, J., & Pang, L. (2016). Discovering region features
based on user’s comments. In Li Y., Xiang G., Lin H., Wang M.
(Eds.) Social Media Processing. SMP 2016. Communications
in Computer and Information Science, vol 669. Singapore:
Springer. doi: 10.1007/978-981-10-2993-6_17

Retrieved from https://github.com/fxsjy/jieba

Tacchini, E., Ballarin, G., Della Vedova, M. L., Moret, S., & de Alfaro,
L. (2017). Some like it hoax: Automated fake news detection in

Yi Zhao, Haixu Xi, Chengzhi Zhang*

Exploring Occupation Differences in Reactions to COVID-19 Pandemic on Twitter

https://doi.org/10.2478/dim-2020-0032
received August 15, 2020; accepted September 15, 2020.

Abstract: Coronavirus disease 2019 (COVID-19) pandemic-related information are flooded on social media, and analyzing this information from an occupational perspective can help us to understand the social implications of this unprecedented disruption. In this study, using a COVID-19-related dataset collected with the Twitter IDs, we conduct topic and sentiment analysis from the perspective of occupation, by leveraging Latent Dirichlet Allocation (LDA) topic modeling and Valence Aware Dictionary and sEntiment Reasoning (VADER) model, respectively. The experimental results indicate that there are significant topic preference differences between Twitter users with different occupations. However, occupation-linked affective differences are only partly demonstrated in our study; Twitter users with different income levels have nothing to do with sentiment expression on covid-19-related topics.

Keywords: occupational differences, COVID-19, Twitter, topic discovery, sentiment analysis

1 Introduction

The outbreak of coronavirus Disease 2019 (COVID-19) is rapidly spreading worldwide and causing a profound effect on various aspects of society. At the time of writing, Johns Hopkins University reported more than 21,056,850 confirmed cases of COVID-19 globally, including 762,293 deaths and 13,100,902 recovered.1 To slow down the spread of virus, social distancing and self-isolation have been implemented globally, and social media has become an important channel for people to post their opinions and attitudes. The outbreak of the COVID-19 pandemic caused a rapid increase in COVID-19-related information on social media platforms, including YouTube, Facebook, Twitter, etc. (Abd-Alrazaq, Alhuwail, Househ, Hamdi, & Shah., 2020). Extensive research has shown that social media is a popular source of data in understanding public concerns and attitudes, and it is an important way to support crisis communication between the public and the government (Jordan et al., 2018; Shah & Dunn, 2019).

Occupation is one of the noteworthy demographic variables of Twitter users (Sloan, Morgan, Burnap, & Williams, 2015). In a recent study, Kern, McCarthy, Chakrabarty, and Rizoiu (2019) automatically assessed the personality of different occupations based on the tweets and found that personality was related to distinctive occupations. In addition, several studies have found an association between personality and perceptions of the COVID-19 situation (Carvalho, Pianowski, & Goncalves, 2020; Zajenkowski, Jonason, Leniarska, & Kozakiewicz, 2020). The above research may imply that Twitter users engaged in different occupations may have different topic preferences and sentiment expressions, when faced with COVID-19 pandemic. This motivates us to provide a comprehensive analysis based on Twitter data from the perspective of occupation differences, which helps policymakers understand the fine-grained public concerns.

Previous research has been focused on gender-linked affective or topic differences in social media platforms (Thelwall & Thelwall, 2020; Thelwall & Vis, 2017; Vegt & Kleinberg, 2020). Few works have been devoted to occupational differences in response to COVID-19, and

---

1 https://coronavirus.jhu.edu/map.html.
the research mainly focuses on a particular occupation, including world leaders (Rufai & Bunce, 2020) and college students (Duong, Luo, Pham, Yang, & Wang, 2020). Additionally, the income level may be another factor that affects topic preference and sentiment expression. The purpose of this study is to demonstrate that Twitter users with different occupation respond differently to COVID-19 and different income levels affect topic preferences and sentiment expressions.

2 Related Works

When it comes to discovering topics from Twitter, a majority of studies utilize the LDA technique and its improved versions. Abd-Alrazaq et al. (2020) utilized LDA to identify the main topic in the language of COVID-19 tweets. Guo, Vargo, Pan, Ding, and Ishwar (2016) conducted empirical research on two Twitter datasets for the 2012 presidential election, and they concluded that the LDA generated meaningful results. A lot of extended version of the LDA was used for topic analysis on social media platforms. N-gram LDA technique was used by E. H.-J. Kim, Jeong, Kim, Kang, and Song, (2016) to investigate the topic coverage of Twitter and news publications about the Ebola virus. Yan, Guo, Lan, and Cheng (2013) proposed a bi-term topic model to capture the word co-occurrence directly and enhance the topic quality. Sasaki, Yoshikawa, and Furuhashi(2014) implemented a Twitter–TTM model based on the topic tracking model that is competent in online inference, and the model can capture the dynamic topic trends on Twitter. To improve the interpretation of the topic model, especially for short text documents, Alkhodair, Fung, Rahman, and Hung (2018) developed a new model that combined Twitter–LDA, WordNet, and hashtags and assessed the effectiveness of the model against two real Twitter datasets.

Sentiment analysis for Twitter textual data is also a hot topic. Two main methods have been utilized to classify the sentiment polarity: supervised machine learning method and rule-based method (Hutto & Gilbert, 2014; Kim & Hovy, 2014). Tang et al. (2014) developed a new method that learned sentiment from specific word embedding and outperformed the previous top-performing system. Ren, Wang, and Ji (2016) used SVM for sentiment classification, and LDA was adopted for the improvement of word embedding. Recently, a large pretrained model was adopted for the classification task. BERT and RoBERTa were used by Duong et al. (2020) to examine the Twitter user’s sentiment expression. Supervised machine learning methods are powerful for sentiment classification, but large numbers of annotation data are needed for model training, which is costly and time consuming. Rule-based methods, including sentiwordnet (Esuli & Sebastiani, 2006) and Valence Aware Dictionary and sEntiment Reasoning (VADER) (Hutto & Gilbert, 2014), are also widely used on Twitter. Pandarachalil, Sendhilkumar, and Mahalakshmi(2014) proposed an unsupervised method that incorporated SenticNet, SentiWordNet, and SentislangNet, and it performed well on large scales. Chaithra (2019) presented a hybrid approach combining VADER and Naive Bayes to use the comments of the mobile unboxing videos to predict the sentiment, and the result concluded that VADER has improved the performance of the Naive Bayes classifier in predicting the sentiment of the comments.

To our knowledge, no previous study that adopted topic modeling and sentiment analysis to investigate the occupation distinctions in reactions to COVID-19 pandemic, which will be addressed in this study. As a result of a trade-off between cost and accuracy, LDA and VADER methods are applied in the study.

3 Material and Methods

3.1 Data Collection

Using the Twitter IDs provided by Chen, Lerman, and Ferrara (2020) on the repository, we collected 8716,289 tweets from May 1 to 15, 2020. In this study, we focus only on the English-language tweets and Twitter user populations of those who can identify occupational information from their self-description field. Since the self-description field is an open text in which Twitter users can choose to write whatever they like, it’s essential to design an appropriate method for occupation extraction. Inspired by Kern et al. (2019), regular expression matching was used to extract occupation information and we obtained 15,984 job titles from their research. The difference is that the job titles are aligned into major occupation groups according to the O*net alternative titles data. For example, the Assignment Editor and Morning News Producer are two job titles that belong to the Producers and Directors group.

3 The Occupational Information Network (O*NET) is the primary source of occupational information in the United States, and the O*NET database is the central of the network, which contains hundreds of standardized and occupation-specific descriptors on nearly 1,000 occupations covering the entire U.S. economy.
Furthermore, only the major occupation for each Twitter user was reserved during the process of occupation extraction. In our subsequent studies, we only considered the major occupation group as a research object. After removing retweets and filtering out corporate Twitter users with more than 50 tweets, we acquired 622,687 unique COVID-19-related tweets, belonging to 373,773 users, representing 800 occupations.

Due to the diversity of occupational categories and the restrictions of space, we selected nine occupations from different income levels. The classification criteria on occupation type and income levels are lacking, so we incorporate the salary data from the Bureau of Labor Statistics\(^5\) and information from Wall Street News (Suneson, 2019) to generate the following classification standard: Occupations in the high-income level group with incomes over $100,000 per year; occupations in the medium-income level group with incomes $30,000 to $100,000 per year; occupations in the low-income level group with incomes less than $30,000 per year. The nine occupations encompass a variety of social occupations, including technical occupations, managerial occupations, service occupations, etc. Of these nine occupations, the smallest one included 1,121 unique Twitter users. For balance, we randomly sampled 1,121 unique Twitter users from each occupation. The selected occupations are shown in Table 1.

### 3.2 Method

Many researchers have utilized Latent Dirichlet Allocation (LDA) to identify topics from social media data (Alkhodair et al., 2018; Asghari, Sierra-Sosa, & Elmaghraby, 2020; Giannetti, 2018). LDA is a three-level hierarchical Bayesian model (Blei, Ng, & Jordan, 2003), it is an unsupervised machine learning technique used to create a representation of documents by topic, where each topic consisted of a set of words. In this study, we employed an LDA algorithm from the Python Gensim library.\(^6\)

To obtain a clean corpus, we conducted data preprocessing at first. Then, we used regular expressions to remove URLs, HTML tags, and Twitter user mentions. Next, we also removed punctuation, stop words, and nonprintable characters from tweets. Finally, all tweets are lower-cased, tokenized, and lemmatized. It is well known that phrases are more meaningful than individual tokens. Hence bigrams and trigrams are created and added to the corpus. Before the corpus was fed into the LDA model (Blei et al., 2003), we used two kinds of document representation methods, bag of words (BOW) and term frequency–inverse document frequency (TF-IDF) (Salton & Yu, 1975), to represent tweets. The number of topics that need to be determined before running the LDA and it is a hyperparameter, so CV coherence measure is used to fine-tune our LDA topic model to obtain the optimal topic number (Röder, Both, & Hinneburg, 2015).

Sentiment analysis was also performed in our study. Sentiment analysis was conducted on the cleaned tweets using the VADER, a lexicon- and rule-based sentiment analysis model for social media text (Hutto & Gilbert, 2014). Hutto and Gilbert (2014) compared VADER with multiple methods, including Affective Norms for English Words (ANEW),\(^7\) Linguistic Inquiry and Word Count

---

\(^5\) [https://www.bls.gov/ooh/home.htm](https://www.bls.gov/ooh/home.htm).

\(^6\) [https://radimrehurek.com/gensim](https://radimrehurek.com/gensim).

\(^7\) [http://csea.phhp.ufl.edu/media/anewmessage.html](http://csea.phhp.ufl.edu/media/anewmessage.html).
Exploring Occupation Differences in Reactions to COVID-19 Pandemic on Twitter

(LIWC), SentiWordNet (Esuli & Sebastiani, 2006), the General Inquiurer, and machine learning techniques based on SVM, Naïve Bayes, Maximum Entropy, and the results show that the VADER model that is superior to these methods. One significant advantage of using VADER is that it can quickly and accurately obtain the sentiment of each tweet on such a large scale. The sentiment scores ranged from −1 to +1, with -1 as the most negative sentiment and 1 as the most positive sentiment; furthermore, when the score is between −0.05 and +0.05, it means neutral sentiment. The sentiment of each tweet was calculated using the library “vaderSentiment” in Python, which is the implementation of the python version of VADER model.

4 Experiment and Result

4.1 Selection of the Optimal Number of Topics

All 622,687 tweets were fed into the LDA model, we repeated the experiment on a different number of topics (ranging from 2 to 32) and reported the coherence value for two document representation methods. As shown in Figure 1, BOW performed better than TF-IDF, and the coherence value selected 14 as the optimal number of topics. Since the words of Topics 3, 5, 7, 8, and 13 are difficult to assign a specific theme, the authors reached a consensus on selecting nine highly relevant topics as the final result, as shown in Table 2.

As shown in Table 2, we assigned the potential themes to nine topics based on the top 15 most relevant words. Topic 0 discussed the preparation for reopening, with words such as “social distancing,” “reopen,” “school,” and “guideline” indicating this overarching theme. Topic 1 was primarily about U.S. President Donald Trump’s lies about the COVID-19. Topic 2 was about the coronavirus new cases and deaths, which were identified in tweets that mentioned the rapid growth in the number of confirmed cases. Topic 4 talked about free online support, and Twitter users on the topic mainly discussed joining a local free online support team to provide useful pandemic-related information. A potential theme identified in Topic 6 was protests against the stay-at-home order, with the words “fuck” and “shit” showing the feelings for the order. Topic 9 primarily referred to the risk caused by COVID-19, which can affect business, jobs and food. Topic 10 was measures to slow the spread of COVID-19, including wearing masks and staying at home. The theme of research on the vaccine and treatment was identified in topic 11. Topic 12 was related to virus misinformation and fake news, and Twitter users debated whether the novel coronavirus originated in Wuhan’s laboratory and spread around the world.
Figure 2 shows the topic distribution of Twitter users engaged in different occupations. Topics 3, 5, 7, 8, and 13 are grouped into a separate group (other Topics).

Overall, there is a significant difference in topic concern between Twitter users with different occupations, while Twitter users at different income levels showed a slight tendency toward some topics. For high-income level groups, Topic 2 is more concerned with Computer and Information Research Scientists (CIRS) and Dentists, General (DEN) and Marketing Managers (MM) cares more about Topics 4 and 9, and it indicates the topic preference for Twitter users with difference occupations. In addition, Topic 4 also attracts more attention to high-income and medium-income occupations than to low-income occupations. Twitter users in low-income occupations are more interested in Topic 1 than Twitter users in high-income and medium-income occupations. Compared with other topics, all of Twitter users who engaged in these nine occupations showed less curiosity about the virus misinformation and fake news (Topic 12).

Table 2
Topics in COVID-19-Associated Tweets

<table>
<thead>
<tr>
<th>Topic</th>
<th>Themes</th>
<th>Top 15 most relevant words</th>
<th>Number of tweets</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Preparation for reopening</td>
<td>social_distancing, reopen, open, school, take, people, back, life, place, economy, say, child, close, measure, guideline</td>
<td>35,336</td>
</tr>
<tr>
<td>1</td>
<td>President’s lies about COVID-19</td>
<td>trump, american, say, president, response, cdc, america, death, lie, white_house, claim, government, administration, medium, covid</td>
<td>40,838</td>
</tr>
<tr>
<td>2</td>
<td>Coronavirus new cases and deaths</td>
<td>case, death, test, new, report, number, positive, total, covid, update, day, state, rate, high, data</td>
<td>50,863</td>
</tr>
<tr>
<td>3</td>
<td>*</td>
<td>get, see, come, work, covid, back, hard, time, like, hit, interesting, one, since, bit, well</td>
<td>35,531</td>
</tr>
<tr>
<td>4</td>
<td>Free online support</td>
<td>help, support, need, crisis, student, work, time, learn, online, community, free, join, impact, new, team</td>
<td>49,828</td>
</tr>
<tr>
<td>5</td>
<td>*</td>
<td>people, know, get, think, covid, like, make, would, good, thing, one, say, die, want, bad</td>
<td>107,280</td>
</tr>
<tr>
<td>6</td>
<td>Protests against the stay-at-home order</td>
<td>state, fuck, vote, governor, house, party, game, get, man, right, play, bbc_news, stay-at-home_order, shit, protest</td>
<td>25,435</td>
</tr>
<tr>
<td>7</td>
<td>*</td>
<td>day, love, read, great, time, watch, video, one, today, story, thank, good, new, share, stayhome</td>
<td>62,833</td>
</tr>
<tr>
<td>8</td>
<td>*</td>
<td>may, week, month, next, plan, ease, due, year, end, last, start, two, restriction, thread, day</td>
<td>29,162</td>
</tr>
<tr>
<td>9</td>
<td>The risk caused by COVID-19</td>
<td>health, business, worker, pay, care, government, public, work, need, risk, job, service, food, staff, help</td>
<td>54,488</td>
</tr>
<tr>
<td>10</td>
<td>Measures to slow the spread of COVID-19</td>
<td>home, stay, people, work, safe, mask, keep, get, wear_mask, order, face, family, back, protect, life</td>
<td>48,237</td>
</tr>
<tr>
<td>11</td>
<td>Research on the vaccine and treatment</td>
<td>vaccine, world, virus, patient, covid, use, disease, could, doctor, study, find, spread, cure, treatment, people</td>
<td>39,989</td>
</tr>
<tr>
<td>12</td>
<td>Virus misinformation and fake news</td>
<td>corona, virus, china, india, fight, wuhan, chinese, world, spread, lab, come, country, war, outbreak, govt</td>
<td>21,210</td>
</tr>
<tr>
<td>13</td>
<td>*</td>
<td>like, look, question, year, well, say, answer, ask, science, london, wow, would, could, break, first</td>
<td>21,657</td>
</tr>
</tbody>
</table>

Note: * indicates that it is difficult to assign a specific theme to the topic and will not be analyzed in a subsequent section.
To understand occupation differences in sentimental responses to COVID-19, we analyze the sentiment distribution toward the different topics in Figures 3 and 4.

Overall, Twitter users in different occupations express different sentiments on different topics, but the sentiment expressed by Twitter users of different occupations toward a topic seems to have nothing to do with their income levels. Overall sentiment trends are almost identical for Twitter users with different occupations for most of the topics. For example, regardless of the Twitter user’s occupation, negative sentiment ratio is greater than positive sentiment ratio toward Topic 1 (President’s lies about the COVID-19), that is, U.S. President Donald Trump admitted to journalist Bob Woodward that he played down the severity of the pandemic11; the number of negative tweets posted on Twitter on this topic is very alarming. Moreover, online support groups offered medical support to the public during the COVID-19 pandemic, enhancing the public’s ability to self-protection (Gong, Xu, Cai, Chen, & Wang, 2020), and thus about 58% of tweets positively responded to the Topic 4 (free online support). At the same time, differences exist in emotional expression toward Topic 2, that is, CIRS, MM, and DEN respond more positively sentiment toward the coronavirus new cases and deaths, whereas, Financial Analysts (FA), Farmworkers, Farm, Ranch, and Aquacultural Animals (FFRAA), Production Workers (PW), and Landscaping and Groundskeeping Workers (LGW) are more likely to express negative feelings on this topic. Additionally, for Topic 6, the positive sentiment ratio of CIRS and LGW is a little higher than that of other occupations, and MA tends to be subject to less positive sentiment on the topic of protesting against the stay-at-home order.

4.3 Topic-based Sentiment Analysis among Different Occupations

In this study, we collected 622,687 tweets from 800 occupations and selected nine occupations with different income levels as topic for the research. We found that there was a significant difference in topic concern between Twitter users with different occupations, but sentiment expression differences only existed toward Topic 2, and income level seems to have nothing to do with emotional expression. These findings only partly demonstrate the hypothesis that Twitter users engaged in different occupations have different sentiment expressions. Furthermore, our study finds significant occupation differences in topic preference.

As a short paper, there are also a few limitations. First, we only compare nine occupations, and since the sample size is not big enough, it is necessary to investigate all the occupations and may provide more convincing results. Second, supervised machine learning can apply to topic

discovery and sentiment analysis to improve the topic quality and classification accuracy, it may provide a better analytical result.

There are some related studies that can be implemented in the future. First, this article mainly focuses on sentiment differences toward topics among groups of Twitter users clustered by occupations; a more nuanced exploration should be conducted on emotional difference, which could provide us with an in-depth understanding of people’s actions. Second, the topics posted on Twitter have been continuously changing with the development of the COVID-19 pandemic, so merely examining the topic preferences of Twitter users for a given period of time may be insufficient. Future researchers can therefore conduct the research throughout the lifetime of the pandemic and explore that whether the phenomena of topic preferences of Twitter users with different occupations still exist.

Acknowledgements: This work is supported by Jiangsu Social Science Fund (grant no. 20TQA001), Humanities and Social Science Research Fund of the Ministry of Education in China (grant no. 18YJC840045), and Postgraduate Research & Practice Innovation Program of Jiangsu Province (grant no. KYCX20_0347). The authors are grateful to all the anonymous reviewers for their precious comments and suggestions.
References


Vegt, I. v. d., & Kleinberg, B. (2020). Women worry about family, men about the economy: Gender differences in emotional responses
Yi Zhao, Haixu Xi, Chengzhi Zhang


Public Sector Employee Perspective towards Adoption of E-Government in Pakistan: A Proposed Research Agenda

1 Introduction

In recent years, governments across the world have shown a growing interest in shaping the public sector using Information and Communication Technology (ICTs) to provide services to their residents and other stakeholders. The United Nations named this process as “e-government” (Samijadi, 2017). Governments in developing and developed countries are engaging more and more in using ICTs as a mediator to resolve citizens’ concerns regarding increased accountability, corporate responsibility, and interaction between government organizations and people to access information (G. F. Khan, 2015).

A wide variety of modern technologies have been developed by various businesses (Hasibuan & Syahrial, 2019). In the present state of COVID-19, where the majority of businesses are affected, employees and citizens are motivated or compelled to do business from home, and the demand for proper technology is growing (Yue et al., 2020). This pandemic encounter is a global transformative event that has intensified the need for complete online work. Although there is an increasing need for intelligent systems, including e-government, e-banking, e-commerce, and other digital systems, the present outbreak has increased the importance of web and automation modes (Sepasgozar, Ramzani, Ebrahimzadeh, Sargolzae, & Sepasgozar, 2020; Twizeyimana & Andersson, 2019). Nonetheless, information system conduct must be investigated in various contexts, and researchers analyze technology adoption from different perspectives (Hasibuan & Syahrial, 2019; Sepasgozar et al., 2020).

TAM was developed on the basis of the Theory of Reasoned Action (TRA) (Davis, 1989). This model, considered as one of the most exquisite models in research to describe e-gov adoption by Carter and Bélanger (2005), suggests that trust, attitude, and Perceived usefulness PU are unwaveringly influenced in the intent to adopt e-gov.
The significance of this study is the impact of public sector employees in Pakistan who had seldom studied or had been exposed to such studies before this pandemic. Since this pandemic, most businesses are affected, and hence the needs for complete online work and e-gov systems have increased. This proposed conceptual model directly affects factors such as PU, attitude, and trust thereby enhancing efficiency, reducing corruption, and increasing transparency in the public sector framework. As an underdeveloped region, Pakistan and its infrastructure are not very well established. No research has been found on the adoption of e-gov among employees in Pakistan's public sector. This theoretical model is useful for addressing the need and interpreting the influences of e-gov adoption among public sector employees in Pakistan. The importance of using new technology like e-gov in the public sector seems to be apparent to everyone, but what is less found in our country is adoption and implementation against e-government in the public sector and quitting traditional ones.

2 Theoretical Background

A revolution in ICT has been witnessed in recent years. This transition has prompted more and more improvements in everyday life around the globe. It has transformed the way governments throughout the world connect with their people, government departments, businesses, stakeholders, and all other employees. These modifications are often stated as e-gov (Sang & Lee, 2009).

2.1 E-government

The word “e-government” means electronic government. Though, digital, online, internet, and the connected government have also been identified and used synonymously in the literature (Grönlund, 2004). The UN describes e-government as the use of the internet and the world wide web to provide government information and public services for people (Unpan, 2002). The World Bank views e-gov as the use of ICTs by the government authorities capable of transforming ties with branches of government, citizens, and other businesses (Khalil et al., 2002). e-Gov is the adoption of ICTs by the government to provide facilities to government authorities, businesses, and citizens (Shah & Lim, 2011). The transition from traditional public services to e-gov allowed governments to improve management efficiency, increase transparency, make information easy to access, and provide efficient services (Kayani, Haq, Perwez, & Humayun, 2011). The distinguished e-gov scholars argue that e-gov adoption will give a nation the following benefits (Hassan & Lee, 2015):

**Improve the Efficiency**
- Cost Minimization and Maximum Output
- Increase Transparency level
- Accountability
- Standards and Policy Enforcement
- Quick Delivery
- Reduce of Corruption
- Improved Public Connections
- To Exchange and Share All Information

2.2 Barriers to E-government in Pakistan

The Government of Pakistan’s very first web portal was launched in 2005 by the Department of Science and Technology, Electronic Government Directorate (EGD). Pakistan’s e-Gov Directorate (EGD) was formed as an IT ministry unit, explicitly focusing on e-gov. Governments are aiming at greater transparency and accountability in policy-making, increasing the efficiency and cost-effectiveness of the provision of public services to people (S. Khan, Rahim, & Maarop, 2018).

However, the e-govt projects could not be successfully implemented in Pakistan, given these critical initiatives. Pakistan thus placed 153rd in the e-government list compared to its neighboring countries, China (45th), Iran (89th), and India (100th) as well as Bangladesh (119th) (E. S. UNPAN, 2020). Pakistan is a developing country that is advancing owing to the CPEC economy, as many countries invest and assist Pakistan in improving its infrastructure. Simultaneously, it faces challenges in developing its ICTs infrastructure and in establishing a framework for educating its large population (Hassan, Madad, Das, Akhtar, Jehan, & Research, 2019).

Pakistan also faces several challenges in implementing e-gov projects in the country, like any other developing country (Khan et al., 2018). In the e-gov sector, failure is defined as incapability in attaining its objectives. Government IT programs in developing countries collapse because of an absence of an internal political will, political dominance, poor management, overall vision, lack of skills, and insufficient technological infrastructure (Kayani et al., 2011). Despite a lot of support in implementing e-gov, numerous challenges need to be tackled. In many countries around the world, particularly in developing countries, e-gov fails. The challenges and impediments to Pakistan’s e-gov adoption process are as follows (Kayani et al., 2011):
Inadequate expertise and human resources
- Abilities and skills deficiencies among senior officials
- Economic crisis and lack of legitimacy
- Accessibility and ICTs infrastructure at low levels
- Delicate interdepartmental coordination
- Official and legal recognition of electronic documents
- The lack of cooperation and interaction among the various departments
- Internal disputes and external challenges
- Insufficient access to the Internet in remote areas
- Conservative-minded bureaucracy and red-tapism
- Instability of political system and corruption
- Scarcity of people in using technologies
- Inadequate allocation of resources
- Awareness about applying and the significance of e-gov
- Insufficient understanding of the benefits of e-gov

2.3 Technology Acceptance Model (TAM)

TAM is the foundation for the adoption of technology models and is used in various fields. Most articles also attempt to evaluate the aspects of technological adoption from the users’ point of view in distinct environments. The findings of specific articles show that the factors impacting technology adoption in diverse organizations are not similar, and the TAMs should always be updated (Rana, Dwivedi, Williams, & Weerakkody, 2015; Sepasgozar, Hawken, Sargolzaei, & Foroozanfa, 2019). TAM is essential in describing and predicting the adoption of e-gov and numerous adoptions of other technologies. As a result, many scholars send TAM to consider adoption through e-government (Rana et al., 2015; Roy, Chartier, Crête, & Poulin, 2015). However, inadequate studies are showing that public sector employees in Pakistan are appropriate. Recently, in some studies, e-gov adoption of Tunisian citizens (Nasri, 2019), Mehr bank employees in Iran (Sepasgozar et al., 2020), and Malaysian small- and medium-sized enterprises (SMEs) have been identified as influential factors by using a TAM model (Soong, Ahmed, & Tan, 2020).

3 Hypotheses and Proposed Model

3.1 Proposed Conceptual Model

Based on prior related research, an interconnected conceptual framework is developed and presented.

This integrative approach is intended to provide a full insight into the factors impacting e-gov adoption by public sector employees in Pakistan. It consolidates Pakistan’s conceptual study model for e-gov adoption, incorporating variables specific to the Pakistan context. The perceived ease of use influence perceived usefulness, trust, and attitude and its effects on Pakistan’s Intention to use new technologies and e-gov in the Pakistani public sector. The conceptual framework expects the adoption of e-gov is presented in Figure 1. The considerations and their corresponding variables found in the literature are presented as follows.

3.2 Perceived Ease of Use and Perceived Usefulness

As the e-gov system is easy to use, employees can easily find useful and meaningful information to help them learn the usefulness of these services. Perceived ease of use (PEOU) has a direct relation to perceived usefulness (Alkali & Abu Mansor, 2017; Xie, Song, Peng, & Shabbir, 2017). PU is another essential TAM construct that expects employee intention to adopt an e-govt system. It suggests that the more employees perceive e-government services that are useful/beneficial in comparison to traditional offline modes, the more inclined they are to use them. Besides, earlier research found PEOU to be a significant antecedent of trust (Wu & Chen, 2005; Xie et al., 2017), besides, PU has a positive impact on trust as well as a significant predictor of intentions of the e-govt adoption (Al-Sharafi, Arshah, Alajmi, Herzallah & Qasem, 2018; Xie et al., 2017).

The theory of TAM indicates that PU and PEOU are associated with the attitude for adopting advanced technology, which influences the intention of adoption. A previous study suggests that PU is a reliable indicator of the adoption of technology (Ökcü, Koksalmis, Basak, & Callisir, 2019). PU has a direct effect on the intention to adopt in some studies and has enhanced its impact on employees through attitude (Iskandar, Subramaniam,
PU and PEOU drive trust and attitudes of employee e-gov adoption, which are impacted by certain external factors. For modern technology or e-gov, such considerations are easy to understand. They can be used extensively to tackle e-gov adoption problems concerning public sector employees (Taylor & Todd, 1995; Xie et al., 2017).

Hence the following hypotheses are:

**H1:** PEOU will be positively affected by the PU of e-gov adoption in employees.

**H2:** PEOU will be positively influenced by the trust for e-gov adoption in public sector employees.

**H3:** PEOU will positively impact the attitude of e-gov adoption in public sector employees.

**H4:** PU will have a positive effect on employees’ trust to adopt e-gov adoption in the public sector.

**H5:** PU will have a positive influence on employees’ attitudes of e-gov adoption.

**H6:** PU will significantly impact the employee intention to adopt e-gov in the public sector.

### 3.3 Trust

Trust is an essential part of human interactions. Trust is defined as honesty, confidence, and reliability; e-govt adoption highly concerns a high level of trust, described as a complicated concept that expresses one party’s willingness (Xie et al., 2017). Trust is considered an emotional and intellectual desire. The shared among stakeholders plays a crucial role in trust and attitude. Specific findings have shown a positive relationship between trust and attitude (Munoz-Leiva et al., 2017; Xie et al., 2017). Numerous researchers have demonstrated the significant impact of trust on e-gov intentions (Khasawneh, Rabayah, & Abu-Shanab, 2013). Therefore, trust has been hypothesized as follows:

**H7:** Trust will significantly impact the attitude of e-gov adoption in public sector employees.

**H8:** Trust will positively influence employees’ intentions toward e-gov adoption.

### 3.4 Attitude

Attitude asserts the negative or positive feelings of a person toward a specific behavior. The most apparent precursor of intent is an attitude that is well described as an individual’s negative or positive emotions about the behavior expected (Yeh & Wu, 1999). A previous study on attitude toward technological content indicates that individuals are motivated to be using a system to assist them in their activities (Yeh & Wu, 1999). Different findings in previous literature expressed the fact that the attitude of employees significantly impacts the intent to adopt new technology (Iskandar et al., 2020). Recent research found a significant association between attitude and intention in public management settings (Zahid & Haji Din, 2019). Some studies (Iskandar et al., 2020; Munoz-Leiva et al., 2017; Xie et al., 2017; Yeh & Wu, 1999) have indicated that attitudes may influence the higher use of intention to adopt.

The hypothesis was thus proposed as follows:

**H9:** Attitude has positively influenced the employees’ intention toward e-government adoption.

### 4 Proposed Methodology

#### 4.1 Survey Administration

A survey with more than 300 public sector employees in Pakistan will be conducted to test the hypotheses. An online study would be designed for data collection because it is a quicker and more efficient way to gather views. Its survey’s general population is all employees with in-depth knowledge of e-gov and their services. Since this study was intended to gather information on the e-gov adoption decision-making experience and opinion, it is therefore initially proposed that the study should be carried out at one public sector organization, and the sample size could be expanded in the future by incorporating some similar organizations (Ahmed, Qin, & Aduamoah, 2018).

#### 4.2 Proposed Measures

Five factors premise the proposed framework that reflects PEOU, PU, trust, attitude, and the intention to adopt e-gov. In a survey designed by Munoz-Leiva et al. (2017), the final questionnaire contained 16 items. Three statements PEOU and five PU statements are expected to be measured updated from Ahmed et al., 2018 and Venkatesh & Davis, 2000 were conducted using the TAM model. Three items are used to measure trust (Munoz-Leiva et al., 2017) in the same way as three items are used to measure attitude adapted from Venkatesh and Davis (2000). Finally, IU has been measured by two items taken from (Munoz-Leiva et al., 2017). A 5-point Likert scale measured each item
(1 = strongly disagree, 5 = strongly agree). There were also numerous questions concerning the adoption of the e-govt system following the outbreak in the survey.

4.3 Data Analysis

This study aims to test the hypothesis and the theoretical framework using a path analysis proposed based on a sequence of analyzing multiple regression. The Pearson correlation coefficients shall detect potential multicollinearity problems for all objects in the survey instrument. Benchmarks with a 0.05 $P$-value and a minimum Chronbach alpha of 0.70 are established for each element. Besides, model fitness will be calculated through Normed Fit Index (NFI), Goodness of Fit Index (GFI), Comparative Fit Index (CFI), and Root Mean Square Error of Approximation (RMSEA). The hierarchical multiple regression is carried out by the R2 values to check the relationships between independent and dependent variables.

5 Conclusion & Future Direction

This study examined the precedents of the intention to adopt the e-government system in Pakistan. Presently, Pakistan faces many critical issues such as the political system, legitimacy crisis, internal disputes, external threat, bureaucratic red-tapism, corruption, and economic crisis. Under these circumstances, Pakistan needs a technology-driven framework to develop and advance the e-government system in the public sector. This study identifies the determinants and influences factors of the intentions of public sector employees to adopt an integrated conceptual model for e-government services. The findings of the survey have meaningful consequences and indicate that PEOU, PU, trust, and attitude are closely connected to e-government system adoption. The more individuals have trust and attitude, the more they intend to use the e-government system. Policymakers will help to improve efficiency, enhance transparency, and reduce public sector corruption to carry out this study. The proposed design model can be used in the public sector and the private sector and is significant to the local context of Pakistan’s e-government framework. In the future, we would like to conduct our empirical research to know the citizens’ opinions on the successful adoption and propagation of the e-government system in Pakistan.

References


Communication Technology (Vol. 843, pp. 1002–1012). Berlin, German: Springer. doi: 10.1007/978-3-319-99007-1_93


Adoption of E-Services and Quality of Life Among Older Consumers in China

Abstract: E-services projects have been successfully implemented and adopted in many countries worldwide. China, one of the developing countries, has also been transitioning itself into this e-services structure. The success of such an effort depends not only on the support of the government but on citizens' readiness to accept these services, also a central feature. For advanced nations, many studies classify citizens' acceptance features of e-services, but for emerging countries, such research works are less; for China, where citizens' adoption of e-services is actually stumpy as compared to developed countries, there are only few studies available in the Chinese context, and this research is aimed at addressing this gap. In this study about e-service adoption, among the available technology adoption models, a further modification in the original “Unified Theory of Acceptance and Use of Technology” (UTAUT) is suggested and used in order to check its credibility and validity in the Chinese context. This exploration is important in numerous ways and will guarantee the development of China by way of supporting e-services policymakers.

Keywords: e-services, older consumers, quality of life

1 Introduction

In spite of there being no commonly known explanation of e-services (Javalgi et al., 2004), the subsequent definition suggested by de Ruyter et al. (2001) is noteworthy: “E-service is a cooperative, content-centered and Internet-based customer service, determined by the customer and combined with related organizational customer support procedures and skills with the objective of firming the customer-service relationship”.

According to Sudbury-Riley et al. (2015), the widespread distribution of information and communication technology (ICT) over the past few years has not only broadened the concept of knowledge and information sharing but also changed the responsiveness of economical and societal environments by turning the whole world into a global village. And because of this technological advancement, this e-facilitation accessible and available regardless of place and time, allowing single citizens, organizations, and other institutes to increase their participation in the diversity of this process (Arduini & Zanfei, 2014).

Unlike direct retail dealings, the physical separation of vendors and consumers in anonymous Internet e-marketplaces creates doubts (Marshall et al., 2012). Transaction-related doubts constitute a dominant dispute in online commerce situations, specifically when e-retailers are unacquainted with the buyer (Featherman & Hajli, 2016). E-commerce has become established, but users still have worries concerning the moral practices of e-retailers (Roman, 2007).
information and services, and escalating the transparency and liability of a business deal. However, unluckily, numerous nation states have been encountering the issue of low adoption rate of e-services by its residents (Carter & Bélanger, 2005).

2.1 E-Services in China

According to the United Nations E-Services Survey (2012), “Countries have positioned E-Services advantages and ICT practices for the individuals to promote and heighten public segment competences and reorganize government structures to upkeep feasible development”. Because of this abundance of e-services for consumers, a healthy growth rate of these services has been observed not only in the developed, but also in developing, nations. China is also showing the same trend.

Till recently, a large number of studies have worked on information systems and a limited number have focused on social e-services (Webster & Watson, 2002). There have been a variety of studies that pinpoint citizens’ adoption of e-services in technically advanced countries, but literature that attempts to find the adoption features in developing countries is very scarce. After a careful evaluation of the published articles on citizens’ adoption of e-services, it is found that there are no research reports that study citizens’ adoption of e-services and its impact on their quality of life (QoL) in the Chinese context. Therefore, this gives birth to an unfilled gap in research work, which attempts to classify the issues that affect citizens’ adoption of e-services in the Chinese context and its effect on their QoL.

2.2 Older Consumers and Aging Population

One of the socio-demographic characteristics that differentiate ethical beliefs is age. Age has been found to be a significant determinant in most of the studies (Sudbury Riley et al., 2012). As people age, they become increasingly dissimilar with respect to lifestyles, needs, and consumption habits (Sudbury-Riley et al., 2015), which of course renders chronological age even more useless in terms of targeting Baby Boomers, as research has explained that aging is multidimensional.

Regardless of the reputation of this marketplace in terms of its dimensions and certainly buying influence, it remains an under researched fragment. This condition is mainly true of the studies carried out on seniors outside the United States (Kohlbacher et al., 2011). Despite many researchers listing demographic change and need satisfaction as measures of the QoL, absence of research in this vast field is observed (Kohlbacher & Chéron, 2012). Some researchers have, however, studied emerging opportunities in connection with the demographic trend; but their research is not based on primary empirical data and does not investigate older people as customers (Kohlbacher et al., 2015). The present study therefore makes a contribution to knowledge by focusing on adoption of e-services and QoL of older people outside the United States, all of whom are important from an older consumer perspective (Sudbury Riley et al., 2012).

Studies have revealed that mature people are extra resilient while adopting technology because of many factors, such as difficulty in understanding, usability, and the perceived complexity of new technology (Kohlbacher & Hang, 2011); however, developing products and technologies for older users is not an easy task (Kohlbacher et al., 2011; Compagna & Kohlbacher, 2015). Surely, technology has brought benefits to many aspects of aged people’s lives and will definitely bring positive changes to their QoL (Kohlbacher & Hang, 2011).

3 Quality of Life

Sirgy (1986) defined QoL as the satisfaction of the hierarchical need of the followers of a specified culture, and satisfaction shows the QoL of the aggregate members of a society. With the increased amount of satisfaction of needs, increased QoL is expected (Sirgy, 1986).

Conversely, the term QoL not only remains the part of day-to-day conversation but is also widely used by researchers of various fields ranging from natural science to social science (Farquhar, 1995). Due to its multidisciplinary practice, it begs the question “where has it come from?” Actually, the definition was first witnessed in the context of physical goods – a home, a car, more and better appliances, and the money to travel and retire (Farquhar, 1995).

Humanity is shifting and QoL is becoming a vital subject for numerous clusters. There is increasing appreciation that QoL of aged consumers is a multifaceted and complicated state (Diener & Suh, 1997) that needs more consideration (Hyde et al., 2003). Studies have also shown that in the retirement ages, people are enjoying longer and healthier lives and have more income at their disposal; so, older age is not just a part of the life course that is viewed as a weak and disabled part of life (Hyde et al., 2003).
4 Theoretical Framework

This research will use the UTAUT model developed by Venkatesh et al. (2003). They invented the model to illustrate a more comprehensive demonstration of the technology-acceptance concepts that had been in use formerly, by combining those (AlAwadhi & Morris, 2009).

This study is suggesting a modification in the UTAUT model initially proposed by Venkatesh et al. (2003), in order to make it suitable to the adoption of e-services in the Chinese context. According to the modified UTAUT model, it is assumed that Trust, Performance Expectancy, Effort Expectancy, Social Influence, and Perceived Risk are considerably prompting the Behavioral Intention of the citizens, and Behavioral Intention and Facilitating Conditions are considerably prompting the Use Behavior of the citizens.

These constructs are elaborated below.

1. **Trust (T):** Deficiency of trust is an obstacle to the implementation of electronic services. The electronic definition of trust could be as “opinion of assurance in the e-marketer’s consistency and truthfulness” (Carter & Belanger, 2005). Several researches proposed that considering trust as a major component of any e-facilitation, many consumers perceive the deficiency of this component as a barrier to the adoption of such services. This study will measure the trust with the supposition that citizens’ Behavioral Intention to adopt the e-services is swayed positively by a higher level of trust.

2. **Perceived Risk (PR):** “The probable for damage in the search of an anticipated consequence of using an e-service”, and it surrounds three paradigms, they are as follows: Financial Risk - “The apparent monetary cost allied with the premature purchase price as well as the later preservation cost of the product”; Time Risk - “individuals may undergo time lose when fetching in a bad buying choice by deteriorating time, examining and creating the buying attitude, or finding a suitable substitute the expectation of a customer are not addressed”; Privacy Risk - “Likelihood of losing hold over one’s personal and private data, such as when statistics about you is used without your awareness or permission”. This study will calculate the perceived risk with the supposition that perceived risk has a straightforward effect on customer’s behavioral intention to adopt e-services.

3. **Performance Expectancy (PE):** “It is the magnitude of a certainty to which a person trusts that using the system will benefit him or her to achieve advantages in career performance”.

4. **Effort Expectancy (EE):** “The amount of easiness linked with the use of a particular system”, and it surrounds two paradigms that formulate the idea of EE. They are as follows: Perceived Ease of Use - “The point to which a respondent imagines that interacting with a particular innovation would be free of exertion”; and Complexity - “The point to which a system is supposed as comparatively tough to comprehend and use” (Venkatesh et al., 2003).

5. **Social Influence (SI):** “The magnitude of a belief to which a person feels significant that others accept as true he or she should use the novel system”, and it covers Image - “The gradation to which use of an invention is observed to improve one’s image or reputation in one’s social circle” (Venkatesh et al., 2003).

6. **Facilitating Conditions (FC):** “The point to which a person trusts that a structural and practical arrangement subsists to help the use of the system”, and it covers three concepts. They are as follows: Perceived Behavioral Control (reveals insights of inner and outer controls on performance and includes self-efficacy, source assisting environments, and technology assisting environments); Facilitating Conditions (impartial aspects in the surroundings that create a stress-free environment to do action, containing the facility of computer provision); and Compatibility (the amount to which an invention is supposed as being reliable with prevailing standards, prerequisites, and capabilities of possible adopters).

**Use Behavior (UB):** According to Venkatesh et al. (2003), “Use behavior is user intentions to use an information system and subsequent usage behavior”. The theory holds that out of the four key constructs, Facilitating Conditions is the direct determinant of user behavior.

**Behavioral Intention (BI):** “Individual’s particular likelihood that he or she will carry out the conduct in question” (Venkatesh et al., 2003). It is the individual’s intention to embrace and make use of the e-services in future. The BI is used in many theories to forecast technology adoption (Dwivedi & Irani, 2009). The BI is believed to have a direct effect on the adoption choice. This study will evaluate the BI by Intention, forecasting, and use of e-services.

The above-mentioned six direct factors are likely to be affected by the controlling elements, namely, Gender, Internet Experience, Education Level, and Satisfaction Guarantees.
1. **Gender**: It is believed that gender has a direct effect on the use of technology; males are expected to be very work focused than females. Similarly, as stated by Venkatesh et al. (2003), females are believed to be oscillated by others’ opinions; hence, females’ intention to use a system is affected by others.

2. **Internet Experience**: The experience in the novel UTAUT model will be substituted by Internet experience since numerous earlier studies, such as that by Agarwal and Prasad (1999), showed that Internet experience has an impact on an individual’s use or belief to use a specific system. Consumers with Internet exposure are more probable to use e-services, which are accessible online; consequently, this aspect should be considered to explain citizens’ adoption of e-services.

3. **Educational Level**: Level of education will be added to the study, as unlike stages of education result in different points of belief that consumers have (AlAwadhi & Morris, 2009).

4. **Satisfaction Guarantee**: AlAwadhi and Morris (2009) claimed that e-services are extremely voluntary, so the Voluntariness of Use is detached from the unique UTAUT model, and a new mediator Satisfaction Guarantees, such as money-back guarantee and post-purchase guarantee, which can affect adoption behavior, will be incorporated.

As this research is concentrating on older consumers of China and will be pointing to particular age clusters, the original AGE mediator from the UTAUT model will be omitted. The modified UTAUT model, which is projected for the assessment of customers’ approach toward the usability of e-facilitation, is shown in Figure 1.

5 **Hypothesis Development**

From the above-mentioned amended UTAUT model, the subsequent hypotheses have been developed:

**H1**: The greater the amount of Trust, the higher will be the chances of Behavioral Intention to adopt e-services, and this link will be moderated by Gender, Education Level, Satisfaction Guarantees, and Internet Experience.
H2: The greater the amount of Performance Expectancy, the higher will be the chances of Behavioral Intention to adopt e-services, and this link will be moderated by Gender, Education Level, Satisfaction Guarantees, and Internet Experience.

H3: The greater the amount of Effort Expectancy, the higher will be the chances of Behavioral Intention to adopt e-services, and this link will be moderated by Gender, Education Level, and Internet Experience.

H4: The greater the amount of Social Influence, the higher will be the chances of Behavioral Intention to adopt e-services, and this link will be moderated by Gender and Education Level.

H5: The greater the amount of Perceived Risk, the lower will be the chances of Behavioral Intention to adopt e-services, and this link will be moderated by Gender, Education Level, Satisfaction Guarantees, and Internet Experience.

H6: The greater the amount of Facilitating Conditions, the higher will be the chances of Use Behavior to adopt e-services, and this link will be moderated by Gender, Education Level, Satisfaction Guarantees, and Internet Experience.

H7: The greater the amount of Behavioral Intention, the higher will be the chances of Use Behavior to adopt e-services.

H8: The greater the amount of Use Behavior to adopt e-services, the higher will be the chances of increased QoL.

6 Methodology

This research will use quantitative techniques based on questionnaire analysis. The quantitative technique will permit the investigator to examine the association among the variables admitted in the model and thus allow the researcher to deliver proofs to approve or refute the hypotheses (Carter & Bélanger, 2005). The population of this research contains all Internet users with or without knowledge and skills in using e-Services in China. China has shown a great upward trend in developing e-services usage. However, because of deficiency of availability of details about Internet surfers, difficulty will be faced in forming and selecting sample data and sample clusters. This problem will be addressed by using volunteer Internet users for answering an online questionnaire, which will be e-mailed to them by an online survey, such as SurveyMonkey, during the data collection stage.

According to Rehman et al. (2012), “every independent variable is predicted to have ten data records”, i.e., if there are ten variables in a study, then there will be 100 respondents that will be used for data collection, but the mockup magnitude will be directed to 500 respondents because “sample sizes larger than 30 and fewer than 500 are suitable for most research data collection” (Sekaran & Bougie, 2010). In an attempt to reduce the prejudice in the data collection stage, the data triangulation method will be implemented. Furthermore, for the online analysis, printed questionnaires will also be used to gather data individually. The respondents for these printed questionnaires will be Internet users representing all districts in China. Respondents will be government and public sector, as well as private sector, employees, university and college students, and so on, who will be selected based on convenience sampling.

7 Conclusion and Significance of the Study

This study is significant for the reason that it tries to contribute to the field of e-service adoption among older consumers in China and its influence on their life. Primarily, this study provides a contribution to the body of knowledge that presently exists in the area of customers’ approach toward the usability of e-facilitation. Moreover, this research is projected to frame a model that can guide about e-service adoption in China. Moreover, it will cover the gap that is present between the struggle to diffuse e-services and citizens’ beliefs for the adoption and use of these services.

Each and every section on this experiment will be handled with abundant carefulness; however, it is comprehensible that there will be certain restrictions. The research is limited to the analysis of Internet users in China. Computer-educated individuals could have been included for this study and also all families with computer access could also have been added. Because of the absence of particulars about such computer-educated individuals or such families that own computers, the study will not be able to add them. Moreover, many of these individuals are probable to become Internet users within a short interval of time. This short span of time and the collection of data once make this study cross sectional. If it was longitudinal, then more factors inducing citizens’ adoption could have been detected and the implications would have been more wide ranging.

Furthermore, it will streamline the main barriers that might stimulate e-service adoption in China. Many articles are available that have examined e-service adoption,
such as Technology Acceptance Model (TAM), Diffusion of Innovation (DOI), and so on, but this research, particularly, will apply a broader model, namely, the UTAUT model, which incorporates eight such technology adoption studies to clarify the adoption elements so that policymakers would better recognize the elements of e-service adoption and arrive at better conclusions accordingly.

**Bibliography**


Factors Influencing User Behavior Intention to Use Mobile Library Application: A Theoretical and Empirical Research based on Grounded Theory

Introduction

With the development of mobile Internet technology, the scale of Internet mobile terminal is expanding rapidly. In April 2020, China’s Internet users reached 904 million, with an Internet penetration rate of 64.5%, mobile Internet had a dominant position in Internet industry, and the type and quantity of mobile application continues to rise (China Internet Network Information Center, 2020). As one of Internet mobile terminal, mobile library application is the extension of library mobile service and an innovative library service mode. Mobile library mainly includes three service mode, such as the SMS mode, the mobile web mode, and the mobile device (tablet PCs and smartphones) application mode. Mobile library application was classified as the mobile device application mode (Parsons, 2010; Liang & Guo, 2013; Vassilakaki, 2014). Mobile library application can integrate the basic functions of the library completely and give full play to the characteristics of mobile Internet to realize intelligent services, such as location-based service (LBS) (Junglas & Watson, 2008), book lending, and bibliographic barcode scanning. Mobile library application can help library to better meet the user needs in the mobile Internet context (Zhao & Gao, 2015). To meet the needs that user can have access to e-journals, e-books, books subscription information, library opening hours, and other e-resources and information at any time and in any place (Parsons, 2010; Rafique, Anwer, Sharmim, Minaei-Bidgoli, Gureshi & Shamshirband, 2018), many libraries independently or with mobile application developer have constructed mobile library application (Lai, Chiu, Huang, Chen & Huang, 2014). There is no doubt that mobile library application is the development direction of library mobile service (ChanLin & Hung, 2016; Fung, Chiu, Ko, Ho & Lo, 2016; Torres-Pérez, Mendez Rodriguez & Orduna-Malea, 2016).
However, the utilization rate of mobile library application is not high; the actual application value has not been reflected (Wei, Chang & Cheng, 2015); and the degree of user acceptance and adoption of mobile library application is relatively low at present (Hu & Zhang, 2016). Thus, it is systematical studying user behavior intention to use mobile library application and figuring out factors influencing user’s usage behavior intention that have an important significance for library to improve mobile service quality (Rafique, Almagrabi, Shamim, Anwar, & Bashir, 2020).

With the development of mobile Internet technology and mobile device form, mobile information services have been developing rapidly. Libraries, as professional institutions providing information services, have carried out the construction of mobile libraries to conform to the trend of The Times. Users can break through the limitations of time and space, and enjoy library services at any time and any place through mobile devices. Users’ behavioral intention is an important measure of users’ adoption, continuous use, and noncontinuous use. To develop and perfect mobile library, it is necessary to study the influencing factors of mobile library user behavior (Liu, Su, Akram & Abrar, 2020). However, through the literature review in Section 2, it can be seen that current researches about user behavior of mobile library application are insufficient and lack systematic construction process of model, whose selection of latent variables of model and generalization of conclusions need to be further verified.

The study reported in this article will improve current user’s usage behavior intention of mobile library application by exploring the following research questions: RQ1. What are the main factors influencing user behavior intention to use mobile library application? RQ2. How do these factors influence user behavior intention to use mobile library application?

To answer the research questions, this article first adopted in-depth interview, focus group interview, and other methods to obtain user interview results. Second, the interview results were coded using grounded theory, and a theoretical model of factors influencing user behavior intention to use mobile library application was constructed. Third, the survey questionnaire was designed to collect sample data, and an empirical analysis was made to validate the theoretical model using partial least squares. Finally, factors influencing user behavior intention to use mobile library application were discussed, and some recommendations were put forward to improve and perfect the mobile service quality of library.

2 Literature Review

2.1 Mobile Library Service and Mobile Library Application

With the unceasing advancement of mobile technologies, the mobile library service has changed dramatically. Library/book on wheel is the oldest mobile library service, and also known as mobile bus library, mobile train library, and bookmobile. It transfers some library resources from library to places where there is no library, such as rural districts or remote areas, and provides library services. Later, short message service (SMS) was popular with library user and widely used to provide user with basic library services, such as the content alert service (Anbu & Mavuso, 2012), the library catalogue service (Goh & Liew, 2009), the reference service (Herman, 2007), and so forth.

Later, the further development of mobile technologies makes it possible that library user can have access to library information resources through using a mobile device and connecting a wireless network, so mobile web service of library emerged and many libraries started designing the service. For example, the Oriental Institute of Technology Library (OITL) in Taiwan provided user with due-day reminder service, renewal-request service, and others. The due-day reminder and renewal-request services were the first two mobile web services (Wang, Ke & Lu, 2012). The OITL user can receive e-mail and SMS notification a day before the borrowed items were due, and extend the due date through access to library website or using SMS. Most member institutions of the Association of Universities and Colleges of Canada (AUCC) created a mobile library website, which referred to any web presence accessible through the browser of mobile device and has been tailored in some way to the mobile context (Canuel & Crichton, 2011). The mobile library website provided user with library catalogue, library hours, contact information, account access, library location, library news, main library web site link, study rooms, and other services. The library catalogue existed in almost all mobile library website of member institutions of AUCC. For the purpose of providing users with a convenient mobile interface to obtain library resources and information, Fu Jen Catholic University library also created a mobile library website, where the most frequently used services included online search, library news, personal circulation status, book recommendations, seats in the learning commons, feedback, opening hours, and contact information (ChanLin & Hung, 2016).
At present, the mobile device application is widely introduced to various types of library, which includes mobile library application, WeChat (a social software developed by Tencent for instant messaging) library (Wei & Yang, 2017), and other social media applications (Zhu, 2016). Mobile library application is the main mobile library service in many libraries. To better design and apply mobile library application, and enhance user awareness about it, a few researchers have explored the applying status, the emerging technology applying, the system and platform designing of mobile library application, and so on (Pianos, 2012; Schnell, 2012; Miller, Vogh & Jennings, 2013; Wei, Chang & Cheng, 2015; Pu, Chiu, Chen & Huang, 2015; Ansari & Tripathi, 2017; Ghuloum & Allamki, 2017; Kerr & Pennington, 2018). For example, Pianos (2012) developed a specialized mobile library application to cater to the needs of researchers in business and economics. The application had four basic functions, such as search, favorites, libraries, and options. The function of libraries was LBS, by which library user can find the nearby library and the library they want to go. Wang, Yang, Li and Wang (2018) analyzed the factors of the mobile library users’ attitude and behavioral intention from the perspective of information ecology. Wei, Chang and Cheng(2015) conducted a usability testing of mobile library application via pretest questionnaires, accomplishing tasks, and posttest surveys. The experiment results indicated that mobile library application was effective but the efficiency of it required improvement. The services provided by it were appealing and useful to users. Pu, Chiu, Chen and Huang (2015) designed and developed a mobile library application using android open source code. The application can be installed on tablet PCs, smartphones, and other mobile devices, and then library user can run the application and quickly search for book collection, read latest news of the library, and check their borrowing record.

2.2 Related Research in the User Behavior of Mobile Library Application

In recent years, drawn from various disciplines, a variety of models have been employed to explain user behavior about information systems (Yoon, 2016), such as technology acceptance model (TAM) (Davis, 1989), unified theory of acceptance and use of technology (UTAUT) (Venkatesh, Morris, Davis & Davis, 2003), and ECM-ISC (Bhattacherjee, 2001), and others. TAM was constructed to pursue better measures for predicting and explaining user accepting information systems or technologies (IST) by Davis (1989). Then, Venkatesh, Morris, Davis and Davis (2003) proposed the UTAUT by integrating eight models, including the theory of reasoned action, the technology acceptance model, the motivational model, the theory of planned behavior, a model combining the technology acceptance model and the theory of planned behavior, the model of PC utilization, the innovation diffusion theory, and the social cognitive theory. TAM and UTAUT were widely used to study the user acceptance or adoption behavior of IST (Moon & Kim, 2001; Gefen, KarafHanna & Straud, 2003; Gruzd, Staves & Wilk, 2012). ECM-ISC was adapted from expectation confirmation theory (ECT) (Oliver, 1980) by Bhattacherjee (2001) and then adopted to study the factors influencing user intention to continue using IST.

In the library and information field, these models (i.e., TAM, UTAUT, ECM-ISC, ECM, or ECT) were applied to explore the user behavior of library emerging IST. For example, Kapoor, Dwivedi, Piercy, Lai and Weerkakody, (2014) explored the factors influencing the use of RFID systems in the library context by empirically testing the relevant constructs from the extended TAM. Chen, Chang, Kao and Huang(2016) proposed a new model that incorporated information system success model into TAM, and analyzed the factors influencing the user’s usage behavior of digital library services in the National Central Library of Taiwan. Yuvaraj (2016) studied the adoption of social media technologies in the recruitment and selection process of library professionals and faculty members by constructing an UTAUT model. Cheng (2014) proposed the model of user continuance intention of digital library based on TAM, ECM, and so on. Joo and Choi (2016) analyzed the factors affecting students’ continuance intention to use online library resources in the context of academic libraries through using an extended ECT.

A few researchers also proposed some user behavior models of mobile library application based on these models. Yoon (2016) constructed a conceptual structural model for user acceptance of mobile library application using TAM and analyzed factors influencing user acceptance of mobile library application in academic libraries. The experimental results showed that perceived usefulness, interactivity, and ease of use had significant effects on user attitude and intention to use mobile library application. Zhao, Deng and Zhou (2015) developed a model of user’s continued usage behavior about mobile library application based on ECM-ISC and theory of information system success. The results demonstrated that perceived usefulness, expectation confirmation, and satisfaction had significant effects on the continuance usage of mobile library application. Chang (2013)
integrated UTAUT with task technology fit to explain user behavior intention to use mobile library application in university libraries. The results showed that performance expectancy, effort expectancy, social influence, and facilitating conditions influenced user behavior intention to use mobile library application. Furthermore, Hu and Zhang (2016) proposed an integrated model to examine behavior intention of Chinese university students toward mobile library application and found that user attitude is the main predictor of behavior intention, and the quality of mobile library application determined students' perception of usefulness.

3 Research Design

3.1 Methodology

To systematically study user behavior intention to use mobile library application, this study mainly adopted structural equation modeling and grounded theory.

3.1.1 The Grounded Theory

The grounded theory is a methodology jointly developed by Glaser and Strauss (1967), and commonly used to systematically develop theoretical understandings from rigorous qualitative data collection and analysis (Parker, 2018). The process of grounded theory is different from the deductive logic commonly used, in which the meaning and knowledge are previously established (Roman, Osinski & Erdmann, 2017). The grounded theory adopts a set of procedures based on codification processes and theoretical sampling (Roman, Osinski & Erdmann, 2017) to construct a model or theory. User's usage behavior intention of mobile library application is an intricate psychological behavior, and the common deductive logic may not fully and systematically analyze user behavior intention to use mobile library application. So the research used grounded theory to construct a theoretical model of factors influencing user behavior intention to use mobile library application. First, the interview results were coded via means of open coding, axial coding, and selective coding, so as to obtain influential factors of user's usage behavior intention about mobile library application. Second, a theoretical model was constructed about factors influencing user behavior intention to use mobile library application. Finally, the theoretical model was tested for saturation.

3.1.2 The Structural Equation Modeling

Structural equation modeling are well-recognized as significant and useful in analyzing intercorrelations among observed and latent variables (Song & Lee, 2004). Therefore, before the structural equation modeling was applied, a theoretical model based on the hypothesized relationship as described earlier was constructed using the grounded theory (Yanuar, Ibrahim & Jemain, 2010). Referring to variables in the theoretical model and related literature, and combining with basic characteristics of mobile library application, a survey questionnaire was designed to obtain sample data coming from college teachers and students using paper questionnaire and network questionnaire. Then an empirical analysis was made to discuss the factors influencing user behavior intention to use mobile library application using the structural equation modeling. Structural equation modeling is referred to as causal modeling, causal analysis, simultaneous equation modeling, analysis of covariance structures, path analysis, or confirmatory factor analysis (Ullman & Bentler, 2003), and this article selected the latter two to assess intercorrelations among the latent variables of the theoretical model.

3.2 Data Collection

Data collection includes two steps: (1) the acquisition of interview data for the model construction in quantitative analysis based on the grounded theory and (2) the acquisition of questionnaire data for empirical analysis in quantitative analysis based on the structural equation modeling.

3.2.1 Interview Data

In view of that the grounded theory requires respondents who have certain cultural level and cognitive ability, this interview took college teachers and students as objects. Furthermore, the study required that the college teachers and students were willing to accept this interview. Thirty-six college teachers and students were interviewed, including 8 teachers and 28 students from Wuhan University, Huazhong University of Science and Technology, Nanjing University, Fudan University, and Shanghai Jiao Tong University. The interview was conducted according to semi-structured interview outline, which mainly sets up guiding question and is launched around the research goal with a view to fully understand
Factors Influencing User Behavior Intention to Use Mobile Library Application

The interview was recorded by recording pen and word software, so as to save the interview results completely. It was divided into individual in-depth interviews and focus group interviews. The individual in-depth interviews of 20 interviewees were an average of 1 h per person, and the focus group interviews of 16 interviewees were conducted four times, which were an average of 2 h per time. After the interview, 120,000 words of interview results were obtained via sorting audio data and text records. The interview results of 28 college teachers and students were randomly selected to make open coding, axial coding, and selective coding, including 16 individual in-depth interview results and 3 focus group interview results. The rest of the eight interview results were used to make a theoretical saturation test, including four individual in-depth interview results and one focus group interview result.

3.2.2 Questionnaire Data

3.2.2.1 Questionnaire Design
Five-point Likert scale is adopted in the questionnaire, which was designed according to the basic hypotheses in Section 5. In the process, the questionnaire is divided into the following parts:
(1) Guide language. It expounds some professional terms that the study purpose and questionnaire involved.
(2) Basic information about the users. It includes gender, identity (student or teacher), and degree of users.
(3) Variable measurement items. It includes accessibility, relevance, system help, screen design, navigation, term, performance expectancy, domain knowledge, social influence, and user’s usage behavior intention. Five specialists and scholars in the library and information field were invited to discuss, revise, and give feedback on the questionnaire to ensure the contents comply with the research purpose (Chiu, Chao, Kao, Pu, & Huang, 2016). Then, the questionnaires were sent to 16 college teachers and students to conduct a small sample of pre-research, and the content, format, items, order, and clarity of the questionnaire were again revised and improved. After revision, 34 questions of the variable measurement items remained in the questionnaire, as were shown in Appendix.

3.2.2.2 Data Sources
Using web and paper questionnaires, the questionnaires were sent to college teachers and students of Wuhan University, Huazhong University of Science and Technology, Nanjing University, Fudan University, and Shanghai Jiao Tong University. A total of 478 questionnaires were collected. After filtering the questionnaire data, 421 valid questionnaires were obtained with a rate of valid return of 87.08%, 183 of which came from paper questionnaires and 238 came from web questionnaires. The main filter criteria are: (1) if all items or many items in a row are the same score, it is invalid and (2) if there are many items not scored, it is invalid. Through descriptive statistical analysis of sample characteristics, it were found that male accounted for 55.82%, female accounted for 44.18%, students accounted for 87.89%, teachers accounted for 12.11%, bachelor’s accounted for 71.97%, master’s accounted for 21.14%, and doctorate accounted for 6.89%. Basic information of the respondents is shown in Table 1.

<table>
<thead>
<tr>
<th>Information</th>
<th>Item</th>
<th>Frequency</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Male</td>
<td>235</td>
<td>55.82</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>186</td>
<td>44.18</td>
</tr>
<tr>
<td>Identity</td>
<td>Student</td>
<td>370</td>
<td>87.89</td>
</tr>
<tr>
<td></td>
<td>Teacher</td>
<td>51</td>
<td>12.11</td>
</tr>
<tr>
<td>Education</td>
<td>Bachelor’s</td>
<td>303</td>
<td>71.97</td>
</tr>
<tr>
<td></td>
<td>Master’s</td>
<td>89</td>
<td>21.14</td>
</tr>
<tr>
<td></td>
<td>Doctorate</td>
<td>29</td>
<td>6.89</td>
</tr>
</tbody>
</table>
4 Model Construction

The model construction consisted of category extraction, theoretical construction, and saturation test. Especially, the category extraction was divided into open coding, axial coding, and selective coding. The coding process was in accordance with predetermined standards, and the standards mainly included the following three: (1) the original data statements were extracted according to the interview results about mobile library application, and other interview results would be filtered, (2) the subcategories and main categories were mainly coded on the basis of the latent variables in the TAM and UTAUT, which have defined these latent variables in detail (Davis, 1989; Lingaard, 1994; Agarwal & Prasad, 1999; Thong, Hong & Tam, 2002; Venkatesh, Morris, Davis & Davis, 2003; Gefen, Karahanna & Straud, 2003; Chang, 2013; Yoon, 2016; Hu & Zhang, 2016), and (3) five specialists and scholars in the library and information field were invited to discuss, analyze, define, and revise the subcategories and main categories to ensure the accuracy and reliability of coding process. After the category extraction, the subcategory, main category, core category, influencing user to use mobile library application and the link between categories, would be gotten. Then a theoretical model of factors influencing user behavior intention to use mobile library application was constructed in the theoretical construction and saturation test process.

4.1 Open Coding

In this study, line by line and verbatim analysis of interview results were conducted to select the original data statements and make it initialized. Then a total of 462 original data statements and the corresponding initial concept were obtained. Since the initial concepts are very complex and exist a certain degree of crossover, the original data statements with the same concept were further integrated and classified referring to the existing literature and the group discussion. Finally, 33 initial concepts and 9 subcategories were gained. Coding process, subcategories, and the original data statement are shown in Table 2.

4.2 Axial Coding

In view of the fact that the subcategories formed in open coding do exist intrinsic links, the nine subcategories were divided into three main categories (system feature, interface feature, and individual difference) according to the interrelationships and logical levels between subcategories. The main categories and their corresponding subcategories are shown in Table 3.

4.3 Selective Coding

Through the step-by-step coding, it was found that all three main categories focus on user behavior intention to use mobile library application. Thus, the user’s usage behavior intention is determined as core category, and the story lines in the core category are summarized as follows: the three main categories namely system feature, interface feature, and individual difference have a positive impact on user’s usage behavior intention; system feature and interface feature are the external driver of user’s usage behavior intention; and individual difference is the internal driver of user’s usage behavior intention. Typical relational structures of the main categories are shown in Table 4.

4.4 Theoretical Construction and Saturation Test

After three coding processes, this study pointed out three main categories, such as system feature, interface feature, and individual difference, which have a positive impact on user behavior intention to use mobile library application, and initially established a theoretical model of factors influencing user behavior intention to use mobile library application, as shown in Figure 1. The saturation test of grounded theory is applied to verify whether the interview results are in a state of saturation, and to comprehensively analyze user behavior intention to use mobile library application. Therefore, based on the rest of eight interview results, the saturation test was conducted on the theoretical model. The test results show that the categories from open coding, axial coding, and selective coding are very rich. There is no new category and connection appearing between categories and the theoretical model is saturated.
<table>
<thead>
<tr>
<th>Subcategory</th>
<th>Original data statement (initial concept)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accessibility</td>
<td>I have access to mobile library application in the office, at home, and at school and can also download journal articles and read e-books online. (Access to library resources via mobile library application)</td>
</tr>
<tr>
<td></td>
<td>When I use the mobile library application, I found that the application responds fast and does not get stuck at all. I am happy to use it. (Access speed)</td>
</tr>
<tr>
<td></td>
<td>Last time, I was unable to download the journal articles outside the school using the website, but I can download it using mobile library application. (Access advantage via mobile library application)</td>
</tr>
<tr>
<td>Correlation</td>
<td>I found that the information resources provided by mobile library application are all that I need and are very useful to me. (Demand-related)</td>
</tr>
<tr>
<td></td>
<td>I am very interested in information mobile library application recommends. (Related recommendation)</td>
</tr>
<tr>
<td></td>
<td>Mobile library application has professional information resources related to my research. (Related resource)</td>
</tr>
<tr>
<td></td>
<td>UI settings of mobile library application are the same as my usage habits. (Related habits)</td>
</tr>
<tr>
<td>System help</td>
<td>If I do not understand the operation of mobile library application, I can consult the staff in the feedback column. (Interaction)</td>
</tr>
<tr>
<td></td>
<td>Mobile library application have the introduction of basic information about the software, I feel very good and happy to use it. (Basic information)</td>
</tr>
<tr>
<td></td>
<td>Once we install a mobile library application, the basic functions of mobile library application will be introduced. (Introduction of first use)</td>
</tr>
<tr>
<td></td>
<td>The function, service of mobile library application are regularly updated and prompted. I think the service is very thoughtful. (Service prompt)</td>
</tr>
<tr>
<td>Screen design</td>
<td>I think UI of mobile library application was well designed, people are willing to use it. (Screen style)</td>
</tr>
<tr>
<td></td>
<td>I think UI of mobile library application is very similar with that of QQ and WeChat and I am willing to use it. (Reasonable user interface)</td>
</tr>
<tr>
<td></td>
<td>Touch sensitivity of mobile library application is very good, which gives people a better experience. (Screen sensitivity)</td>
</tr>
<tr>
<td>Navigation</td>
<td>I think navigation bar design of mobile library application is very reasonable, the resources have been classified according to different disciplines. (Resource classification)</td>
</tr>
<tr>
<td></td>
<td>We can design our own navigation bar according to our habits of using mobile library application. (Personalized navigation)</td>
</tr>
<tr>
<td></td>
<td>Home page of mobile library application set up: “bookshelf,” “audio books,” “newspaper,” “open class,” “academic resources,” and other navigation bars. (Conventional navigation)</td>
</tr>
<tr>
<td>Term</td>
<td>The relevant terms on mobile library application are set very well, making it easy to understand what is said. (term name)</td>
</tr>
<tr>
<td></td>
<td>In my search for journal articles, my first reaction was to have access “academic resources” and I found what I was looking for. (terminology)</td>
</tr>
<tr>
<td></td>
<td>I think term should be set, so users can understand the meaning of term and are very convenient to use. (Term meaning)</td>
</tr>
<tr>
<td>Performance</td>
<td>I think using mobile library application can help my study and work, and I often use it. (Helps to learn)</td>
</tr>
<tr>
<td>expectations</td>
<td>I think mobile library application can help me get information resources outside the school, such as journal articles, e-books, etc. (Usefulness)</td>
</tr>
<tr>
<td></td>
<td>Personally, it is a lot easier to use a mobile library application than not to use it. (Convenience)</td>
</tr>
<tr>
<td>Domain knowledge</td>
<td>I had used mobile library application and have some related knowledge about it, so I feel very easy when using mobile library application. (Related knowledge)</td>
</tr>
<tr>
<td></td>
<td>I think my operations are very clear when I search for some journal articles, which may be related to my knowledge. (Knowledge)</td>
</tr>
<tr>
<td></td>
<td>I think I can find out the resources I want from the mobile library application. (Related experience)</td>
</tr>
<tr>
<td></td>
<td>I found out that when I have used the mobile library more frequently and became more familiar with it, my using mobile library application was more comfortable. (Use experience)</td>
</tr>
</tbody>
</table>
In this article, quantitative research was carried out by proposing basic propositions and hypotheses. According to the model construction processes and the theoretical model, the following basic propositions and hypotheses were discussed.

Table 3
Categorization of Axial Coding

<table>
<thead>
<tr>
<th>Subcategory</th>
<th>Original data statement (initial concept)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social influence</td>
<td>My teacher recommended me to use mobile library application, so I use it. (Teacher influence)</td>
</tr>
<tr>
<td></td>
<td>Libraries have given lectures on mobile library application before, so I think mobile library application do help my learning. (Lecture effects)</td>
</tr>
<tr>
<td></td>
<td>There were some news about mobile library application in the library website of my university, so I use it. (Website impact)</td>
</tr>
<tr>
<td></td>
<td>My friends and classmates had used the mobile library application and recommended me to use it. (Friend influence)</td>
</tr>
</tbody>
</table>

Notes: Original data statements are on behalf of the original words of teachers and students interviewed, and the statement in parentheses is the initial concept extracted from the original data statements. Only the original data statement covering all the content was selected for each subcategories. QQ is a social software developed by Tencent for instant messaging.

Table 2
Categorization of Open Coding

<table>
<thead>
<tr>
<th>Main category</th>
<th>Corresponding subcategory</th>
<th>Category connotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>System feature</td>
<td>Accessibility</td>
<td>Accessibility is about access performance of mobile library application, whether it can be quickly accessed, and whether there is no load situation.</td>
</tr>
<tr>
<td></td>
<td>Correlation</td>
<td>Correlation is about that whether the services and contents of mobile library application are consistent with user’s individual needs.</td>
</tr>
<tr>
<td></td>
<td>System help</td>
<td>System help refers to the degree to which mobile library application provides user with the interaction, service prompt, basic information, and other services.</td>
</tr>
<tr>
<td>Interface feature</td>
<td>Screen design</td>
<td>Screen design is about that whether the screen style, interface size, sensitivity of mobile library application meet users.</td>
</tr>
<tr>
<td></td>
<td>Navigation</td>
<td>Navigation is about information classification design of mobile library application.</td>
</tr>
<tr>
<td></td>
<td>Term</td>
<td>Term is about the name of each button in mobile library application, which corresponds to the function of button.</td>
</tr>
<tr>
<td>Individual difference</td>
<td>Performance expectation</td>
<td>Performance expectation refers to the degree to which individuals think that the use of information technology can help them gain a sense of effectiveness in job performance.</td>
</tr>
<tr>
<td></td>
<td>Domain knowledge</td>
<td>Domain knowledge refers to the degree to which individuals have the cognitive level of basic concepts and professional knowledge about mobile library application.</td>
</tr>
<tr>
<td></td>
<td>Social influence</td>
<td>Social influence refers to the degree to which individuals think they should use mobile library application.</td>
</tr>
</tbody>
</table>

5 Model Validation

5.1 Model Elaboration and Hypothesis

In this article, quantitative research was carried out by proposing basic propositions and hypotheses. According to the model construction processes and the theoretical model, the following basic propositions and hypotheses were discussed.

Basic proposition 1

System feature positively affects user behavior intention to use mobile library application, and it is the external driver of user’s usage behavior intention which mainly exerts the influence through the following three paths:
1. Accessibility of mobile library application positively affects user behavior intention to use mobile library application. The stronger the access performance of mobile library application is, the stronger the user’s usage behavior intention is.
Factors Influencing User Behavior Intention to Use Mobile Library Application

(2) Correlation between mobile library application and users positively affects user behavior intention to use mobile library application. The more relevant service and content provided by mobile library application is related to the user’s needs, the stronger the user’s usage behavior intention is.

(3) System help of mobile library application positively affects user behavior intention to use mobile library application. The better the system help of mobile library application is designed, the stronger the user’s usage behavior intention is.

Proposition 1. System feature of mobile library application is positively correlated with user’s usage behavior intention.

H1a. Accessibility is positively related to the user behavior intention to use mobile library application.

H1b. Correlation is positively related to the user behavior intention to use mobile library application.

H1c. System help is positively related to the user behavior intention to use mobile library application.

Basic proposition 2

Interface feature positively affects user behavior intention to use mobile library application and it is the external driver of user’s usage behavior intention which mainly exerts the influence through the following three paths:

(1) Screen design of mobile library application positively affects user behavior intention to use mobile library application. The better the screen design of mobile library application is, the stronger the user’s usage behavior intention is.

(2) Navigation of mobile library application positively affects user behavior intention to use mobile library application. The better the navigation of mobile library application is, the stronger the user’s usage behavior intention is.

(3) Term of mobile library application positively affects user behavior intention to use mobile library application. The more the term design of mobile library application is reasonable, the stronger the user’s usage behavior intention is.

Table 4
The Typical Relational Structure of the Main Categories

<table>
<thead>
<tr>
<th>Typical relationship structure</th>
<th>Connotation of relational structure</th>
<th>Representative statement (extracted from the typical structure)</th>
</tr>
</thead>
<tbody>
<tr>
<td>System feature → user’s usage behavior intention</td>
<td>System feature is the external driver of user’s usage behavior intention, which affects user’s usage behavior intention by self-performance of mobile library application.</td>
<td>When I use the mobile library application, I find that my system responds fast and does not get stuck at all, I am happy to use it. (Accessibility → user’s usage behavior intention) I find that the information resources provided by mobile library application are all I need and useful to me, I will use it. (Correlation → user’s usage behavior intention) Mobile library application have the introduction of basic information about the software, I feel very good and happy to use it. (System help → user’s usage behavior intention)</td>
</tr>
<tr>
<td>Interface feature → user’s usage behavior intention</td>
<td>Interface feature is the external driver of user’s usage behavior intention, which affects user’s usage behavior intention by self-performance of mobile library application.</td>
<td>I think user interface style of mobile library application of our school is perfect, people want to use it. (Screen design → user’s usage behavior intention) We can design our own navigation bar according to our habits of using mobile library application. (Navigation → user’s usage behavior intention) I think term should be set, so users can understand the meaning of term and are very convenient to use. (Term → user’s usage behavior intention)</td>
</tr>
<tr>
<td>Individual difference → user’s usage behavior intention</td>
<td>Individual difference is the external driver of user’s usage behavior intention, which affects user’s usage behavior intention by user perception.</td>
<td>I think using mobile library application can help my study and work, and I often use it. (Performance expectation → user’s usage behavior intention) I find out that when I have used the mobile library frequently and became familiar with it, my using mobile library application is more comfortable. (Domain knowledge → user’s usage behavior intention) Many people around me have used the mobile library application and they say it works well, so I install it later. (Social influence → user’s usage behavior intention)</td>
</tr>
</tbody>
</table>
Proposition 2. Interface feature of mobile library application is positively correlated with user’s usage behavior intention.
H2a. Screen design is positively related to the user behavior intention to use mobile library application.
H2b. Navigation is positively related to the user behavior intention to use mobile library application.
H2c. Term is positively related to the user behavior intention to use mobile library application.

Basic proposition 3

Individual difference positively affects user behavior intention to use mobile library application and it is the internal driver of user’s usage behavior intention which mainly exerts the influence through the following three paths:
(1) Performance expectancy of mobile library application positively affects user behavior intention to use mobile library application. The higher the user’s performance expectation to mobile library application is, the stronger the user’s usage behavior intention is.
(2) Domain knowledge of mobile library application positively affects user behavior intention to use mobile library application. The more the user’s domain knowledge is abundant, the stronger the user’s usage behavior intention is.
(3) Social influence of mobile library application positively affects user behavior intention to use mobile library application. The larger the number of people using mobile library application around user is, the stronger the user’s usage behavior intention is.

Proposition 3. Individual difference of user is positively correlated with user’s usage behavior intention.
H3a. Performance expectation is positively related to the user behavior intention to use mobile library application.
H3b. Domain knowledge is positively related to the user behavior intention to use mobile library application.
H3c. Social impact is positively related to the user behavior intention to use mobile library application.

5.2 Data Analysis

5.2.1 Reliability and Validity Test

Reliability is a measure of the consistency and stability of questionnaire results. In this study, Cronbach’s a coefficient was used to test the reliability of formal scales. In the evaluation of internal reliability, Cronbach’s a is greater than 0.7, indicating that the data has considerable credibility (Nunally, 1978). The results shown in Table 5 manifested that the Cronbach’s a coefficients for all variables are greater than 0.8, so the sample data selected in this study had good internal consistency.

Validity refers to the fact that the items can indeed measure the variables and have validity, including the validity of content and the validity of construct. Questionnaire items were designed via adapting the relevant literature, so the questionnaire has a good content validity. In this study, exploratory factor analysis was used to test the validity of formal scales. The results shown in Table 5 manifested that all KMO values are greater than 0.7; all p values of Bartlett’s spherical test are
Factors Influencing User Behavior Intention to Use Mobile Library Application

less than 0.001; the factor load of each variable is greater than 0.5; and all values of explanations variance is greater than 60%, so the questionnaire has good validity.

5.2.2 Structural Model Validation

By the test of validity and reliability, it was found that the questionnaire had good reliability and validity and could be further analyzed. This study used partial least square method to verify the theoretical model of factors influencing user behavior intention to use mobile library application and selected Amos 21 software to analyze the path coefficient of the theoretical model. The data of 425 questionnaires were imported into Amos 21 and processed to obtain path diagram of structural equation modeling, as shown in Figure 2.

According to verification results of the structural equation modeling, the impact of all factors influencing user behavior intention to use mobile library application has reached a significant level as shown in Table 6.

In system feature, the path coefficient of accessibility to user’s usage behavior intention is 0.304; the parameter estimation value is \( p < 0.001 \); and accessibility positively affects user behavior intention to use mobile library application. It can be seen that the better the access performance of mobile library application is, the more the user is willing to use it. The path coefficient of relevance to user’s usage behavior intention is 0.213; the parameter estimation value is \( p < 0.01 \); and relevance positively affects user behavior intention to use mobile library application.

In interface feature, the path coefficient of screen design to user’s usage behavior intention is 0.296; the parameter estimation value is \( p < 0.01 \); and screen design positively affects user behavior intention to use mobile library application. It can be seen that the better the sensitivity, screen style, and other screen design of mobile library application are, the more the user is willing to use it. The path coefficient of navigation to user’s usage behavior intention is 0.312; the parameter estimation value is \( p < 0.01 \); and navigation positively affects user behavior intention to use mobile library application. It can be seen that the better the navigation design of mobile library application is, the more the user is willing to use it.

In individual difference, the path coefficient of performance expectation to user’s usage behavior intention is 0.356; the parameter estimation value is \( p < 0.01 \); and term positively affects user behavior intention to use mobile library application. It can be seen that the higher the similarity between the services provided by mobile library application and the user needs is, the more the user is willing to use it.

Table 5
Reliability and Convergence Validity Analysis

<table>
<thead>
<tr>
<th>Variables</th>
<th>Items</th>
<th>Cronbach’s a</th>
<th>KMO</th>
<th>Bartlett’s Cartesian</th>
<th>Bartlett’s sphere test</th>
<th>Factor loading</th>
<th>Explanation variance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accessibility</td>
<td>3</td>
<td>0.809</td>
<td>0.771</td>
<td>633.699</td>
<td>0.00</td>
<td>0.784–0.895</td>
<td>70.546</td>
</tr>
<tr>
<td>Correlation</td>
<td>4</td>
<td>0.853</td>
<td>0.810</td>
<td>753.543</td>
<td>0.00</td>
<td>0.851–0.869</td>
<td>73.620</td>
</tr>
<tr>
<td>System help</td>
<td>4</td>
<td>0.804</td>
<td>0.792</td>
<td>525.235</td>
<td>0.00</td>
<td>0.748–0.826</td>
<td>69.860</td>
</tr>
<tr>
<td>Screen design</td>
<td>3</td>
<td>0.832</td>
<td>0.728</td>
<td>674.970</td>
<td>0.00</td>
<td>0.841–0.911</td>
<td>75.872</td>
</tr>
<tr>
<td>Navigation</td>
<td>3</td>
<td>0.864</td>
<td>0.731</td>
<td>677.968</td>
<td>0.00</td>
<td>0.869–0.902</td>
<td>79.715</td>
</tr>
<tr>
<td>Term</td>
<td>3</td>
<td>0.835</td>
<td>0.732</td>
<td>605.518</td>
<td>0.00</td>
<td>0.843–0.884</td>
<td>76.643</td>
</tr>
<tr>
<td>Performance expectation</td>
<td>4</td>
<td>0.917</td>
<td>0.719</td>
<td>505.846</td>
<td>0.00</td>
<td>0.874–0.937</td>
<td>71.401</td>
</tr>
<tr>
<td>Domain knowledge</td>
<td>3</td>
<td>0.823</td>
<td>0.833</td>
<td>1,248.509</td>
<td>0.00</td>
<td>0.855–0.864</td>
<td>76.342</td>
</tr>
<tr>
<td>Social influence</td>
<td>4</td>
<td>0.874</td>
<td>0.835</td>
<td>835.507</td>
<td>0.00</td>
<td>0.835–0.871</td>
<td>75.280</td>
</tr>
<tr>
<td>User’s usage behavior</td>
<td>3</td>
<td>0.864</td>
<td>0.730</td>
<td>606.381</td>
<td>0.00</td>
<td>0.864–0.902</td>
<td>76.712</td>
</tr>
</tbody>
</table>
and performance expectation positively affects user behavior intention to use mobile library application. It can be seen that the higher the user’s performance expectation is, the more the user is willing to use it. The path coefficient of domain knowledge to user’s usage behavior intention is 0.251; the parameter estimation value is $p < 0.01$; and domain knowledge positively affects user behavior intention to use mobile library application. It can be seen that the richer the user’s domain knowledge is, the more the user is willing to use it. The path coefficient of social influence to user’s usage behavior intention is 0.367; the parameter estimation value is $p < 0.01$; and social influence positively affects user behavior intention to use mobile library application. It can be seen that the larger the number of people using mobile library application around user is, the more the user is willing to use mobile library application.

In summary, system feature and interface feature are the external driver of user usage behavior intention, and individual difference is the internal driver of user usage behavior intention. This result is consistent with conclusions of the grounded theory, so the theoretical model of factors influencing user behavior intention to use mobile library application was verified.

### 6 Conclusions and Implications

This article proposed the theoretical model of factors influencing user behavior intention to use mobile library application based on grounded theory, and explored the factors influencing user behavior intention to use mobile library application. The results showed that system feature (i.e., accessibility, relevance, and system help), interface feature (i.e., screen design, navigation, and term), and individual difference (i.e., performance expectancy, social influence, and domain knowledge) were three main categories influencing user behavior intention to use mobile library application. The user behavior intention to use mobile library application is mainly influenced by system feature, interface feature, and individual difference. Moreover, system feature and interface feature are the external drivers of user usage behavior intention, and individual difference is the internal drivers of user usage behavior intention.

According to the above conclusions, mobile library application may be built and improved from three aspects such as system feature, interface feature, and individual difference, so as to enhance the willingness of user to use mobile library application.

In terms of system feature:

1. The library can enhance accessibility of mobile library application, such as enabling user to access and acquire library resources ubiquitously, and enhance access speed and user experience of mobile library application.

2. The library can enhance correlation between mobile library application and users. For example, librarians can analyze the needs of different user to library resources and provide them with relevant resources. Chief librarian can arrange some programmers to optimize recommendation service in mobile library application according to the user’s usage habits.
The library can perfect system help services of mobile library application. For example, librarians or other staff can set system help column in mobile library application to help new and old user make better use of the app. Facebook, Twitter, and other social media can be set in the feedback column of mobile library application for user interaction.

In terms of interface feature:

1. The library can improve screen design of mobile library application, such as designing the interface style of mobile library application according to the cultural background of library and the design style of library official website, and combine with the interface layout of social media application (i.e., Facebook, Twitter, and others) to design mobile library application UI layout.

2. The library can enhance navigation of mobile library application, such as classifying library resources according to different classification rules and designing corresponding navigation bars, and set up personalized sections, so that the user can design navigation bars according to his preference.

3. The library should improve term design of mobile library application. For example, different mobile library applications have different terms about the same content, so most used term should be selected to facilitate user’s understanding. The library can also add the introduction of term, such as setting the meaning of various terms in the system help column of mobile library application.

In terms of individual difference:

1. The library can enhance user’s performance expectations, such as introducing the influence of mobile library application to learning and working through lectures and other promotional methods, and user’s perception of usefulness about mobile library application.

2. The library can enhance domain knowledge of user, such as promoting mobile library application through social platforms such as Facebook and Twitter, and promote user’s understanding to mobile library application, such as carrying out lectures about mobile library application in the library.

3. The library can propagandize mobile library application to the public. For example, incentives, promotions, and other means can be used to encourage mobile library application user to recommend it, so as to influence more users to use mobile library application. Facebook, Twitter, and other social media can be applied to propagandize mobile library application, so as to enhance the popularity of mobile library application.

There are still many limitations in this study. The sample selection is mainly concentrated on students from five schools. The relationships between dependent variable and independent variable were discussed, but all kinds of complex correlations among the factors cannot be expressed. Limited by time, finance, energy, and other factors, this study did not discuss the difference of system feature, interface feature, and individual difference.
difference influencing user usage behavior intention from the perspective of weight. In the future, we will expand the diversity of the research sample, explore to find deep relationships between main categories, and start from the importance of system characteristics, interface characteristics, and individual differences to users’ behavioral intention to use mobile library applications so as to guide libraries to improve mobile library applications and enrich the user behavior research of mobile library applications.

References


## Appendix

### Option Design of Variable Measurement Items

<table>
<thead>
<tr>
<th>Factor</th>
<th>Variable</th>
<th>Content</th>
</tr>
</thead>
</table>
| System feature          | Accessibility           | H1a1: I can easily access information resources in mobile library application anytime.                                                                                           
|                         |                         | H1a2: The information resources in mobile library application are easily accessible.                                                                                                                    
|                         |                         | H1a3: I can easily access the information resources provided by mobile library application on my mobile phone.                                                                                          
| Correlation             |                         | H1b1: The information provided by mobile library application is relevant to information I need.                                                                                                          
|                         |                         | H1b2: Mobile library application can provide enough resources to meet my needs.                                                                                                                         
|                         |                         | H1b3: Mobile library application provides up-to-date information on my research area.                                                                                                                  
|                         |                         | H1b4: I think using a mobile library application is very important for me to learn and work.                                                                                                             
| System help             |                         | H1c1: Mobile library application can provide user consultation services.                                                                                                                             
|                         |                         | H1c2: Mobile library application provides a demonstration about how to use it.                                                                                                                     
|                         |                         | H1c3: The help information provided by mobile library application is easy to find.                                                                                                                 
|                         |                         | H1c4: Mobile library application can promptly recommend new information to user.                                                                                                                    
| Interface feature       | Screen design          | H2a1: UI of mobile library application was well designed, people are happy to use it.                                                                                                                  
|                         |                         | H2a2: User interface of mobile library application is similar to that of QQ, WeChat, and other mobile application frequently used.                                                                  
|                         |                         | H2a3: Screen sensitivity is good and touch responses fast.                                                                                                                                              
|                         |                         | H2b1: Academic resources, open class, and other information resources are systematically classified.                                                                                                   
|                         |                         | H2b2: I can set personalized navigation bar and customize related services.                                                                                                                     
|                         |                         | H2b3: Navigation design is reasonable, attracting users' attention.                                                                                                                              
| Term                   |                         | H2c1: Column names allow users to easily understand the contents.                                                                                                                              
|                         |                         | H2c2: Column terms are designed to prompt users to find relevant resources quickly.                                                                                                               
|                         |                         | H2c3: Term should have a specific meaning and not be ambiguous.                                                                                                                                  
| Individual difference   | Performance expectations | H3a1: Mobile library application is useful to me.                                                                                                                                                    
|                         |                         | H3a2: Mobile library application helps me with study and work.                                                                                                                                    
|                         |                         | H3a3: The benefits of using mobile library application are much better than not using it.                                                                                                          
| Domain knowledge        |                         | H3b1: I had knowledge of mobile library application before.                                                                                                                                         
|                         |                         | H3b2: I have experience of using mobile library application.                                                                                                               
|                         |                         | H3b3: I have a certain amount of expertise.                                                                                                                                                        
|                         |                         | H3b4: Domain knowledge will help me use mobile library application.                                                                                                                    
| Social influence        |                         | H3c1: My classmates, friends, teachers, and others are using mobile library application.                                                                                                        
|                         |                         | H3c2: Some students or colleagues recommend me to use mobile library application.                                                                                                              
|                         |                         | H3c3: Some classmates, friends, and colleagues say it is convenient to use mobile library application.                                                                                              
|                         |                         | H3c4: There are often friends sharing mobile library application in WeChat circle of friends.                                                                                                     
| User’s usage behavior   | BI1:                    | I am happy to use mobile library application.                                                                                                                                                     
| intention               | BI2:                    | I will use mobile library application.                                                                                                                                                              
|                         | BI3:                    | I would recommend mobile library application for other people.                                                                                                                                   |
Abstract: The paper explores the regional cooperation model and the differences among artificial intelligence research teams. It is helpful to reveal the status and strategies of scientific cooperation models across regions or within regions. We identified the world of artificial intelligence research teams with co-authorship network, and then identified the leading team based on the Number of Publications, Number of Citations, H-index, Weighted Degree Centrality, Betweenness Centrality, and Closeness Centrality. Based on the identified artificial intelligence research leading teams, this paper divides different types of cooperation models by region and comprehensively analyzed the three aspects of geographical distribution, cooperation indicators, and cooperation topics in the research teams from the perspective of comparisons. In order to find the international gap between China and other countries, we still highlight the difference between China’s participation and non-participation in cooperation. The research results show that Chinese and their foreign research maintain close ties with major scientific research countries; international cooperation is widespread and is conducive to crossing into the leading team; China’s domestic cooperation is higher than in other countries, and their domestic cooperation research is mainly manifested in the data processing and application level, while the core technology and basic algorithm levels need to cooperate with foreign countries.

Keywords: research cooperation, cooperation model, artificial intelligence, research team comparative study

1 Introduction

Scientific research cooperation is an activity carried out between scientific researchers through the sharing of scientific and technological resources and teamwork. In the environment of big data, the comprehensiveness, complexity, and interdisciplinary nature of scientific research are becoming more and more obvious, and scientific cooperation has become an inevitable trend (Ma, Li, & Chen, 2015). Analyzing the attributes and relationships of cooperation from different perspectives, the modes of scientific research cooperation show different expressions and characteristics, which in turn reflect different scientific cooperation trends and scientific research development laws. Investigating the scientific research cooperation model from the perspective of the location of the collaborator can discover the distribution, characteristics, and differences in scientific research cooperation between different regions or countries, and distinguish the research status of a country in the world or in certain regions. In addition, it provides a basis for promoting cross-regional or regional scientific communication and policy formulation.

As a comprehensive discipline that has attracted attention from all over the world and is rapidly developing, artificial intelligence (AI) not only produces a large number of scientific research cooperation in the research process, but also has a variety of cooperation modes across regions and fields. Therefore, it is necessary to explore
and compare the cooperation models of AI research teams from the geographical perspective, analyze the country and geographical distribution of different cooperation models, study the size and performance of the teams under each cooperation mode and their research topics, and provide an in-depth analysis of the AI field. The state of cooperation and development has important research significance.

At present, the international cooperation research situation mainly takes a single paper as the unit of consideration, from the perspective of co-authors and the related attributes of the paper. Zhou and Glänzel (2010) studied the overall and representative characteristics of China's international scientific and technological cooperation using bibliometrics and social network methods; Jin, Richard, Zhang, Cao, Wang and Zhou (2007) used bibliometrics to analyze the cooperative relationship between Chinese and American scientists in scientific papers; Hu, Zhu, Zhang and Chen (2009) used complex network theory to conduct scientific research in the field of supply chain management. The cooperation network conducts empirical research and compares the characteristics of scientific research cooperation models in this field; Guo, Yang and Kuang (2013) used the cooperation rate and cooperation degree to compare and analyze Chinese and foreign library and information cooperation models.

The influencing aspects are to explore the internal and external factors that affect international cooperation, including the nature of the subject field, national research strength, geography, politics, history, society, culture, and language (Pu, Yuan, Yue, & Liu, 2015); for example, Melin (1999) compares Northern Europe and the United States, Eustache Mêgnigbêto (2013) analyzed the scientific and technological cooperation model of multiple countries in western Africa and found that the impact of scientific and technological cooperation in western Africa is mainly reflected in the effectiveness of scientific and technological cooperation in western Africa. Regarding the impact of international scientific research cooperation on scientific research production, for example, Qiu & Zeng (2013) took the field of computer science as an example, using correlation analysis to verify whether international cooperation can improve the scientific research influence; Lee and Bozeman (2005) focused on the author’s individual level, analyzing the impact of cooperation on scientific research performance. The existing international cooperation models have multiple research perspectives. However, these studies are mainly based on the co-authoring relationship of single papers to characterize the cooperation models. Most of them use individual scientists as the research base. There is a lack of international cooperation research based on the team of scientists as a whole. There is less research on the content level of cooperation. What is more, a comprehensive and comprehensive discussion on the international AI research field has not yet been conducted.

2 Data and Methodology

2.1 Data Source and Preprocessing

2.1.1 Data Source

This research uses Web of Science (WoS) data for analysis. WoS is one of the most authoritative databases for scientific metrological analysis, and it provides comprehensive coverage of all metadata information from journals, books, and conference proceedings. AI is a complex emerging field. In order to retrieve all relevant documents on this topic, keyword retrieval will encounter the problem of insufficient retrieval rate. In the subject classification of the WoS database, there is an AI subcategory under the computer category, which covers all papers closely related to AI. Therefore, we download all the documents in this subcategory, wherein the inspection rate and accuracy rate are well balanced.

We use the WoS core set as the data source and set the search formula as WC = “Computer Science, Artificial Intelligence”, and the index includes SCI-EXPANDED, CPCI-S, CCR-EXPANDED. The time span is 2009–2018, and the data collection date is 2019. On January 16, 421,148 records were obtained. It is well known that conference papers in the computer field are of significant importance. Our search results include journal papers, conference papers, and so on, among which are important conferences in AI-related fields, such as CONFERENCE ON ARTIFICIAL INTELLIGENCE, and INTERNATIONAL CONFERENCE ON INTELLIGENT ROBOTS AND SYSTEMS. These data can represent research in the AI field as a whole. We download them as basic data and perform data cleaning.

2.1.2 Data Cleaning

Data cleaning mainly includes the name disambiguation of three parts: institution, author, and country. The entire process involves manual review, rule-making, and automated batch processing. First, the amount of
manual tasks for national cleaning is small, and the name expression is easy to summarize. We unified the country names and merged the regions into the corresponding country expressions. After that, processing of the organization name is differentiated according to the country name.

We extract the name of the institution and country to which the document belongs. If the name of the institution is the same but the name of the country is different, it is regarded as a different institution. We manually checked the top 1% of the organizations that issued the documents, combined the different expressions of the same organization based on the aforementioned rules that could not be found, and used the name of the organization with the higher number of documents as the standard. Finally, on the basis of institutional data cleaning, we extract the information about the author's institution and co-authors, and perform author cleaning. The main rules are as follows: authors with the same name in the same institution are regarded as the same author; if there are identical co-authors between authors with the same name, they are judged to be the same author entity; and authors with the same name that do not meet the above two conditions are judged to be different authors.

2.2 Identification of Target Team

After disambiguation of author names, an overall co-author network of all co-authored articles in the data set was constructed by the Louvain method in Pajek. The Louvain algorithm is a community discovery algorithm based on modularity, and its optimization goal is to maximize the modularity of the entire community network (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008).

In all, 94,347 communities were detected through this way. These communities are closely connected and separated from other communities, so they are regarded as research teams formed in a co-authored relationship. The largest of these research teams contains 1,553 authors, and only one of the top 10 teams is under 1,000 (996). We chose to delete the author nodes with a volume of 1 and a citation less than 100 in the original co-author network, in order to retain the research team members who have a certain influence in the field of AI. The overall co-author network has been greatly reduced while maintaining the original important nodes. In order to identify a research team with a reasonable scale and available for analysis, it is necessary to constantly adjust the parameters and compare and evaluate the results of the research team identified under different parameters. When the parameters are set to Resolution=290, Max Level=13, and Max Iteration=13, the granularity that is more suitable for observing the specific situation within the team is reached. The team size is less than 100 people. If the team size is too large, internal connections may be loose. If the team size is too small, some important connections may be missed. It should be noted that there are no strict criteria for the selection of team size and granularity, but the degree of closeness of the revealed research team will be different.

In order to reveal the diversity of the leading team, we use six indicators to measure the research team from different perspectives. They are Number of Publications, Number of Citations, H-index, Weighted Degree Centrality, Betweenness Centrality, and Closeness Centrality (Bonacich, 1972). Among them, the first three indicators measure the strength of the research team from the node attributes, and the latter three indicators measure the strength of the research team from the network structure (Abbasi, & Hossain, 2013). The top 100 teams in each index were screened out, and the top 100 research teams in each index were merged to get 342 teams. Some countries cannot be traced to specific country names and expressed as none, so they are regarded as unknown countries. This research does not consider teams with unclear data for the time being, so as to facilitate the calculation and analysis of the concentration of countries in the later period. Therefore, after removing this part of the team data, 333 teams are obtained as the research object of this article. According to the WoS number, the corresponding text is extracted from the cleaned text data for topic analysis.

2.3 Definition of Related Concepts

2.3.1 Concentrative Ratio

Considering the concentration degree of the author’s country in the team, the market concentration measurement indicator (HHI) is applied to this research to calculate the Concentrative Ratio (ConR) in the team as formula (1). The formula is expressed as a ratio of the number of authors in each country to the total number of authors in the team, and then the sum of the squares is calculated. The larger the ConR value, the more concentrated the author’s country in the team. The lower the degree of concentration, the more scattered the country the authors belong to in the team. The value of the ratio ranges from 0 to 1. A team with a value of 1 is a single-country cooperative team:
\[
\text{ConR} = \sum_{i=1}^{N} \left( \frac{x_i}{X} \right)^2
\]

where \( i \) is one of the countries in the list set \( \text{Country} = \{1, 2, \ldots, i, \ldots, N\} \), \( x_i \) represents the number of authors in the country, and \( X \) represents the total number of authors in the team.

2.3.2 The Dominant Country in the Team

Based on the previous processing, the research target team we obtained was extracted from the co-author’s relationship network, and the network was based on multiple papers. Usually, scholars take the co-authors of a single paper as a research group, so scholars will study the order of the authors and whether they are corresponding authors. This research is based on a community group formed from multiple papers, and it is studied from a macro perspective, so the above two aspects are not considered at present.

In addition, considering whether there are differences in cooperative teams led by different countries, this research introduces the concept of the dominant country. The country with the largest number of authors in the team is regarded as the dominant country in the team. The country with the largest number of authors in the team is regarded as the dominant country in the team. We identified the dominant country of each team, and conducted statistics in the international cooperation team to analyze the leading position of each country in the world.

2.4 Division of Team Cooperation Mode

In this study, the research team is divided into two aspects according to the geographical distribution pattern. On one hand, according to the proportion of the number of countries that the authors belong to in the team, it can be divided into two types: single-country domestic cooperation and multinational international cooperation. On the other hand, this study takes the comparison between Chinese and foreign research as the starting point, and calculates the proportion of Chinese collaborators in the team to further divide it into four types, namely, single-country cooperation with China, international cooperation with China’s participation, single-country cooperation without China, and international cooperation without China’s participation. The specific division is shown in Table 1.

Table 1 gives team cooperation modes and their quantities. The cooperation model of a single country involves 27 teams. Among them, there are 11 teams for domestic cooperation with China and 16 teams without China. There are a total of 306 international cooperation teams, of which there are 148 international cooperation teams without China’s participation, and 158 teams are involved in Sino-foreign international cooperation. It is worth noting that this research uses the leading AI team as the research object, and the purpose is to explore the regional cooperation model situation of the international leading team. The main investigation is the research situation of the leading team rather than the overall research situation, and each model is used as a measure. Overall, the difference in the number of leading teams in each mode does not have an impact on this research.

2.5 Research Design

According to the above research assumptions, based on the identification of AI research teams in the early stage, the main analysis framework designed in this paper is as follows: First, we carry out research on the country distribution and co-occurrence of AI research teams, use Python self-compiled programs to realize the country distribution map of the research team, and analyze cooperation between countries with co-occurrence relationships; second, we carry out comparative research on the cooperation performance of AI research teams, and compare the indicators of each model team from the perspective of multi-index calculation; Finally, a comparative study of the collaborative content of AI research teams is carried out. From the perspective of collaborative knowledge content, spectral clustering is used to form the research topics of different types of teams and conduct comparative analysis. The overall research framework process is shown in Figure 1.
3 Results and Analysis

3.1 Cooperation Status

3.1.1 Overall Distribution of Countries

A total of 82 countries are involved in statistics on the participation of countries in the team. The statistical results of the number of teams participating in the country are mapped to the world map according to the geographical location, and the statistical distribution of the number of teams in the world is shown in Figure 2. The lightness of the color in the picture indicates the number of participating teams, and the darker the color indicates that the country has more teams participating. The United States has the largest number of teams (201, 58.77% of the total), followed by People's Republic of China (173, 50.58% of the total). As a whole, the main distribution areas are the United States and Canada in North America, China and India in Asia, Australia and developed Western Europe.

At the same time, we conducted a distribution study on the dominant countries. There are 37 countries that dominate the team. Although 45 countries participate in teamwork, they do not occupy the dominant position. Count the number of teams led by countries and display them on the international map as shown in Figure 3. China dominates the largest number of teams (100, accounting for 30.03% of the total), followed by the United States (48, accounting for 14.41% of the total). The number of other leading countries is relatively small, and the number of international teams participating in non-dominant countries is also relatively small. On the whole, the country’s dominant position in the team is related to the participating teams.

3.1.2 Multinational International Cooperation

International cooperation is divided according to whether or not China participates. It can be divided into two situations: international cooperation with China’s participation and international cooperation without China’s participation.

3.1.2.1 International Cooperation without China’s Participation

The co-occurrence analysis of multi-country cooperation in this model is shown in Figure 4. The color of the map in the figure indicates the number of teams involved in the country, and the thickness of the line indicates the degree of co-occurrence between countries. It can be seen from the figure that the United States occupies a dominant position in international cooperation under this model, and there are more connections between it and many European countries. According to existing theory (Wang, 2014), countries with the number of co-occurrences with the United States greater than $0.749 \sqrt{n_{\text{max}}}$ are selected, where $n_{\text{max}}$ represents the maximum number of co-occurrences with the United States [11]. A total of 14 major co-occurring countries are obtained: Germany (32), UK (28), India (23), France (19), Switzerland (19), Spain
(18), Netherlands (14), Italy (13), Austria (12), Canada (10), Japan (9), Belgium (8), and South Korea (8). The number of co-occurring teams is in parentheses. It can be seen that the cooperation between Europe and the United States in the international cooperation of AI is relatively close, and it is mainly manifested in the cooperation between the developed countries in Western Europe and North America. India’s AI research also occupies a certain position in international cooperation.

### 3.1.2.2 International Cooperation with China’s Participation

This model represents some groups that conduct academic exchanges between China and other countries. The distribution and co-occurrence of countries in this model are shown in Figure 5. In the figure, China occupies a core position. In addition to China, the United States, Australia, Canada, and Europe are the main connection points, and the connection between China and the United
States is the thickest, indicating that in this model, China mainly cooperates with the United States. We select countries that have a number of co-occurrences with China greater than $0.749 \sqrt{n_{\text{max}}}$ where $n_{\text{max}}$ represents the maximum number of co-occurrences with China. There are 14 major co-occurring countries, namely, the United States (111), UK (57), Australia (43), Canada (36), Singapore (25), India (21), Germany (20), France (19), Japan (17), South Korea (16), Saudi Arabia (14), Malaysia (11), Brazil (10), and Spain (10). It can be seen that China’s international cooperation covers a wide range, mainly in North America, Oceania, and Europe. Among the Asian countries that China cooperates with, Singapore is the first.

We identify the dominant country in this model, and count the number of teams and their proportions, as shown in Figure 6. Among the international cooperation teams with China’s participation, the number of teams led by China is the largest, followed by the teams led by the United States. There are 8 cooperative teams led by the United Kingdom, 7 cooperative teams led by Australia, 6 cooperative teams led by Canada, 4 cooperative teams led by France, and 3 cooperative teams led by Singapore, India and Japan., Germany, South Korea, and Spain are led by two teams, and the rest are one.

### 3.1.3 Single Country Domestic Cooperation

The statistics are divided by country, and the number of teams in the country under the domestic cooperation model is ranked. Tracing the specific countries in the single country cooperation model team, there are 14 countries with single country cooperation model, namely Peoples Republic of China (11), Brazil (3), USA (2), and Canada (1), Czech Republic (1), Germany (1), India (1), Israel (1), South Korea (1), Spain (1), Portugal (1), Mexico (1), United Kingdom (1), and Turkey (1). The numbers in parentheses is the number of teams in this type of mode. The Chinese domestic cooperation model team refers to a cooperative group of domestic experts in China, which also belongs to the cooperation model of a single country. Compared with the number of groups cooperating with other single countries, China has the largest number of cooperating groups. This confirms the large number of Chinese scholars, and the number of groups involving Chinese scholars under this model; on the other hand, it also reflects that some Chinese scholars have limited cooperation.
3.2 Cooperation Performance

3.2.1 Cooperation Performance of Different Cooperation Modes

This study counts the number of authors involved in the team under different modes and calculates the number of publications, citations, and H-index in each team based on the principle of score counting. The results of each index obtained are shown in Table 2. As a whole, all indicators present the characteristics of “domestic cooperation < international cooperation”. Regardless of whether it is international cooperation or domestic cooperation, the indicator values of the team under the model with Chinese participation are greater than the indicator values of the team under the overall international cooperation model.
The indicator values of the team under the model without Chinese participation are all lower than the overall international cooperation.

It can be seen that China’s participation in international cooperation can help improve the research performance of the team, and Sino-foreign cooperation has promoted the output of scientific research results to a certain extent. For China, the performance of international cooperation is higher than that of domestic cooperation, indicating that international cooperation in the field of AI research is of certain importance for entering the leading team.

3.2.2 Cooperation Performance of Different Concentration Ratio

We calculate the concentration ratio of the author’s country in each team based on formula (1). In order to more concisely show the performance of teams with different levels of concentration, we further divide the range of concentration levels and calculate the indicators in each range. The calculation results are shown in Table 3. On the whole, the values of the indicators in different ranges are not much different, but there is still a small gap. When the concentration ratio is 1, there are fewer publications and H-index, which is a single country domestic cooperation; when the degree of concentration is in the range of 0–0.2, the values of the three indicators are all low, and the team is too scattered. It can be seen that teamwork needs to have a certain degree of concentration, but it cannot be completely concentrated in one country, that is, international cooperation with major cooperating countries is more conducive to the output of scientific research results.

3.3 Cooperation Contents

3.3.1 Multinational International Cooperation

This part of the research separately conducts thematic analysis of the presence or absence of China’s participation.

3.3.1.1 International Cooperation without China’s Participation

The research topics in this mode are shown in Table 4. The main collaborative research includes various algorithms, such as Differential Evolution, Clustering Algorithm, Support Vector Machine, Neural Network, and Deep Neural Network. The differential evolution algorithm guides the optimization search by imitating the heuristic group intelligence generated by the cooperation and competition between individuals in the biological group. t-s fuzzy system and fuzzy rough set are often used in system simulation research (Yuan & Wang, 2006). Robust stability is an evaluation of the model. In the application field, the leading team of major foreign cooperation models mainly focuses on the related research of face recognition. Based on the above research themes, the cooperative research of these teams is mainly at the level of basic algorithm research and basic model level, focusing on basic technology research and model performance.
The research topics of the cooperative model team are shown in Table 5, and the research topics are more detailed. The main research topics in the application field are Humanoid Robot, Face Recognition, Wave Energy Facility Location, and so on. The algorithm level mainly includes the Memetic Algorithm and the Convolutional Neural Network. The former is an optimization algorithm based on simulated cultural evolution proposed by Pablo Moscato. It is a global search based on the population and individual based a combination of the local heuristic search. Research on computer technology includes Multi-Agent System (MAS), Intrusion Detection, Computational Modeling, and so on. The Parameter-Exploring Policy Gradient method in the parameter space is the most effective and powerful policy search method. Based on the above research themes, the collaborative research of AI teams with China’s participation highlights the research in the field of robotics. The research at the algorithm level and search strategy level is more in-depth than other models, the research at the data mining and computer technology level is richer, and the overall presentation is that of a computer hardware-based situation.

### 3.3.2 Single Country Domestic Cooperation

In this mode, we select only China as the research object and analyze the research topics of China’s domestic cooperation. The research topics of the team are shown in Table 6. The research topics cover a wide range. Color Space is an important part of image recognition research (Yin, 2017). Face Reconstruction research mainly involves face analysis, model fitting, and image synthesis. These two research themes indicate that domestic cooperation is in progress. Research on face recognition, principal component analysis, and adaptive neural network are commonly used algorithms in the field of AI. The former is often used for dimensionality reduction in data mining. Departure Warning System and Fuzzy Risk Analysis are applied researches at different levels. Rough set and point matching are also commonly used mathematical tools and methods in data processing. It can be seen that the collaborative research of China’s domestic AI leading teams is mostly data processing level and application level research, highlighting face reconstruction and color space in face recognition research, and focusing on the effectiveness of algorithm learning.

### 4 Discussion and Conclusions

Driven by data science, this article analyzes the cooperation model of the leading AI research team from the perspective of the regional cooperation model, and focuses on the comparison of the AI leading team with or without China participating. On the basis of identifying the leading team, divide the research team cooperation mode by region, comprehensively compare and analyze the cooperation mode of AI research team from the geographical distribution, cooperation performance indicators, and cooperation theme content, and highlight the comparative analysis between China and foreign...
countries. The research conclusions obtained are as follows.

From the perspective of cooperation status, international AI research maintains close ties with the major scientific research countries. Teamwork is mainly established between European and American countries, with the United States as the leader, and North American and Western European developed countries as the main cooperating countries; in addition to the major scientific research foreign countries, more Asian countries cooperate with China.

From the perspective of team performance, it is mainly divided into two aspects. On one hand, the performance of international cooperation is better than that of single-country domestic cooperation, and the performance of international cooperation teams with China’s participation is higher than that of international cooperation teams without China. On the other hand, teams with different concentration degrees in different countries have different performances. There is no obvious linear correlation between the concentration degrees of countries in the team and performance. However, the performance is relatively poor in both cases of complete concentration, that is, a single country state, and complete decentralization.

From the perspective of cooperation content, international cooperation focuses on research at the level of core technologies and basic algorithms. International cooperation without China’s participation is mainly reflected in basic research. International cooperation with China’s participation is more detailed and in specific algorithms and computer technologies it is more abundant.

In addition, to support the perspective of China, we integrated the analysis of three aspects of cooperation status, cooperation performance, and cooperation content. China’s domestic cooperation is higher than that of other countries, and some Chinese researchers need to expand their circle of cooperation for international cooperation. Increasing international cooperation is crucial to joining a leading AI research team, and leading international cooperation is more conducive to team development. The subject of China’s domestic AI cooperative research is mainly reflected in the data processing and application level, while most of the core technologies and basic algorithms are cooperated with foreign countries.

Compared with previous studies, this research starts from the perspective of teamwork, combining country distribution and the presence or absence of China’s participation to classify teamwork models. At the same time, it introduces a market concentration index to indicate the concentration index of the country where the authors belong to the team, and perform performance considerations on the identified teams. However, this article still has certain limitations. It only selects literature data as the research object, and uses only China as the comparative country. The future research will conduct multisource data analysis and supplement the research of multiple data sources such as political reports and fund projects. We will try to further explore the differences among the Chinese participating with each of the other countries and work hard to find other supporting materials to enrich the comparison, with a view to more comprehensively and accurately examine the cooperation and research status of international AI research teams.

**Acknowledgements:** This study is supported by the major bidding project of the National Social Science Fund of China (16ZDA224) and the China Scholarship Council “2019 National Construction High-level University Public Postgraduate Project” (201906840116).
References


Yuanye Ma*

Language and Intercultural Information Ethics Concepts: A Preliminary Discussion of Privacy

https://doi.org/10.2478/dim-2020-0027
received August 15, 2020; accepted September 15, 2020.

Abstract: This paper introduces the perspective to understand privacy via language as an intercultural information ethics (IIE) concept. This research perspective carries two goals: to understand privacy as an IIE concept and to do so via natural language. The paper suggests that studying privacy through language answers the challenge faced by IIE work; in addition, studying privacy as an information ethics concept through language seems most appropriate considering that language both embodies and shapes meaning. Specifically, this paper briefly discusses privacy and some of its language expressions in the Chinese and English languages, through which it hopes to reveal the richness and possibilities of using natural language as a research instrument to understand privacy in intercultural settings, which is an area of researching privacy that has attracted little discussion so far.

Keywords: information ethics, intercultural information ethics, natural language, privacy

1 Introduction: Information Ethics and Intercultural Information Ethics

Information ethics (IE) is still a young area, although it has enjoyed development and growth in the past three or four decades. Froehlich (2004) discussed the history of IE in the United States and how it is a “multi-threaded phenomenon” in which multiple disciplines, including library and information science, computer science, and media and journalism, all contribute to the discussions. There are two ways of understanding the scope of IE.

Yuanye Ma, School of Information and Library Science, University of North Carolina at Chapel Hill, Chapel Hill, USA, Email: yuanye1@live.unc.edu

The first way sees IE as the applied ethics of computers (Johnson, 1985; Heersmink et al, 2011). Currently, many IE courses at iSchools still introduce IE as applied ethical issues, which receives a topic-by-topic discussion; for instance, privacy and intellectual property are some of the most common themes of study (Mattern & Gunn, 2019).

However, it is important to note that even in its early days of development, researchers were suggesting a second and broader conceptualization of IE that is beyond just ethical considerations in the face of specific computer applications, as stated clearly by Moore (1984, p.267): “I am arguing for the special status of computer ethics as a field of study. Applied ethics is not simply ethics applied. However, I also wish to stress the underlying importance of general ethics and science to computer ethics. The ethical theory provides categories and procedures for determining what is ethically relevant.” It is through this second and broader conceptualization that we aim to understand IE in this paper.

Similarly, Nishigaki (2006) differentiates between the search for ethical norms in the context of new information technologies on one hand, and the changes “on our views of human beings and society” becoming “necessary to accompany the emergence of the information society” on the other hand (p.237). The second way of considering IE jumps out of the specific application areas and focuses more on the overarching themes across all areas of application because of digitalization, which proposes a much broader scope, that is, which was summarized well in Hausmanninger & Capurro (2002) as “the reconstruction of all possible phenomena in the world as digital information and the problems caused by their exchange, combination and utilization” (p.10). This broader way of conceptualizing IE can also be found in Floridi (2006), where IE is conceived of as macroethics, “an ethics that concerns the whole realm of reality” (p.25).

Floridi’s conceptualization of IE is situated in his conceptualization of the Philosophy of Information (PI), where PI is defined as “the philosophical field concerned
with the critical investigation of the conceptual nature and basic principles of information, including its dynamics, utilization, and sciences, and the elaboration and application of information-theoretic and computational methodologies to philosophical problems” (2002, p.123). PI considers information as the “primary ontological category and constituent” (Ess, 2009, p.160), “to be is to be an informational entity” (Floridi, 2008, p.199). The perspective of PI broadened the reach of ethics, and it moves beyond the limits of anthropocentric ethical theories (Stahl, 2008). Therefore, IE, by Floridi’s (2014) conceptualization is concerned with the “infosphere” (term in reference to biosphere), which consists of “inforgs” (short for informational organisms) and which include human and nonhuman informational entities with equal moral standing.

This shift to consider nonhuman agents is considered as a major change from modern Western emphasis on the human as a moral agent; it recognizes that “moral actions are the result of complex interactions among distributed systems integrated on a scale larger than the single human being” (Floridi, 2008, p.198). This recognition of the larger social and ethical systems that individuals are situated in connects us with the issue of Intercultural Information Ethics (IIE). In other words, it seems that until quite recently, intercultural consideration that has been underlying and remains quite implicit.

The history of IIE as a separate discipline of its own is only about a decade old (Jared, 2015). IIE was first introduced by Rafael Capurro (2006), and it was summarized as a “contemporary reflection on especially a computer ethics oriented towards global dimensions of Information and Computing Technologies (ICTs) and their use” by Ess (2008, p.89-90). The study of IIE finds its motivation in this concern of IE studies relying solely on Western philosophical and ethical traditions, where the non-Western is largely ignored. Hence, the aim of IIE “is to provide shared norms for different societies with different cultures and distinct moral systems; and, at the same time, maintain the cultural diversity and respect the distinctiveness of various moral systems” (Wong, 2009, p.52). Additionally, the aim of IIE, as Ess puts it, is to: “1) address both local and global issues evoked by ICTs, and 2) to do so in ways that both sustain local traditions/values/preferences, etc. and 3) provide shared, (quasi-) universal responses to central ethical problems” (2007, p.102).

The research exploration of IIE spans across two stages/parts: one descriptive and the other more normative (Wong, 2009). The descriptive part of the IIE research will require substantive empirical work that can help with explicating the norms and values embedded in cultures. The first part of the work will help “provide the basis for formulating the (quasi-) universal moral principles,” which is the central concern in the second part of the work. The second stage is challenging for its normative nature (Brey, 2007); it needs to answer the question, what should different cultures share, as explicitly asked by Wong (2009): “Since IIE aims to investigate ICTs-related ethical problems from various cultural perspectives and attempts to settle them interculturally or cross-culturally; therefore, it must employ both empirical findings of different cultural perspectives as well as normative analysis to determine what can, and should, be agreed upon” (p.51).

Wong (2009) also discussed in more detail the challenges of pursuing this second stage of research and how two existing understandings are inadequate when it comes to the question of what should be shared. The two understandings critiqued by Wong are, first, cultures have shared norms but the interpretations of these norms could differ. Wong criticized this view as it relies on a “minimal denominator” as the “foundation” across cultures; this perspective was questioned as it risks “pushing IIE towards metaethical moral relativism” (p.53). Another understanding questioned by Wong was that cultures could have shared norms with different justifications. Again, Wong doubted this view as it downplays the scope and significance of IIE, because although IIE is concerned with many ICT applications and problems, its core consideration is about morality. This second understanding risks reducing IIE to a justification issue, which significantly downplays the study of IIE to pragmatics.

So the articulation of the research scope and actual research operations of IIE are intertwined, and sometimes it boils down to the consideration of what rightfully falls under the research scope of IIE. For example, Brey (2007) suggested that studies of the effects of information technology in non-Western cultures are more appropriately delegated to the social sciences (including communication studies, cultural studies, anthropology and science and technology studies), where IIE primarily focuses on the comparative study of moral systems (p.8). However, even when one scopes the study of IIE as comparative ethics, it still remains a question in terms of what might be considered as an adequate instrument through which the IIE issues and questions can be approached.

With these understandings, this current paper aims to argue that natural language can be considered as such an instrument through which the topic of IIE can be approached. We will not review the relation between language and philosophy, as it is a topic of too broad scope; instead this paper will use the concept of privacy
and its existence or expression in the Chinese and English languages as concrete demonstrations. Specifically, this paper is structured as follows: in the literature review section, we discuss the existing work of privacy as an IIE concept. Next, we move on to see how natural language can be used to understand privacy, and how natural language can offer valuable support to existing work of privacy research, which largely remains within theoretical discussion.

2 Related Work: Privacy as an IIE Subject

In contrast to the abundance and systematic production of privacy research in the Western and especially American academia, discussions of privacy in non-Western cultures are scattered and sometimes produce conflicting conclusions. More importantly, it is unclear yet where comparative illustrations of the concept of privacy in intercultural context are going. Research with a comparative mindset started with questions about how privacy might be conceptualized in non-Western cultures in comparison to the so-called Western conceptualizations. On closer look, existing intercultural discussions comparing the understanding of privacy of a non-Western culture with a generalized Western understanding of privacy touch upon two specific questions: (1) how to describe the differences and (2) what might be some of the contributing factors to the differences. More often, these two aspects are mixed in the actual discussion.

Privacy in recent Chinese society was considered to be hindered by the Chinese culture’s collectivist nature (Lv, 2005); in other words, it could be quite difficult to have privacy in China, at least not the privacy known by its typical individualistic understanding. Lv (2005) did touch on this question of on what grounds privacy might be supported in Chinese society. In Western cultures, privacy was advocated for its association with individual autonomy, while it is likely that China, because of the mixture of its traditional culture and influx of Western ideas, will have to come up with a story of advocating privacy from “both individual and collective perspectives” (p.14). This question will soon lead to a more fundamental question, which is, whether any traditional Chinese school of thought (Confucianism, or Daoism, etc.) ever had a comparable idea of individual or individual autonomy. This question is definitely beyond the scope of IE or IIE, but a good understanding of this question should not be neglected by the IE and IIE research. This recognition is essential in that it prepares the working ground for the conceptualization of any IIE concepts, including privacy.

Lv’s (2005) observation perhaps has captured some true aspects for understanding privacy in the Chinese culture; however, it risks oversimplifying not only the culture in which privacy exists, but also the concept itself. Perhaps an alternative and more cautious way to put it is that both the Chinese culture and the conceptualization of privacy itself is multidimensional, in that there can be a range of possibilities for understanding. Drawing together the collectivist aspect of the Chinese culture and the individualistic characteristic of the conception of privacy is putting two of the most apparent incompatibilities together, which is interesting and valuable to kickoff the discussion. However, by no means does this indicate the completion of discussion.

Attention needs to be directed to domains that are murky; moreover, rather than identifying what might make the conceptualization of privacy diverge in different cultures, an equally meaningful task is to map out if and what common grounds are shared across cultures in terms of the conceptualization of privacy. The discussion in Ma (2019) is an early attempt in this direction, where the discussion was directed toward compatibilities between the conception of “relational person” that can be found in both the Confucianism tradition and the feminist philosophy.

Similar to the discussion of privacy in China, privacy was argued as an imported concept for Japan (Nakada & Tamura, 2005), and as a Western conceptualization it is incompatible with the Japanese worldview trichotomy (which consists of Ikai, Seken, and Shakai). Specifically, in the Japanese trichotomy worldview, Seken refers to the aspect of the world that consists of traditional and indigenous worldviews or ways of thinking and feeling; Shakai includes modernized worldviews and ways of thinking influenced in many respects by the thoughts and systems imported from “Western” countries; and Ikai is the world of “the other(s), i.e., the hidden or forgotten meanings or values in Seken or Shakai,” as well as where spiritual meanings originate (p.27). In addition, another challenge for understanding privacy in Japan is that Japanese society in general grapples with two systems of understanding; the traditional Japanese culture (influenced by Buddhism, Confucianism) and that of the West, which was imported to Japan more recently.

It was suggested that a typical Western understanding of privacy might be only applied within the realm of Shakai, while Seken is related to the social relationships and the social community that an individual finds him/herself. Based on the above illustrations, Nakada & Tamura (2005)
suggested that Japan might have imported privacy only partially. It was further suggested that not only is privacy a concept less valued by the Japanese people, but Japanese culture might be equipped with an alternative concept that is more dominant in terms of regularizing the social roles and expectations of individuals, for instance, the concept of **Bun** (仏, Bun in general refers to the different roles depending on one's relationship with other people) (p.31).

Considering **Seken**'s social community perspective, what might be understood as an invasion of privacy in Western society is considered as not so problematic in the Japanese society. For instance, in the case of revealing a criminal's personal information, personal information is not just about an individual, but about his broader social duties: “people need information about the victims' personalities and relationships in order to understand the meanings of this homicide ...”; “What may seem like a violation of privacy to Westerners is thus justified from the perspective of Seken” (p.30).

In addition, the conceptualization of privacy is highly impacted by other concepts that may be foundational to the meaning of privacy, for instance, the concept of public and private. **Nakada & Tamura** (2005) discussed the concepts of public (Ohyake) and private (Watakusi) and how traditional meanings attached to these concepts may have an impact on privacy. The traditional Japanese understanding would say, “things related to Watakusi are less worthy than things related to Ohyake” (p.32). The downplay of Watakusi is perhaps one of the reasons that privacy is not as important in Japan as it is in the West. In short, a comparison of the concept of privacy seems inevitably to invoke a comparison of also the broader social and cultural contexts in which privacy and its related concepts come to have meaning.

An alternative framework of understanding the incompatibilities of the concept of privacy across cultures can be found in **Mizutani, Dorsey, & Moor** (2004); when trying to decipher privacy in the Japanese and American societies, they make this distinction between **descriptive** privacy and **normative** privacy. **Descriptive** privacy is understood as the presence or absence of privacy as a matter of fact (p.121); it describes the actual situation regarding privacy. For example, in fact, generations of Japanese families live together and do not have their own separate rooms. In contrast, **normative** privacy refers to the situation that, regardless of the actual situation of privacy, there can be expectations of privacy, or normative rules about privacy always exist. **Mizutani et al.** (2004) argue that the lack of the former does not indicate a lack of the latter, and the existence of the former may not be a guarantee of the latter. Indeed, **Mizutani et al.** (2004) consider the lack of privacy in Japanese society as largely a result of practical constraints (e.g., close and limited physical living spaces), while privacy as a concept of inherent value still exists.

Based on the distinction of **descriptive** and **normative** privacy, it seems possible for Japanese and American cultures to share a “minimal conception” (p.124) of privacy, while a full equivalence of the richness of privacy in the two cultures is less likely. **Mizutani et al.** (2004) concern resonates with **Ess** (2005) in that it would be a rush to conclude on an absolute foreignness of the concept of privacy to Japanese culture. However, what seems less clear is how the minimal conception of privacy evolves over time.

**Mizutani et al.**’s (2004) discussion is insightful in that it recognizes the complexity of privacy, in that descriptive privacy is only one aspect of the concept, and how the concept can be impacted by real-world affordances. **Mizutani et al.**’s (2004) discussion, in particular, also cautions that language can both guide and mislead the understanding of privacy, in that the absence of a certain language object does not indicate the complete absence of meaning: “the absence of a single word to describe a concept does not mean the concept is totally lacking, it does suggest that the contours of that concept and its discursive role may be different” (p.121).

Apart from comparing China or Japan’s conceptualization of privacy with that of the West, the comparison between Chinese and Japanese privacy conceptions is as interesting as the Western and non-Western comparison, starting with the word privacy in the Chinese and Japanese languages. Although the modern Chinese language has borrowed many Western concepts from Japanese vocabulary, for instance, **democracy** was translated in Japanese first as 民主, which was then imported to China so that in both languages, they use the same Chinese characters for representing democracy (Ruitenber, Knowlton, & Li, 2016). However, when it comes to the vocabulary for privacy, the Japanese language only has katakana (プライバシー) for the concept, while the mandarin Chinese language equivalent for privacy seem to have evolved from yin1si (隱私) to yin3si (隐私) – both are compound words consisting of two characters. Hence, when the Chinese and Japanese languages are put in contrast, it seems that privacy in the Japanese language appears more like an import, because katakana is used for transcription of foreign-language words into Japanese.

To summarize, in intercultural discussions of the meaning of privacy, first, the concept of privacy and its understanding hinge upon other related concepts and...
more foundational cultural preferences. Second, it is highly likely that intercultural understandings of privacy can reach at least partial correspondence, in other words, but it seems to be more interesting and important to specify the grounds on which common understandings were obtained. Third, in any correspondence, the lack of which shall not be considered as static, they need constant examination as individuals and the society continues to grapple with the multidimensionality of cultures and concepts. Overall, we have seen two different conclusions when comparing the concept of privacy across cultures: one suggests that privacy as a construct is more common in one culture and can be completely foreign to another culture; moreover, a different view suggests a more nuanced understanding, which suggests some common ground between cultures. Ultimately, the purpose of having such discussions lies not in reaching a clear-cut diagnosis of whether privacy is imported or not, but rather, through the explicating of relevant factors and how they matter, a better understanding of privacy might be reached.

3 Privacy in the Two Languages: Chinese and English

In this section, we start with the most relevant word, which is the word representing the concept of privacy in these two languages. We discuss the two characters that constitute the Chinese word yinsi (隐私), and we also see the negative connotation that was once associated with the word “private” in English. The examinations of the characters and words are revealing in that they could tell some of the underlying constructs/concepts that give rise to the concept of privacy, and how privacy as a concept has always been in constant change. Hopefully, these brief discussions can help make the point that language can be a promising tool for understanding a concept, as “language is most productively conceptualized as a semiotic tool, namely, a tool for meaning-making and meaning exchange in imagined or real social interaction” (Holtgraves & Kashima, 2008, p.73).

3.1 Privacy in the Chinese Language

A period after the founding of the People’s Republic of China in 1949 appeared to be when the notion of privacy was harshly attacked, and it was when “private property was banned, and personal desires, including the desire for a space of one’s own, were strictly abhorred among Chinese citizens, old and young alike” (Ong & Zhang, 2008, p.6; Naftali, 2010, p.301). However, the several decades since the 1980s seem to have been witnessing a growing appreciation of privacy (Gao & O’Sullivan-Gavin, 2015), till the last decade of the 20th century was when yinsi (privacy) became “an independent concept that did not need to be contrasted with the ideal of public service” (McDouggall, 2005, p.112).

The word “privacy” in the Chinese language as it is accepted nowadays is a compound word that consists of two Chinese characters, yin (陰) and si (私), each character with its meaning (yin could mean hidden or hide, while si could mean private, selfish, etc.). The choice of the two characters will inevitably reflect and reveal the meaning of the word as we will see shortly. Compound words began to appear during the Han dynasty (206 BC-220 AD), but did not increase substantially until modern times, from roughly 20% of the written lexicon before the Qin dynasty to more than 80% today (Shi, 2002).

Because Chinese is a tonal language, in this paper we use numbers to indicate different tones: 1 indicates the first tone and 3 indicates the third tone. Privacy as yin3si with yin in the third tone was an occurrence in recent decades (Gao & O’Sullivan-Gavin, 2015); before yin3si, it was yin1si that was used in the context of Chinese language. For example, “The Criminal Procedure Law (National People’s Congress, NPC, 1979), which came into effect in 1980, used the term “dark secrets (yin1si阴私) when stipulating that trials involving personal secrets should not be open to the public” (Gao & O’Sullivan-Gavin, 2015, p.235). It was suggested that the first bilingual dictionary appearance of privacy in the Chinese language as yin1si has occurred in a Chinese-English Dictionary compiled by the English Department of Peking Foreign Languages College in 1979 (McDouggall, 2005, p.113). Even earlier, Yinsi (regardless of the first or third tone) appeared nowhere to be found acknowledging that more search in this area is needed: it did not exist in the 1947 encyclopedic dictionary Ci hai (A sea of terms), or in the 1952 revised American edition of Mathew’s Chinese-English Dictionary, or in Lin Yutang’s 1972 Chinese-English Dictionary (McDouggall, 2005).

The two tonal variations of the character yin, especially yin1, can bring in negative connotations for yinsi. In contrast to yin3, which literally means hidden and appears mostly neutral, yin1 could have meanings that are neutral like shade or feminine, but it could also mean negative, or even sinister. Hence, in contrast to yin3si, yin1si appears to carry a more derogatory sense, and the shift from yin1si to yin3si also reflects a change in the understandings of privacy in the Chinese language. In other words, it is a change in shedding the negative connotations. The shift
from *yin1si* to *yin3si* also appeared to have occurred in parallel to an expansion and clarification of the meaning of privacy. In *Opinion on Several Issues Regarding the Implementation of the General Principle of Civil Law (Trial)*, China’s Supreme People’s Court (1988) formally expanded the concept of privacy beyond cases of sexual crime and categorized it under the right to reputation (Gao & O’Sullivan-Gavin, 2015, p.235).

Despite the arguable lifting of negative connotation with the change from *yin1* to *yin3*, the second character in *yinsi*, or *s1* since it is in first tone, may be a defining character that contributes to the meaning of privacy (Farrall, 2008; Huang, 2000; McDougall, 2005; Naftali, 2010; Zarrow, 2002). *Si* is the character that refers to what is private, a realm of understanding that researchers seek to clarify when trying to understand privacy in China (Farrall, 2008; McDougall, 2005; Moore, 1984; Zarrow, 2002). The two parts in public–private (*gong–si*) help define each other; to understand gong well can reveal the meaning of *si*, and only when the meaning of *gong–si* is carefully delineated, can one grasp the meaning of privacy that relies on top of it.

The negative connotation of *si* (private) originates from its antonym to *gong* (public); another antonym of *si* in traditional Chinese language is *guan* (official); in either situation, the character *si* bears a negative connotation, with implications of “disreputable actions carried out in secret and/or from disreputable motives” (Farrall, 2008, p.2). McDougall (2004) also named several compound words common in Chinese that consist of the character *si* and their derogatory meanings, including *si xin* (private mind), which can be easily associated with selfishness, and *si tong*, which means adultery. McDougall (2004) tried to explain the possible cultural and philosophical sources of the negative sense of the character *si* by tracing it back to Confucianism: the ancient Chinese Confucian classics *Li Ji* (Book of Rites), in which one of the chapters (The Great Harmony) begins with, “Tian xia wei gong,” which can be literally understood as “under heaven (everyone/ everywhere) was (held in) common”, notice the presence of gong here and how it is interpreted as “the common good,” which might be interpreted to overshadow (Huang, 2000) whatever that is not gong, namely, *si* (p.2).

The discussions of the two characters that constitute *yinsi* are an initial demonstration of how the meaning of privacy is situated in a larger cultural and social context, and a closer look at the two characters could reveal shifts the concept had gone through. In addition, a closer look at language reveals hopefully how a simple equation of “privacy” with *yinsi* is difficult and less meaningful, if not impossible.

### 3.2 Privacy in the English language

Since Warren and Brandeis’ (1890) seminal paper, privacy has been recognized and supported on multiple grounds, including, but not limited to, its moral value, its protection of personhood, and many more (Moore, 2017; Solove, 2002). Privacy is something of a positive value, it is to be preserved and defended, and the slightest cast of doubt toward privacy itself may appear odd and ridiculous today. Such a well-established appreciation of privacy is a result of years’ evolution; unlike the shifts in the Chinese language that seem to have occurred within recent decades, privacy in the English language may have centuries to trace through.

The word privacy had its initial appearance in the English vocabulary, as discussed in Huebert (1997, p.28), in the mid-fifteenth-century in the Oxford English Dictionary. However, privacy as it is understood in today’s English language has not always been considered this way, and by no means will its current interpretation remain unchanged. The etymology of the word *private* could, first of all, help reveal where a negative sense of privacy may have existed. The word *private* comes from the Latin word *privatus*, meaning “to be deprived” or “limited” (Baldwin Lind, 2015, p.51-52). Similar to the discussion of privacy in the Chinese language, the sense of negativity of *private* was also revealed by its opposing relation with the *public*. The sense of dispossession that the private space originally conveyed meant, “withdrawing from the public body or restricted to one person or a few persons as opposed to the wider community; largely in opposition to public” (Baldwin Lind, 2015, p.51–52).

It was argued that “before 1700, private was essentially a negative term…” (Longfellow, 2006, p.315), which suggests that the shift in the meaning of privacy from negative to positive was accompanied by a shift in the meaning of private; in addition, such a shift occurred during the past three centuries or so. As summarized by Huebert (1997), “... there is a progression from suspicion of or hostility to privacy in the earlier texts to acceptance of and even cherishing of privacy in the later ones. This would by no means be a smooth linear development, but the overall trend would hold nonetheless” (p.35).

It appears that the opposing relation between private and public/official as we have discussed earlier within the Chinese language context can also be found in English. “...the early modern public, often opposed to the private, was strongly linked to office-holding ... an official persona was almost always a public figure with public responsibilities in a specific sphere. Within this defining context in which the public was understood, the private became the sphere
of those who were subordinate or had to obey those exercising office” (Baldwin Lind, 2015, p.58). As far as we could tell today, the sense that private inevitably obeys or is subordinate to the public has largely alleviated, or perhaps even subverted.

A quick contrast of how privacy was understood a century ago in English to how it is constructed in language today could also reveal how the concept has dramatically changed. Warren and Brandeis’ (1890) conception of the right to privacy was responding to a few socio-technical changes’ intrusion on a person, in a time when portable photography became available to more people and when sensationalistic journalism was rising (Solove & Schwartz, 2015). Privacy as the right to be let alone emphasizes the status of a person not being disturbed, regardless of the specific content of information involved. This conception of privacy appears to be already different from how privacy is considered today, and such difference may be best revealed through the language of privacy. For example, privacy as the right to be let alone conveys this sense of the avoidance of disturbing external forces, as if it is preserving an original status each person already contains until being intruded upon. However, for today, privacy is preserved only as a result of certain action, as indicated by commonly used words like “control,” “consent,” and “opt-in/out.”

4 Conclusion

This paper is an initial demonstration of the language used to represent the concept of privacy in the Chinese and English languages. These discussions of concrete characters and vocabularies, hopefully, have demonstrated how natural language is not only a container of ready meaning, but by itself can be a means to gain deeper understanding. Recognizably, the discussions in this paper are still quite limited, both in terms of understanding privacy as an IIE concept, and in terms of using natural language to conduct the IIE research.

In terms of understanding privacy as an IIE concept through language, a comprehensive understanding of related language corpora would add great value, which includes a review of existing discussions of privacy in various literatures (for example, like Longfellow (2006) and McDougall (2005)); moreover, a comprehensive understanding also requires a grasp of privacy as it exists in actual natural language on a daily basis today. When it comes to using language to study IIE, future work could proceed with reviews on linguistics and psycholinguistics features of language and their effects on understanding and expression. In addition, because the use of language is capable of conveying meaning at two different levels (Tng & Lee, 2016), one at the semantics and vocabulary levels, and the other at the cultural and social levels, a review of the interaction of language and social/cultural understandings could help with clarifying where the study of privacy via natural language interfaces with other social research on privacy.

The negative connotations that have existed in the languages about privacy, as it turns out, is not really a differentiating factor for understanding privacy in these two languages; the underlying public–private (gong–si) division turns out to be more like a common theme on top of which privacy and its meanings were established. yin1si, yin3si, right to be let alone, consent, and so on are working vocabularies of privacy at different times and contexts; getting a handle on these vocabularies helps provide a better grasp of the meaning of the concept of privacy as it continues to evolve and change.

References


Examining Cooking Hobbyists’ Information-Seeking Behavior in Different Situations and in Different Stages from a Serious Leisure Perspective

Tien-I Tsai*, Hsuan-Yi Chen

Abstract: The current study investigates cooking hobbyists’ information behavior with emphasis on the sources used and information activities performed in different situations and in different stages. Drawing upon Stebbins’ serious leisure theory, Hektor’s framework of information activities, and Hartel’s cooking stages, a web survey was designed including a serious leisure scale, sets of questions asking various information sources consulted, and frequencies of information activities performed when “trying new recipes” and “making dishes once made before” in the “exploring stage” and “planning stage.” Four hundred eighty-seven cooking hobbyists participated in the survey. Results show that while most cooking hobbyists tend to obtain cooking information through online sources, but some cooking hobbyists with higher scores on the serious leisure scale also value print resources. They also engage in various information activities more frequently. Comparing to “dishes made before,” cooking hobbyists are more likely to seek information when “trying new recipes.” Nevertheless, the home recipe was the only source consulted more frequently when making dishes people once made before. Cooking hobbyists tend to browse and exchange information more frequently in the exploring stage when gathering cooking ideas rather than in the planning stage. Suggestions for food ingredients online shopping websites, recipe-recommendation systems, culinary bookstores, and special libraries were also provided and discussed based on the findings.

Keywords: information behavior, information activities, cooking, hobbyists, serious leisure

1 Introduction

Cooking for food consumption can be one of the physical survival needs as described in Maslow’s Hierarchy of Basic Needs. Eating home-cooked meals more frequently can help develop a better dietary quality (Mills, Brown, Wrieden, White, and Adams, 2017). During the COVID-19 pandemic, home cooking has been encouraged by several international organizations (WHO, 2020). In addition to fulfilling a basic need for food, cooking can also be a hobby that helps fulfill personal goals and live a healthy life.

According to Stebbins’ (2006, 2009) theoretic framework of serious leisure perspective, serious leisure, casual leisure, and project-based leisure are three major forms of leisure; among which, serious leisure is the systematic pursuit of highly substantial and fulfilling activities involving a combination of skills, knowledge, and experience. In the vein of serious leisure, according to Stebbins, amateurs, hobbyists, or volunteers acquire and express special skills, knowledge, and experience; specifically, while hobbyists have a lasting pursuit, they may not like amateurs who peruse a professional goal to become an expert. For cooking hobbyists, gathering or exchanging recipes and other information can be important to their leisure purposes either for skill development or experience sharing. Investigating cooking hobbyists’ information behavior can not only help them better developed their skills and fulfill their goals but also...
indirectly help further promote cooking itself, especially when highly engaged cooking hobbyists who love to share cooking information with others.

However, despite a series of research done by Hartel (2006, 2010, 2011), not much about cooking hobbyists’ information behavior was studied.

Hartel (2006) proposed a nine-step process of a gourmet cooking episode, including exploring, planning, provisioning, prepping, assembling, cooking, serving, eating, and evaluating. Based on Hartel’s findings, all stages, except for eating, may involve some information activities. Nevertheless, most stages manifest information use and reuse (e.g., prepping, assembling, cooking) rather than information seeking (e.g., planning). While Hartel also stated that exploring is “a quest for inspiration, not information,” the examples regarding what individuals do in the exploring stage included browsing cookbooks, websites, magazines, or personal recipe files, and consulting menus or other people imply that individual need to acquire information that helps inspire them. And acquiring information may involve information seeking. Therefore, the current study considered exploring stage and planning stage, the two stages that typically involve information-seeking activities. The former usually needs information to help generate cooking ideas; the latter usually needs to gather and identify recipes to help make the dishes.

Hartel, Cox, and Griffin (2016) used Hektor’s framework to mainly examine three different types of serious leisure activities, including the liberal arts hobby, amateur musicianship, and the hobby of running. Although the article did not discuss the hobby of cooking, they agree that Hektor’s information activities framework can be used to discuss other types of serious leisure activities. Hektor’s (2001) framework of information activities proposed eight information activities in four categories—information seeking, gathering, communicating, and giving. It depicts many aspects of individuals’ information behavior by examining search and retrieve, browse, monitor, unfold, exchange, dress, instruct, and publish. Among which, the search and retrieve, browse, monitor, unfold, and exchange are typically included in the discussions of individuals’ information behavior.

Although relevant information behavior research did not formally examine cooking situations, information behavior research tends to agree that novice and experts may demonstrate different information needs and its consequent information-seeking behavior (Case & Given, 2016). It would be interesting to examine how cooking hobbyists try new recipes for the first time and make dishes that they already had experiences with. The current study consolidated the above literature and designed a survey to investigate cooking hobbyists’ information behavior focusing on their source use and information activities in two different situations and in two different stages.

2 Methods

2.1 Data Collection

The current study used a web survey to investigate cooking hobbyists’ serious leisure characteristics and their information behavior. Cooking hobbyists’ serious leisure characteristics were measured by a serious leisure scale adopted from Gould, Moore, McGuire, and Stebbins (2008) developed based on Stebbins’ theory. Since Stebbins (2006) emphasized the six qualities in serious leisure—perseverance, career, effort, durable benefits, unique ethos, and identity—the current study adopted 18 items (three items in each of the six dimensions) from the short form of Gould et al.’s serious leisure inventory. The current study investigates information sources used by cooking hobbyists when they try new recipes and when they made dishes they once made before, and further investigates how frequently cooking hobbyists engage in four information activities (i.e., searching/retrieving, browsing, unfolding, and exchanging information) in exploring and planning stages of cooking. The information activities were derived from Hektor (2001), and the cooking stages were derived from Hartel (2006). The above instrument design of the current study only focused on information activities that are more relevant to cooking practices (e.g., search and retrieve, browse, unfold, exchange) and the cooking stages that may involve information behavior (i.e., exploring and planning stages). All items on an ordinal scale in the questionnaire used a 7-point Likert scale.

After two rounds of pilot testing, the current study distributed a web survey through various cooking blubs on Facebook, online forums, and other social media platforms. In addition to online platforms, the web survey was also distributed through researchers’ personal networks. Only those who have not been trained in a culinary arts vocational school or university and have not been a professional cook were recruited. Among the 506 participants, 487 valid responses were collected.
2.2 Participants

Among the 487 survey participants, most were female (76.39%). Slightly over half were young adults between 18 and 24 years old (53.8%), and only 15.19% were over 35 years old. Almost all participants were with a bachelor’s or a higher degree (94.87%), but only 39.84% had taken cooking classes for personal learning purposes. Given that the web survey was mainly distributed through cooking-relevant online platforms, and thus the participants were mostly students or young adults. This could possibly explain that even if participants all consider themselves as cooking hobbyists and slightly over one-fourth (26.9%) considered themselves cooking hobbyists for more than 5 years, only 11.91% cook every day and 60.57% cooked 2–3 days per week or once per week.

2.3 Data Analysis

After cleaning the data, Cronbach’s a was used to test the reliability of the serious leisure scale. The overall serious leisure scale reliability of the 18 items used in the current study is 0.922. K-means cluster analysis was then used to identify groups of cooking hobbyists based on their serious leisure characteristics. Two groups (high-score and low-score) were identified in the sample. Among which, 55% of the participants were in the high-score group (n = 269) and 45% in the low-score group (n = 218). The high-score group scored an average of 5.97 on the 7-point Likert scale with less variation (SD = 0.46), whereas the low-score group scored 4.69 with slightly larger variation (SD = 0.55). The high-score group scored highest on effort (M = 6.13, SD = 0.61) and the low-score group scored highest on perseverance (M = 5.06, SD = 0.87); the high-score group had a much higher score across all serious leisure qualities than the low-score group. The quality that received the lowest score from both groups was identity. However, the high-score group (M = 5.84, SD = 0.73) still rated much higher than the low-score group (M = 4.15, SD = 0.90).

Descriptive statistics were used to depict an overview of cooking hobbyists’ serious leisure characteristics and information sources consulted as well as frequencies of information activities performed. Chi-square tests were used to examine what sources were used in different situations (see Section 3.1); paired t-tests were used to investigate the differences in the frequencies of information activities performed between cooking hobbyists with high and low scores on serious leisure (see Section 3.2); and two-way ANOVA tests were used to examine the differences of the frequencies each information activity was performed in different situations (i.e., when trying new recipes and making dishes they once made before) and in different stages (i.e., exploring and planning stages) (see Section 3.3).

3 Results

3.1 Sources and Media Consulted in Different Situations

An overview of what sources cooking hobbyists consulted and the chi-square results regarding sources consulted in different situations are presented in Table 1. It is not surprising that cooking hobbyists consulted more sources when trying new recipes. Most cooking hobbyists consulted online sources, especially when trying new recipes. While most cooking hobbyists consulted recipe books for new recipes, less than half consulted recipe books when making dishes they once made before. When trying new recipes, slightly over half acquired information from TV programs (52.16%), about one-fourth consulted menu from restaurants and magazines/journals (26.90% and 23.20%, respectively); few consulted newspapers (13.14%) or other references (14.99%). According to the chi-square results (see Table 1), online sources, recipe books, menu from restaurants, magazines/journals, newspapers, and other references were more likely to be used when trying new recipes than making dishes once made before. While almost all the sources significantly tended to be used when cooking hobbyists try new recipes, home recipes were more likely to be used when making dishes they once made before.

When discussing cooking hobbyists’ information-seeking behavior based on media consulted, a similar phenomenon can be revealed. No matter social media, print materials, or traditional media, most cooking hobbyists tended to use them in a situation where they would like to try new recipes rather than making dishes they once made before (Table 2). Specifically, almost all participants in the current study consulted social media (93.01%), most consulted print materials (83.37%), and slightly more than half consulted traditional media (52.16%) when trying new recipes. On the contrary, only more than half consulted social media (65.50%) and print materials (58.52%), and slightly over a quarter consulted traditional media (26.49%) when making dishes once made before.

While most cooking hobbyists used social media and print materials in both situations, social media (c² (1, N = 487) = 36.63, p < 0.001, φ = 0.26), print materials (c² (1, N = 487) = 39.48, p < 0.001, φ = 0.27).
Tien-I Tsai, Hsuan-Yi Chen

Table 1
Sources Consulted in Different Situations by Cooking Hobbyists (N = 487)

<table>
<thead>
<tr>
<th>Sources</th>
<th>Try New Recipes</th>
<th>Dishes Made Before</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency (%)</td>
<td>Frequency (%)</td>
</tr>
<tr>
<td>Online sources***</td>
<td>453 (93.02)</td>
<td>314 (64.48)</td>
</tr>
<tr>
<td>Recipe books***</td>
<td>354 (72.69)</td>
<td>192 (39.43)</td>
</tr>
<tr>
<td>TV programs</td>
<td>254 (52.16)</td>
<td>129 (26.49)</td>
</tr>
<tr>
<td>Home recipes***</td>
<td>97 (19.92)</td>
<td>136 (27.93)</td>
</tr>
<tr>
<td>Menu from restaurants***</td>
<td>131 (26.90)</td>
<td>50 (10.27)</td>
</tr>
<tr>
<td>Magazines/journals***</td>
<td>113 (23.20)</td>
<td>41 (8.42)</td>
</tr>
<tr>
<td>Newspapers***</td>
<td>64 (13.14)</td>
<td>23 (4.72)</td>
</tr>
<tr>
<td>Other references***</td>
<td>73 (14.99)</td>
<td>35 (7.19)</td>
</tr>
</tbody>
</table>

***p < 0.001.

Table 2
Media Consulted in Different Situations by Cooking Hobbyists (N = 487)

<table>
<thead>
<tr>
<th>Media</th>
<th>Try New Recipes</th>
<th>Dishes Made Before</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency (%)</td>
<td>Frequency (%)</td>
</tr>
<tr>
<td>Social media***</td>
<td>453 (93.01)</td>
<td>319 (65.50)</td>
</tr>
<tr>
<td>Print materials***</td>
<td>406 (83.37)</td>
<td>285 (58.52)</td>
</tr>
<tr>
<td>Traditional media (e.g., TV)**</td>
<td>254 (52.16)</td>
<td>129 (26.49)</td>
</tr>
</tbody>
</table>

***p < 0.001.

3.2 Information Activities Preformed

Cooking hobbyists tended to perform several information activities quite frequently. Search and retrieve (M = 5.63, SD = 1.32), unfold (M = 5.51, SD = 1.22), and browse (M = 5.49, SD = 1.28) all ranged very high. This shows that cooking hobbyists not only actively look for information through searching and retrieving, but also consume information through browsing and reading or other ways to unfold the information. It seems that cooking hobbyists may not necessarily exchange information (M = 3.96, SD = 1.52) as frequently as the above information activities.

In general, cooking hobbyists with a higher serious leisure score tended to exhibit various information activities and be more engaged with information than those with a lower score (Figure 2). Comparing to the low-score counterparts, high-score cooking hobbyists tended to search and retrieve information more frequently (t(485) = −4.59, p = 0.000), browse information more frequently (t(485) = −5.49, p = 0.000), monitor information more frequently (t(485) = −3.99, p = 0.000), unfold information more frequently (t(485) = −4.48, p = 0.000), and exchange information with others more frequently (t(485) = −5.71, p = 0.000).

3.3 Information Activities Performed in Different Situations and Stages

The situations and stages were explained in the questionnaire so that participants can accurately reflect on the information activities performed in different situations and stages. The scope of stages was adopted from Hartel (2006). The exploring stage is when cooking hobbyists haven’t decided what dishes they would like to make and needed to look for inspiration; the planning stage is after decided what to make and need to plan for how to make the dishes. In both exploring and planning stages, cooking hobbyists perform all information activities more frequently when trying new recipes (Figure 3).
The two-way ANOVA results showed that the only significant interaction results happened in browsing ($F_{(1,486)} = 7.845, p < 0.01$). When in different situations, different stages have an effect on cooking hobbyists’ frequency of browsing information; when in different stages, different situations can also have an effect on their browsing. Although other items did not reach significant results in the interaction between situations and stages, there were many significant main effect results when it comes to different situations. Cooking hobbyists tended to perform the information activities, including search and retrieve ($F_{(1,486)} = 256.570, p < 0.001$), unfold ($F_{(1,486)} = 268.247, p < 0.001$), and exchange ($F_{(1,486)} = 17.829, p < 0.001$), more frequently when trying new recipes. However, there were not that many significant differences in the two stages. It seemed that cooking hobbyists typically performed similar information activities in both stages. The only significant results were in browse ($F_{(1,486)} = 22.408, p < 0.001$) and exchange ($F_{(1,486)} = 5.931, p < 0.05$). That is, cooking hobbyists tended to browse and exchange information more frequently in the exploring stage than in the planning stage. Since browsing and exchanging information are more likely to help to get new ideas, and it requires more inspirations when deciding what dishes to make, it is quite reasonable that cooking hobbyists tend to browse and exchange information more frequently in the exploring stage than in the planning stage.

When further analyzing cooking hobbyists’ serious leisure characteristics with the information activities performed in different situations and stages, we found that cooking hobbyists with higher scores performed all information activities more frequently in both situations and in both stages (Figure 4). This reflects the nature of serious leisure. Those who scored higher on the serious leisure scale are more frequently engaged in all types of information activities regardless of the situation or stage.

### 4 Conclusion

The current study investigates cooking hobbyists’ information behavior with emphases on the sources used and information activities performed in different situations and in different stages. While cooking hobbyists tend to obtain cooking information through online sources, cooking hobbyists with higher scores on the serious leisure scale value print resources more than those with lower scores. They also engage in various information activities more frequently. Compared to “dishes made before,” cooking hobbyists are more likely to seek information when “trying new recipes.” They tend to browse and exchange information more frequently in the exploring stage when gathering cooking ideas rather than in the planning stage.

The findings can help provide suggestions for various stakeholders that targeting cooking hobbyists as their audiences, such as food ingredients online shopping websites, recipe-recommendation systems, culinary bookstores, and special libraries. For food ingredients online shopping websites and recipe-recommendation system, since the current study found that browsing and exchanging information can be helpful when gathering cooking ideas, it is important for the website to proactively recommend recipes and relevant information to the users, and it could also be important to provide customized functions or interactive forums for users to exchange recipes or cooking ideas. For culinary bookstores and special libraries, continue providing high-quality print resources is essential, especially to those highly engaged
cooking hobbyists. Given that information services can be even more helpful when people would like to try new recipes, culinary bookstores and relevant special libraries can also try to design different types of cooking-theme exhibits so that it may help generate new cooking ideas and thus help promote relevant resources and information services.

The current study found that home recipes could be an interesting source that was the only source consulted more frequently when making dishes people once made before. Future research can further explore information behavior regarding how cooking hobbyists create and use home recipes and take personal cooking notes. This would also help investigate cooking hobbyists’ other information activities beyond the scope of the current research design (i.e., Hektor’s [2001] other three information activities not discussed in the current work—dress, instruct, and publish).

Additionally, the current study found that the interaction between situations and stages only had a significant result in browsing, but it is somewhat difficult to provide in-depth explanations. Mixed-method research designs using a survey with in-depth interviews or focus-group interviews may also help explain cooking hobbyists’ information behavior. It would be easier if cooking hobbyists’ personal accounts were also collected.

Finally, the ultimate goal of the current research stream aims to help cooking hobbyists promote cooking to help more people live a healthy life. With this in mind, future research can expand the scope to compare and contrast cooking hobbyist versus noncooking hobbyists’ information behavior, and further examine the information needs of people with different individual differences. This way, it could help the aforementioned stakeholders more directly design and provide various forms of relevant information services, and broaden the scope of culinary information services.

Acknowledgments: The authors would like to thank Drs. Muh-Chyun Tang and Ling-Ling Lai for their helpful feedback on an earlier version of this work.
References


Quality Assessment for Digital Stories by Young Authors

Abstract: Digital storytelling, an innovative way of writing, has been introduced to young learners who are taught to construct stories with digital tools to convey their knowledge and ideas. In 2018 and 2019, 31 digital stories created by Hong Kong primary school students were published on a digital story writing platform and linked from an online gamified reading platform. Each book on average gained 4,000+ views from across the globe and received 3,000+ favorable comments in total. While the digital stories are popular in these platforms, their quality and education value are uncertain. A review of the literature shows there is a lack of robust tools for assessing digital stories by young authors. The research team for this paper thus constructed their own framework in evaluating digital stories. An assessment of the stories has been done by two capable assessors, who found that the stories overall were of good quality and suggested room for improvement. This paper made three contributions: (1) “invention” of a digital story assessment framework; (2) it shows that stories created by students (with support from educators) can be an enjoyable and useful educational resource for their peers; and (3) digital storytelling can help foster the development of young authors.

Keywords: digital storytelling, e-books, digital literacy, quality assessment

1 Introduction

Writing can be a daunting task. In fact, it can also be an enjoyable task. Digital storytelling, integrating the use of multimedia into traditional story writing, has become a powerful technological tool in educational settings and has proved to be an effective way to engage both educators and learners (Wang & Zhan, 2010). This innovative way of storytelling has been introduced to young students, enabling them to write and publish digital stories online, sharing their ideas and work across the globe.

In recent years, around 40 Hong Kong primary schools have been using a gamified online reading platform “Reading Battle” to try to enhance students’ reading interest and abilities. In 2018 and 2019, top users who read more than 200 books and achieved at least 80/100 in answering the questions in these books were invited to create their own digital stories with support from a team of educators on an online platform “Storyjumper.” In all, 31 digital books were published by these students with a total of 127,094 reads as of April 20, 2020.

In this paper, we constructed an assessment framework to evaluate the quality of these 31 stories and investigated the educational value and practicality of digital storytelling for young students.

2 Literature Review

2.1 Digital Storytelling for Educational Use

Digital storytelling combines traditional storytelling with digital multimedia, including images, animations, recorded audio, and music (Lambert & Hessler, 2018). Like traditional storytellers, digital storytellers select a topic and create stories with their purposes and points of view. Digital stories are typically short which readers can finish reading within minutes, and is enhanced by...
technology while the story content remains as the primary focus (Boase, 2013).

In the previous decade, digital storytelling has been more commonly practiced in classrooms and has the potential to become a powerful tool for teachers and students’ collaboration and communication (Robin, 2008). Digital stories help teachers introduce ideas and knowledge more effectively by arousing students’ interest and curiosity toward topics (O’Byrne, Houser, Stone & White, 2018), while multimedia elements capture students’ attention, encouraging them to explore further. Robin (2006) also suggested that teachers can create stories that tailor their students’ needs and demonstrate how ideas can be conveyed using technology.

Under the teacher’s guidance, students create digital stories to share their ideas with peers. In addition to brainstorming ideas, conducting research, developing storyline and script, they also learn how to collect, analyze, and incorporate multimedia elements with text when creating digital stories (Robin, 2008). This equips students with twenty-first century literacy skills – digital, global, technological, visual, and information literacy, enhancing students’ effective communication skills in this innovative technology-driven world (Brown, Bryan, & Brown, 2005). Developing digital stories can also trigger students’ creativity, in transferring ideas into artifacts by combining elements to make them interesting for the audience, and enhance critical thinking through analysis and reflection during the process (Boase, 2013).

### 2.2 Current Models for Constructing Digital Stories


Robin (2008) from the University of Houston modified Lambert’s 7 elements into 10 for educational use. He intended to encourage students to consider the story’s purpose, quality of multimedia elements, and language use, on top of the seven elements, to make it more practical for teaching and learning purposes in the classrooms. These elements provide important references for teachers and students on producing quality digital stories, with the reminder that story literacy should be the top priority in digital stories (Ohler, 2013).

### 2.3 Assessment for Digital Storytelling

After constructing digital stories, it is essential to conduct evaluations to build stronger reflective and grading systems at schools. This can be done during the design, development, and completion stages, by both authors, peers, and teachers (Ohler, 2013). Teachers are found lacking professional knowledge on digital story evaluation (Aagaard, 2014). Assessment frameworks are therefore important to help them assess students’ work more effectively and objectively, and for students to comment on peers’ work and improve their own stories. Ohler (2013) proposed digital story evaluation traits for teachers’ development of rubrics in the classrooms, including story, writing, research, digital craftsmanship, media grammar, and creativity. Samples of digital storytelling rubrics are found and used to assess high-school projects (Barrett, 2006).

However, there are no robust models and most rubrics were designed for secondary to higher education. Guidelines or assessments for younger authors are lacking and therefore we developed a rubric to assess and evaluate their work.

### 3 Methods

#### 3.1 Participants

A total of 30 Chinese students ranging from Grades 3–7 (age 8–11) from different primary schools in Hong Kong participated in the study. These students were the top users and high performers in a gamified online reading platform “Reading Battle,” who read more than 200 books and attained 80/100 score in the reading game.

#### 3.2 Instrument

A digital storytelling assessment framework was developed in this study to examine the quality of digital stories produced by young learners. The three-step development was based on elements of digital storytelling (Lambert & Hessler, 2006; Robin, 2008) with adaptations made to accommodate the needs of young learners: first, a
draft was built from theoretical models. Assessment items were modified to be more age-appropriate after analyzing digital stories produced by authors at a similar age in the online story writing platform “Storyjumper.” For example, “emotional content” in the seven elements of digital stories was excluded in our framework, while language use was taken into consideration for our scoring. Second, this was reviewed by experts in language education, early years education, and digital learning. Third, five randomly selected stories were rated by two independent raters who discussed the scorings and revised the rubrics.

The final scale was structured in three domains: story content and language (includes purpose, organization and structure, language usage), the use of multimedia elements (includes digital effects, selection and quality of illustrations, audio narration, sound effects), and final product & presentation (includes overall length, creativity elements). Each item was rated on a 4-point scale: 1 = Needs improvement, 2 = Satisfactory, 3 = Good, and 4 = Excellent. Expected performance under each item was presented in detailed rubrics for scoring (Appendix A).

We applied different equations for the overall scores for different genres. For picture books, the weighting for the three domains (Story content and language: The Use of Multimedia Elements: Final Product & Presentation) was 6:2:2. Chapter books which were more textual based were weighted 8:1:1. Some items were only applicable to particular genres, for example, we did not score dramatic questions for non-fiction books.

### 3.3 Procedures

All students participated in 1–2 one and a half-hour training workshops on producing the digital stories and using the digital platform. Students were invited to write a digital story for students of similar or younger age, and they knew that these stories would be displayed on “Reading Battle,” which would be read by other students. They were free to choose whether they wanted to write the story in Chinese, English, or bilingual in both languages. They were provided with the same electronic platform “Storyjumper” to construct and present their digital stories with.

### 4 Results

#### 4.1 Story Content and Language Use

Students’ digital stories were first evaluated on story content and language use, which are the heart of effective digital stories according to Ohler (2013). Six areas were rated using the rubrics (Table 1):

Every story is created with a purpose, with reasons why authors decided to write and share (Greene, 1996). According to Fuhrken (2009), the major purpose of authors for elementary students included to persuade, inform, entertain, and describe. Among the 31 digital books, the most popular purpose is entertainment and enjoyment, telling an imagery story or real experience, focusing on characters and events, for example, “Friendship with a Dragon” and “The Tortoise’s Adventure” show adventurous journeys of the characters facing dangerous situations. Some stories have moral messages, for example, “A Fairy Tale Shuffle” sends the message of “All things in their being are good for something” and “The Ocean was Angry!” calls for environmental protection. Few students expressed the purpose explicitly, for example, the author of “My Invention Journey” wrote “I wish you could find your favorite hobby and keep doing, and there could be some surprising results”.

For the story structure and organization (Items 1b and 1c), all students included core story elements, which are the setting, characters, and plot. Some wrote about conflicts, struggles, climax, and resolution, making it more attractive and arousing readers’ interest. Students also demonstrated good language skills (Items 1c and 1d), most had appropriate vocabulary choices with accurate sentence structures and grammar, while there are two chapter books written with complex structures and advanced vocabularies. We noted that English books contained more grammatical mistakes than Chinese books.

Lastly, these books had a relatively low score on Dramatic question (Item 1f), which is the key question

<table>
<thead>
<tr>
<th>Item</th>
<th>Mean score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a. Point of view – purpose</td>
<td>3.16</td>
</tr>
<tr>
<td>1b. Story organization</td>
<td>3.42</td>
</tr>
<tr>
<td>1c. Story structure</td>
<td>3.33</td>
</tr>
<tr>
<td>1d. Language use – vocab and sentence structure</td>
<td>3.39</td>
</tr>
<tr>
<td>1e. Language use – grammar</td>
<td>3.68</td>
</tr>
<tr>
<td>1f. Dramatic questions</td>
<td>2.65</td>
</tr>
</tbody>
</table>

Average score under this domain 3.26

Scale: 4 Excellent, 3 Good, 2 Satisfactory, 1 Needs improvement
kept in the readers’ mind on how things will turn out, either by asking or prompting the audience to think about it at the beginning (Lambert & Hessler, 2018). Some stories are quite straightforward and do not raise questions to arouse readers’ curiosity, while most of the realization is barely different from the expectation.

Overall, the average total score for story content and language use was 3.26, which shows that students have strong capabilities in story construction.

### 4.2 The Use of Multimedia Elements

Thirty-one digital books were assessed on multimedia elements, considering Principles of Multimedia Learning developed by Mayer and Fiorella (2014). The results are shown in Table 2.

We examined the quality of the written texts. For the 31 digital books, some books showed texts in a conventional way, while some authors added text effects connected to the plot to facilitate understanding. For example, important vocabularies were highlighted in color with a bigger font size, and some dialogues are displayed with speech bubbles (see Table 3). This is consistent with the signaling principle that suggests that people read and learn better if we highlight significant information to direct readers’ attention (Mayer & Fiorella, 2014).

Multimedia and image principles suggested that students learn better with relevant images. Most books attained high scores in image selection and quality (Items 2b and 2c), with abundant images that enhanced engagement. They connect to the text and are relevant to the story and one author even drew her own pictures (see Table 4).

Audio including human voiceover in an informal and conversational tone positively helps the audience to get into the presentation, as stated in the voice and personalized principle (Mayer & Fiorella, 2014). Score for the audio components was low (items 2d-2g) because most books do not have a voiceover, as this is not compulsory. However, even though it was not introduced during training, six students added voiceover narratives. For example, authors of Friendship with a Dragon and The Adventure of a Monkey added their own voiceovers, the pace matched with the storyline and punctuation, facilitating the audience to engage in the story. They even added sound effects to the background to make stories become more engaging and unique.

### 4.3 Final Product and Presentation

Lastly, we reviewed the digital books in terms of duration (how long is the story), economy (the amount of information), and creativity (see Table 5).

To facilitate an effective presentation, authors combine content with multimedia to communicate ideas with readers and keep it as a short digital story which can be completed within 5 min (Harun & Shiratuddin, 2009). While 10 books are either too short or too long, the remaining ones can be finished within 5 min, with 10–25 pages in total. Within these pages, students were able to select and include the optimal amount of materials in the story without overloading the readers, which is always regarded as a challenging task for the authors (Lambert & Hessler, 2018).

Creativity is another bonus element of the digital stories. While all stories are original, some students managed to present their ideas in an unordinary way with the application of multimedia elements. Students’ creativity was shown from the surprise plot in the storyline, creative combination of pictures, and different sound effects made by authors (see Table 6).

### 5 Discussion and Conclusion

Realizing students’ capability to produce high-quality digital stories, it is worthwhile to encourage them to engage in digital story writing from an early age, using this innovative form of storytelling to communicate and share ideas with others.

First, it turns writing into an enjoyable task for students, giving them a more motivating and engaging way to communicate thoughts by adding different digital
media and skills into stories instead of just words alone. While writing has always been regarded as a traditional assignment in language classes, digital storytelling demonstrates a gamified way for everyone to tell stories and express themselves using various channels like pictures, songs, graphics, and so on. This fun way of learning helps motivate students, even for those who are struggling with language, to create stories and to practice their language skills and communicate with others in a less stressful environment.

Second, it helps foster students’ development and broaden their horizon by providing opportunities and channels to exchange their ideas and work with people all around the world, forming a globalized peer-centered learning environment. Students can create stories, publish them on online gamified platforms, and interact with their peers internationally. They read, answer questions, play games designed by the peers, comment, and rate the stories. This process turns students’ work into valuable educational resources which allow them to learn from peers online. It is believed this could enhance students’ engagement and produce higher quality digital stories in the long run.

Furthermore, the assessment framework developed can benefit teachers and students in several ways. The framework provides a well-defined and systematic outline...
for good digital stories and assists the trainers to educate students on writing digital stories. Trainers can instruct students on how to write good digital stories through the three domains and allow them to target the areas in which the students might need to improve. Students can refer to the guideline to improve their writing and also help assess other online stories, which can in turn facilitate peers’ learning. Lastly, this can also be used as an instrument for teachers to assess and grade their students’ work and offer suggestions for improvements.

6 Limitation and Improvement

The assessment framework was produced and used to analyze the 31 digital stories as explained in the paper. More digital stories should be chosen and assessed in order to further examine and modify the rubric. Besides, in order to validate the reliability of this new instrument, comparison of the scores rated by experts such as school teachers and librarians is needed to ensure that the framework is able to help differentiate the quality of digital stories in an effective and objective way.

Reference


## Appendix A

### Assessment Framework for Digital Stories Written by Young Authors

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>Item</th>
<th>Excellent (4)</th>
<th>Good (3)</th>
<th>Satisfactory (2)</th>
<th>Need improvement (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Story Content and Language</td>
<td>1a. Point of view – purpose</td>
<td>Purpose of the story is expressed clearly and explicitly, and remains the main focus throughout.</td>
<td>Purpose of the story is expressed clearly and remains the main focus for most of the time.</td>
<td>Purpose of the story is fairly clear but with a few lapses in focus.</td>
<td>The story does not have a purpose/message.</td>
</tr>
<tr>
<td>1b. Story Organization</td>
<td>The ideas are connected/organized coherently and cohesively. Paragraping is reasonable and logical.</td>
<td>Most of the ideas are connected/organized coherently and cohesively. Paragraping is mostly reasonable and logical.</td>
<td>The story is fairly organized, some ideas are not connected. Paragraping needs improvement.</td>
<td>The story is poorly organized, most ideas are not connected. Paragraping is not reasonable.</td>
<td></td>
</tr>
<tr>
<td>1c. Story Structure</td>
<td>The story includes the core elements: setting (When and where does it happen?), characters (Who are in the story?), plot (What happens in the story?). And it also has conflicts/struggle, climax and resolution.</td>
<td>The story includes the core elements: setting (When and where does it happen?), characters (Who are in the story?), plot (What happens in the story?). And it has at least one of the following: conflicts/struggle, climax and resolution.</td>
<td>The story includes the core elements: setting (When and where does it happen?), characters (Who are in the story?), plot (What happens in the story?). But there are no conflicts/struggle, climax and resolution.</td>
<td>It lacks one or more elements in the story structure: setting (When and where does it happen?), characters (Who are in the story?), plot (What happens in the story?).</td>
<td></td>
</tr>
<tr>
<td>1d. Language use – vocabulary and sentence structure</td>
<td>The author uses a wide range of advanced vocabulary in the story with a variety of sentence structures.</td>
<td>The author uses appropriate vocabulary in the story with accurate sentence structures.</td>
<td>The author uses appropriate but limited vocabulary in the story.</td>
<td>There are a lot of errors in the use of vocabulary and sentence structures.</td>
<td></td>
</tr>
<tr>
<td>1e. Language use – grammar (only applicable for English books)</td>
<td>There are few or no errors in grammar, spelling, capitalization, and punctuation.</td>
<td>There are some errors in grammar, spelling, capitalization, and punctuation.</td>
<td>There are lots of errors in grammar, spelling, capitalization, and punctuation.</td>
<td>There are too many errors in grammar, spelling, capitalization, and punctuation, which affect the writers’ understanding of the story.</td>
<td></td>
</tr>
<tr>
<td>1f. Dramatic questions</td>
<td>The story raises a question or prompts the audience to raise a question on how things will turn out. Realization is totally different from expectation.</td>
<td>The story raises a question or prompts the audience to raise a question on how things will turn out. Realization is barely different from the expectation.</td>
<td>The story raises a question or prompts the audience to raise a question on how things will turn out. Realization is the same as the expectation.</td>
<td>The story does not attempt to raise any questions on how things will turn out.</td>
<td></td>
</tr>
<tr>
<td>II. The Use of Multimedia elements</td>
<td>2a. Written text</td>
<td>More than one effect on the text is added, including size, font, color, or conversation box. They are connected to the plot and facilitate the understanding of the story.</td>
<td>One effect is added on the text, including size, font, color, or conversation box. They are connected to the plot and facilitate the understanding of the story.</td>
<td>No effect on the text is added to facilitate the understanding of the story.</td>
<td>No effect on the text is added to the story. The text is edited badly, e.g., the size of the words is too big or too small, the color is too light, or the font is difficult to read.</td>
</tr>
</tbody>
</table>
### Assessment Framework for Digital Stories Written by Young Authors

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>Item</th>
<th>Excellent (4)</th>
<th>Good (3)</th>
<th>Satisfactory (2)</th>
<th>Need improvement (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>II. The Use of Multimedia</td>
<td>2b. Images – selection</td>
<td>All images selected add impact to the story and enhance engagement. They are</td>
<td>Most of the images add impact to the story and enhance engagement. They</td>
<td>Some of the images add impact to the story and enhance engagement. Some of</td>
<td>There is no image in the story. Or most of the images are not relevant to the</td>
</tr>
<tr>
<td>elements</td>
<td></td>
<td>connected to the text and relevant to the story.</td>
<td>are connected to the text and are relevant to the story.</td>
<td>them are not connected to the text and are relevant to the story.</td>
<td>text and the story.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2c. Images – quality</td>
<td>All images are clear with high resolution, consistent in style, and placed in</td>
<td>Most of the images are clear with high resolution, consistent in style,</td>
<td>Most of the images are clear with high resolution, but some of them are not</td>
<td>Most of the images are fuzzy with low resolution, inconsistent in style and placed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>the appropriate position.</td>
<td>and placed in the appropriate position.</td>
<td>them not consistent in style or placed in the appropriate position.</td>
<td>in an inappropriate position.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2d. Audio – voiceover’s quality</td>
<td>The voiceover is clear and consistent.</td>
<td>The voiceover is clear and consistent in most parts of the story.</td>
<td>The voiceover is clear and consistent in some parts of the story.</td>
<td>There is no voiceover. Or the voiceover is in low quality.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2e. Audio – voiceover’s style</td>
<td>The author speaks in a conversational style (either colloquial or written</td>
<td>The author speaks in a conversational style (either colloquial or written</td>
<td>The author mainly uses a monologue or automated style.</td>
<td>There is no voiceover.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>language for Chinese version) for most part of the story.</td>
<td>language for Chinese version) for some parts.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2f. Audio – voiceover’s pace</td>
<td>The pace including rhythm and punctuation fits and supports the storyline and</td>
<td>The pace including rhythm and punctuation relatively fits and supports the</td>
<td>The pace including rhythm and punctuation partly fits and supports the</td>
<td>There is no voiceover. Or the pace including rhythm and punctuation does not fit or</td>
</tr>
<tr>
<td></td>
<td></td>
<td>facilitates the audience to engage in the story.</td>
<td>storyline and facilitates the audience to engage in the story.</td>
<td>storyline.</td>
<td>support the storyline.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2g. Audio – music/sound effect</td>
<td>Appropriate music or sound effects are added to the story. All of them match</td>
<td>Music or sound effects are added to the story. Most of them match with the</td>
<td>Music or sound effects are added to the story. But only some of them match</td>
<td>No music or sound effects are added into the story. Or they are all inappropriate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>with the storyline and help set up an emotional mood related to the story.</td>
<td>storyline and help set up an emotional mood related to the story.</td>
<td>with the story.</td>
<td>and distracting.</td>
</tr>
<tr>
<td>III. Presentation/Final product</td>
<td>3a. Duration/number of pages</td>
<td>It can be finished within 5 min. It contains 15–20 pages.</td>
<td>It can be finished within 5 min. It has 10–14 or 21–25 pages.</td>
<td>It cannot be finished within 5 min. It has more than 25 pages.</td>
<td>It can be finished in less than 5 min. It has fewer than 10 pages.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3b. Economy</td>
<td>The story is told with optimal amounts of materials. It does not overload the</td>
<td>The story composition is good, but there are a few extraneous, distracting</td>
<td>The story composition is fair, but there are lots of extraneous and distracting</td>
<td>The story composition is poor. The materials are either extraneous and distracting</td>
</tr>
<tr>
<td></td>
<td></td>
<td>audience.</td>
<td>materials, or more materials can be added to support the story.</td>
<td>materials, or more materials should be added to support the story. Further</td>
<td>and distracting to overload the readers or not enough to convey the message of the</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>editing is needed to make it better.</td>
<td>story. Extensive editing is needed.</td>
</tr>
</tbody>
</table>
### III. Presentation/Final product

#### 3c. Creativity

The story is original and fulfills two or more of the following creativity requirements:

- (a) New ideas/surprising plot in the storyline
- Creative ways in the application of multimedia elements:
- (b) Creative combination/edit of pictures
- (c) Author’s voice is used to make different sound effects
- (d) Music/Sound effect is added
- (e) Others: (please specify)

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>Item</th>
<th>Excellent (4)</th>
<th>Good (3)</th>
<th>Satisfactory (2)</th>
<th>Need improvement (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>III. Presentation/Final product</td>
<td>3c. Creativity</td>
<td>The story is original and fulfills two or more of the following creativity requirements:</td>
<td>The story is original and fulfills one of the following creativity requirements:</td>
<td>The story is original and the ideas are presented in an ordinary way with the application of multimedia elements.</td>
<td>Ideas of the story are copied from others, and it is presented in an ordinary way with the application of multimedia.</td>
</tr>
</tbody>
</table>
Analysis of User Social Support Network in Online Tumor Community

Hui Lin, Shijuan Li*

Abstract: With the development of Internet technology, online health forums have become indispensable for people who seek non-professional health support. This research focuses on the content posted by cancer patients and their relatives in online health forums and social networks to raise the following research questions: What is the overall view of the social support network in the online tumor community? What are the information behaviors of the online tumor community in different identities of users? How users interact in this community and build this network of social support? What are the topics users would like to share and talk about? What kinds of users could be the key users in this community? Method: Using the post and comment data of the Oncology Forum of Tianya Hospital in 2019, combined with social network analysis and word co-occurrence network analysis, the following conclusions are obtained: (1) There are some central points in the overall social support network, and there are central users consistent with other social networks. (2) Positive users are more likely to comment on others, and it is easier to get others' comments, while negative users are more likely to share personal information and do not want to participate more in social interaction. (3) Users focus on posting emotional and emotional content in content sharing. Information-based social support information. The social support experience that this type of information brings to users can be positive and negative. (4) The most active group in the patients' online health community, followed by the patients' children. (5) The relationship between users and patients is diverse and there are two types of singularity. Users with diverse relationships are more likely to be commented on, and they are more willing to comment on users who also have diverse relationships.

Keywords: online tumor community, cancer survivor, social support network

1 Introduction

With the improvement of living standards, people pay more and more attention to health. In addition to regular visits to the hospital for consultations, the use of mobile Internet technology to query health-related information online is becoming more and more frequent, which has also promoted the demand for health information research. From the perspective of health information content, health information needs can be divided into knowledge type, news type, data facts, and data type information. Through individual information queries, people can obtain more types of health information, which can make up for people and professional medical services. The knowledge gap between personnel serves as a knowledge supplement for medical information. The sources and channels for people to obtain health information mainly included authoritative resource databases, related health websites, discussions, and consultations resulting in the social circulation of health information, interpersonal communication, organizational communication, and mass communication. With the help of social media, the interpersonal communication of health information has developed rapidly.

The development of online communities and social platforms and the continuous improvement of citizens' health awareness have promoted the rise and development of online health communities. Faced with diseases that require long-term care, people who are left with only a limited number of exchanges with professional medical
personnel, and also because of various problems faced in the nursing cycle, their health information needs are not met in time. Under the imbalance of supply and demand, the online health community provides an effective informal communication and interaction platform. People can obtain health information more quickly through the online health community, and at the same time can share their experience and exchange information with users with similar experiences to meet health information needs.

Social support is an important part of users' health needs in online health communities. The theory of social support was born in the 1970s. It is a part of the relationship between the individual and the social environment. It helps individuals adapt to the environment. This model can be integrated into ecological theories and life models, focusing on obtaining resources from personal relationships. With the circulation of various resources in the social system, including formal social resources and informal social resources, formal social resources come from various institutional support from the government and formal social organizations, whereas informal support mainly comes from family, relatives, and friends, and some informal organizations who come forward to guide with a comprehensive support network for the vulnerable groups. Huang (2013) mentioned in the research on the social support network of cancer patients from the perspective of social work that the current social support network provided for cancer patients has a lack of institutional guarantees, lack of professionalism, single-sided support content, and insufficient degree of socialization of support forces. The existing social network is far from enough to help cancer patients cope with multiple pressures.

Based on this, this article raises the following research questions: What is the overall view of the social support network in the online tumor community? What are the information behaviors of the online tumor community in different identities of users? How users interact in this community and build this network of social support? What are the topics users would like to share and talk about? What kinds of users could be the key users in this community?

2 Literature Review

Reviewing users' information behaviors in online health communities, there are several kinds of behaviors by Zhang and Xue (2020), their literature review of influencing factors of user participation behavior in the online health community, such as information privacy concerns, seeking, asking, hiding, sharing, acceptance, using, responding, patients choosing doctors and social support behaviors. These behaviors of users in online health communities are researched by a number of scholars who are repeatedly analyzing data from a single aspect. But personal activities in online health communities are various. We could not think of research questions from one side, but it should be many-sided, just like information sharing behaviors, it's about information flowing from person to person, but this is essentially a one-way output of information. As social support behaviors, these are interactions in networks of people, which include the multi-directional flow of energy, such as information, emotion, and so on.

Research on the categories of information participants in online health communities. Huh et al. (2016) found that there are four main types of roles in online health communities: Caretakers, Opportunities, Scientists, and Adventurers. Zhang and Xue (2020) believe that the information subjects in the online health community can be roughly divided into two categories. One is the non-professional knowledge of patients and their families, and their information needs and participating behaviors are relatively similar. The second category is professional users such as doctors and nurses, who are the main source of community expertise. There are various types of participating behaviors: diving behaviors, information seeking or searching behaviors, information disclosure behaviors, sharing behaviors, adoption behaviors, usage behaviors, and social behaviors among users. The above behaviors are different types of user behaviors, contributing behaviors, social support behaviors, and patients' doctor selection behavior, etc.. Yang and Zhu (2020) believe that based on the information ecosystem, online health communities mainly include information people, information, information environment, and information technology. Research by Sharma and Khadka (2019) showed that information support is the main type of support people seek from online social health support communities. Realizing the complexities of information behavior and the needs in the online health community, it makes us to ask more and more questions. In our data set, how much information needs to be focused on? How many people keep connections with others? When people were commended by others what topics they wanted to discuss? When they commended others, what topics they wanted to ask?

Now that we have paid a lot of attention to information behaviors and information participants, let's discuss these themes more specifically in social support backgrounds.
Reviewing the related literature on the types of social support, Liu, Li and Wang (2004) defined social support as three dimensions in the survey of the social support status of cancer chemotherapy patients: objective support, subjective support, and support utilization. Bambina (2005) proposed that the types of social support in online support groups mainly include emotional support, information, and partnership. The social network of the online community forms an effective combination, consisting of a dominant actor, responsible for connecting the highly forked network. Shi, Li, Qian, Zhou and Zhang (2019) proposed that the main types of users’ social support needs in health question-and-answer communities at home and abroad are: information support, emotional support, and practical support. Wang (2015) found through an investigation of a health group on Douban.com that the network structure formed by information support is relatively loose, and though members participate in topic discussions the interaction between the network members is not high, and the network formed by emotional support and companion support has closer membership; in the network formed by the topic initiated by the dentist, all that the network members exchanged are information support content, and the patient in addition to information support, members of the initiated network will also provide emotional support and companionship support, and the patient network will be more closely connected. Besides, a study on Baidu “HIV Bar” found that the demand for social support is far greater than the supply of social support; and compared with emotional and realistic social support, information-based social support is the most frequently sought and provided type.

Those studies of social support focus on categories and structures of social support and its network. How about the people’s behaviors in online health communities? As we all know that as users in one community, they must have information exchange and build some connection with others. When we dig deeper, we could find that people have different roles to play in the online health community.

Recalling the important role of social support networks in online health communities, Yang, Zhong, Kumar, Chow and Ouyang (2018) proposed that in online health forums, both the provision and acceptance of social support by users greatly encourage continuous social support exchanges. Liu, Fan, Ji and Jiang (2019) proposed that community support has a positive effect on the interaction between health topics and general topics. Yang, Du, Wang and Wu (2019) proposed that the scale of social support networks and the activities of individual and group members within the network are positively correlated with the level of users participating in online health tasks, and the interactive relationship between social support networks, individual user activities, and user participation are complementary. Fullwood, Chadwick, Keep, Attrill-Smith, Asbury and Kirwan (2019) explained that the people lurking in the online health support group may get more benefits than those who actively participate, and achieve their goals without contributing.

We have already done more work to review others works about information behaviors and social support related topics. What about those research in specific populations such as cancer patients and their caregivers?

In response to related research on cancer patients in online health communities, Westmaas, Fallon, McDonald, Driscoll, Richardson, Portier and Smith (2020) investigated the behavior of American cancer survivors in online support communities and found that Conclusions Engagement in online survivor communities may increase support perceptions that promote well-being, but benefits may accrue more to survivors reporting low offline social support. Research by Weiss, Berner, Johnson, Giuse, Murphy and Lorenzi (2013) suggests that the online support community for cancer survivors should resolve the interdependence between online support and actual support, reduce social isolation and enhance people’s social interaction in the real-world a sense of control of activities and the reintegration into the social environment may require the use of larger loosely connected networks. One should also simultaneously handle personal existing social networks such as family and friends at the same time. Social support in the existing social networks has shown a positive impact on health, while negative support in social networks is related to the negative impact on health.

From the above literature reading, it can be found that although the existing research has covered user reviews and behaviors in online health communities and research on user social support, there is a lack of research combining multiple perspectives. This research is based on a cross perspective study. Set out with the user comments and the commented behaviors of users in social support groups. At the same time, while constructing the user social support network, it further explores the relationship between users and patients, the overall structure of the social support network, and user comments and comments, the influence of behavior and content.

The theoretical contribution of this research lies in the integration of a single research problem domain that has been sufficiently researched into other problem domains for cross-research, and discovering the patterns of user behavior in different contexts. In social practice, this
research can help professionals improve the personnel structure in the online cancer community, publish professional knowledge through more influential users, provide information assistance in the social support network, and promote more users who can participate in the forum and find the information you need, and get better social support through emotional expression.

3 Data Collection and Analysis

In 2020, Tianya Community announced that there were more than 130 million registered users and more than 250 million monthly users. On September 17, 2001, the Tianya Hospital section was opened in Tianya Community, with the aim of “Netizens ask questions here, and doctors popularize science here.” Tianya Hospital focuses on patient consultation posts, including internal medicine, surgery, obstetrics and gynecology, pediatrics, infectious diseases, oncology, dermatology, ENT, traditional Chinese medicine, psychology, beauty, health, urology, and laboratory. In the content section, users share their own or others’ treatment experience and open posts, and other users can follow and comment in the main post.

In the forum section of Tianya Hospital, the author used the python web crawler to crawl all the posts published by the Oncology Department in 2019. There were 476 main posts, 7,859 posts, and 1,530 participating users. Users in the forum, commenting and posting each other from the perspective of existing literature, are all types of social support, whether it is from an emotional perspective or an information perspective, in this informal communication environment, through the identification of the structure of interaction between users can produce an understanding of the overall structure of the connections between users in the online health community. Use python for data analysis, follow the posts and comments between users, call the Pychars python visualization library to build a social network graph, and view the relationship between user comments and the commented. Use the word co-occurrence network to find out the relationship between user comments and the words in the comment, and find the correlation between the words most frequently used by users, extract the words after the topic according to TF-IDF, and perform word frequency statistics, and the frequency of occurrence will be greater after selecting the user clusters with rich interaction and establish the association between user sets and words.

Automatically segment the collected text data to describe the user’s identity. An example is illustrated as follows: A father is suffering from cancer, and the user can be judged as the patient’s child. He seeks social support in the forum. After segmentation, the user’s self-description is assessed by manually judging and determining the identity of the user, including children, parents, friends, relatives, patients, brothers, and sisters.

Among the 1530 users, 961 were thus successfully identified. The rest of the data is difficult to identify or treated as advertising information, and so it is not considered for this study. In the identification, a user can have multiple identities. This is because, in the process of description, the user may mention different members of the family or friends suffering from cancer in different replies. Based on the user’s posting data, the number of posts posted by each identity is counted, and a picture is drawn. At the same time, based on the relationship between the comments in the online health community and the commented, combined with the user ID to construct a social network, focusing on users who actively comment on others, a network is constructed, and also focusing on users commented by others, networks are built to discover the similarities and differences between the two types of networks.

4 Conclusions

4.1 User Social Support Network

4.1.1 User Social Support Network Overview

The overall user relationship network in the online health community is shown in Figure 1 below. Among the collected users, the behaviors of users’ postings and other users’ posting comments are established. Users who have posted a large number of posts have a larger central point, but fewer posts. Every user in social support network as a red point, which big or small point means the frequencies of communication. As we can see, users in the community have built multiple central social support networks. Users with a large number of posts are often connected with others, and users with a fewer number of posts are often isolated or at the end of the network.

4.1.2 Examples of Key Users of the User Social Support Network

From the perspective of users who choose to be in the core position of the network in the social support network, as
shown in Figure 2, the connection between users is not only directly related through postings and comments, but also through their interactive behaviors with other users who are willing to participate in the interaction to establish contacts to expand their social support radiation network.

### 4.1.3 User Comments and Commented Networks

In the user social support network, a social network view is established for the subject of users actively commenting on other users, as shown in Figure 3, and a social network view is established for the subject of the user being commented by other users, as shown in Figure 4. By comparing the two images, it can be found that the social support network in Figure 3 has fewer isolated points, which means that in the online health community, users who actively comment on others are more willing to actively respond to comments and establish contact with other users.

There are more isolated points in Figure 4, which indicates that such users are in a passive state when using the online health network. Although they are commented, they are unwilling to actively respond to the content of the comments and have less contact with other users.

### 4.2 Users Posting Content Combined with Their Social Support Network Analysis

#### 4.2.1 User Posting Content Statistics

From the content point of view, the data collected by users in the data set is distinguished, as shown in Figure 5. The most discussed type of cancer is lung cancer. In the 2019 Cancer Report released by the National Cancer Center, lung cancer is the most susceptible cancer type for men, while breast cancer is the most susceptible cancer type for women. In terms of the number of cancer types, in 2019, the online health community in Tianya Hospital has the largest number of discussions about lung cancer, which may reflect that the forum uses more male users and their relatives and friends.
4.2.2 User-posted Content Word Co-occurrence Network

A word co-occurrence network is established for the user’s posting information in Tianya Hospital, as shown in Figure 6.

In the word co-occurrence network, users discuss the treatment information of cancer itself. You can see that words with high co-occurrence rate are doctor, surgery, chemotherapy, cancer, tumor, patient, Chinese medicine, hospital, health, and other words, which is reflected online. The main type of social support in the health forum is the information support, and the words that rank second in the co-occurrence rate are transfer, father, mother, liar, patient, thank you, treatment, medical treatment, common sense, and other words, and is indicated in the online health community. In addition to the patients themselves, their relatives, especially their children, will pay more attention to disease-related information, and the appearance of words such as liars also indicates that some information in the online health community is doubtful. The appearance of words such as thank you indicates that the user’s communication with other users in the community can meet some of their social support needs and has received a good response, indicating that there is a benign social support network in the online health community.

4.2.3 Identified Users and User-posted Content Word Co-occurrence Network in Social Networks

Combine the user and content dimensions as shown in Figure 7. It shows that the connections between users are connected with the vocabulary commonly used by the main users. You can see that the main content words of users who have been in contact with other users are Pathology, Cancer, Lung Metastasis, Western Medicine, Thank you, Help, Support, Stories, etc. These words indicate that users who actively participate in social interactions are more willing to share disease-related information, or anti-cancer stories and other contents. In this sharing, the high frequency of thank you and mutual assistance shows that users are actively contacting others in the community, which is beneficial to their acquisition of social support.
4.3 Relationships of Supporters and Patient, and Social Network Analysis

4.3.1 User-patient Relationship

From the 1,530 users posted by the Oncology Department of Tianya Hospital Forum 2019, the program was used to segment user identity description fragments, and after manual identification, 961 users who can identify the relationship with the patient were distinguished. According to the details and description in each content posted by each user, the relationship of the patient is summarized and obtained: the patient himself, the patient’s friends, the patient’s relatives, the patient’s spouse, the patient’s children, and the child’s spouse, and they belong to five main types. Count the number of users who posted, and get a classification diagram of the relationship between the users who posted and the patients, as shown in Figure 8. In the Oncology Forum of Tianya Hospital, among the users who participated, patients accounted for the most, accounting for 81.89%, and the patient’s parents accounted for the least, at 3.02%. It is an irony that the number of posts posted in the forum by the patient’s friends is more than those posted by the patient’s spouse.

4.3.2 Posting Users Are Commented on Network Analysis

After identifying the relationship between the poster and the patient, from the perspectives of actively commenting on others’ posts and being commented on by others, the proportion of users who are differently associated with the patient is considered. As shown in Figure 9, as shown in Figure 10.

Users who are more commented by others have diverse relationships with patients. As shown in Figure 9, the relationship between core user 12363744 and patients is follows: children, friends, patients, relatives, he is a patient, and his parents also have cancer. As shown in Figure 10, the relationship between the core user 136658289 and the patient is as follows: children, friends, patients, parents, indicating that he is a patient, and his parents and children I have also experienced cancer with my friends. The size of the user’s dot in the figure represents the number of times the user has been replied to the comment. The connection with other dots indicates that the user is replied by other users and a connection is established between the two.

The selected example is the user who has received the most comments from users in the whole picture. Observing the similarities of the two, we can find that two core users have a diverse relationship with the patient, while the other non-core users associated with them often have a single relationship with the patient.
Analysis of User Social Support Network in Online Tumor Community

4.3.3 Posting Users Actively Comment on Other People’s Network Analysis.

Establish a network with users actively commenting on others, and observe how users who have different relationships with patients in this network actively contact others. As shown in Figure 11, the core user is 12363744, whose identities are as follows: children, patients, friends, relatives, that is, the user himself, his parents, and relatives and friends have all experienced cancer.

In the network of actively commenting on others, the size of the dot represents the number of times that they have actively commented on others, and the connection with other dots represents the association between the user and the user who is commenting. The difference from the network of the commented user is that among other users commented by core users, users with diverse associations with patients account for the majority, while users with single associations with patients account for a minority. The core user characteristics in Figure 12 are the same as those in Figure 11.

5 Discussions

5.1 User Social Support Network

5.1.1 Multiple Centers

Based on the above analysis, it can be found that in the data concentration, the overall social support network has some central points, and there are central users consistent with other social networks. The posting and contact of these users improve the accessibility of the social support network, and some are at the end of the network. Users can find more social support network nodes by establishing contact with users with strong centrality.

5.1.2 Two Ways of Interaction: Positive and Negative

There are two types of user social support networks, positive and negative users. Positive users are more likely to comment on others and get comments from others, while negative users are more likely to share personal information and do not want to participate more socially. This may be because different users have different attitudes toward online health communities. Actively participating users use online health communities more as a social platform, where they find people with similar experiences, provide social support to others, and get social support from others. Support brings more beneficial experiences to active users so that they are willing to maintain an open attitude, and the types of social support they receive are more likely to be emotional.

However, users who passively participate in social interactions use the online health community as a record-type blog platform, posting their anti-cancer experience or their relatives and friends, or publishing popular science information, etc., focusing more on the continuity of the content in the community, the online health community. The types of social support provided to such users may be more informative, which prevents such users from actively establishing contact with others. Zhu, Guan and Donovan (2020) analyzed the
communication behaviors of opinion leaders in two large cancer-centric online health communities and found that the typical responses of opinion leaders to others are often a combination of opinion support, emotional support, and network support. In this study, active users are willing to share information about cancer history and treatment options and provide positive emotional support and comfort.

5.2 Analysis of User Release Information in User Social Networks

It can be seen from the content that users focus on publishing emotional and informational social support information in content sharing. The social support experience that this type of information brings to users can be positive and negative, especially in the community’s treatment of Chinese medicine for cancer. Similar issues may indicate that emotional social support in online health communities is more likely to bring positive reviews, while information-based social support may cause users’ bad evaluation of social support networks because of the unevenness of the information itself. Davis, Du and Wang. (2017) found that online health communities may not have a positive impact on health because anyone can provide information and the information provided may be invalid. In this study, when sorting out the posts of online health community users, we also saw many posts or comments about Chinese medicine or folk prescriptions or a doctor’s treatment that have a good effect on cancer treatment, which has caused controversy.

5.3 Association Analysis between User-patient Relationship and User Social Support Network

After identifying the relationship between the user and the patient, it was found that in the Tianya Hospital Oncology Forum, the relationship between each user and the patient is diverse, indicating that this community has gathered people related to cancer patients to discuss. Patients themselves are the group that participates in the most discussions, indicating that patients are most concerned about their physical conditions and expect to receive social support from people related to their conditions in the online health community. Secondly, the children of patients have posted more posts in the health forum, indicating that caregivers of cancer patients also need to seek social support.

At the same time, because cancer patients are more common in the middle-aged and elderly people, their children have a certain degree of information and communication skills. They can use online health forums to seek social support to give better care for parents with cancer. In the course of the research, it is surprising that the patient’s friends participate in discussions more frequently than the patient’s parents and spouses. There are two potential possibilities. One is that the patient’s friends participate in discussions in online health forums out of concern for friends and also because of their curiosity to know about cancer. The other is when patients or their family members talk about cancer in the online health community. For the protection of personal information, they usually use the identity of the patient’s friend to participate in the discussion to hide their true identity. Therefore, some patients and relatives of patients may be classified as friends of patients due to textual reasons.

Recognize the relationship between user comments and the commented relationship and the relationship between users and patients, and eliminate unidentified users or users who provide advertisements. After establishing a commented network for the identified users, it is found that there is a variety of online health forums with patients. The users of the personalized relationship can often get more comments and the attention of more participants. This may be because users with diverse associations with patients have published more posts in health forums, with richer content, and more experience in cancer. Such users have more content and experience that can be of use for other users. Based on their reference and credibility, other users are more willing to reply to

![Figure 12. Example 2 of the user who posted actively commenting on others' networks. (Part of Figure 3)](image-url)
messages to provide social support or seek help. However, users with a relatively single identity may not publish as much content and follow as users with diversified identities; therefore, the number of replies and comments is less compared with others.

However, users with a single relationship may be able to solve their problems after paying attention to users with diversified relationships, and therefore they are less proactive in establishing contact with others. In the network of actively commenting users, it can be found that users who have diversified relationships with patients are more willing to comment on others, which is consistent with the situation in the commented network. More comments on others may get more comments from others. Yes, the users they comment on are more users who have diverse relationships with patients. This may be because users who have diverse relationships with patients have a certain similarity, and their social support needs are similar, and so they are more likely to have connections. It may also be that such users have diversified connections with patients, who are more experienced and are willing to proactively provide help to similar users.

6 Summary and Future Works

Here, we mainly studied the content of the posts posted by users in the oncology forum of Tianya Hospital in 2019, analyzed their content and the relationship between comments and commented, and conducted research using social networks, artificial coding, and word co-occurrence network methods. The research on the association between users and patients in social support networks in online health forums has been done.

However, in the process of research, the identification and filtering of users and patients’ identities is a combination of procedures and manuals. Some users that are not identified by the procedure are excluded. In the research, these users are listed as other users and have not been studied. User information includes the types of specialized science popularization and medical advertisements. In the user’s social support network, specialized science popularization is also an important type of social support. In this study, this part of the data was not studied.

In the following research, we will supplement this part of the data, analyze and research the advertising information and controversial information in the online health community, and further expand the time range of the research data and expand the research platform to improve this research.

Acknowledgements: This work is funded by Beijing Social Science Fund, PR China (18XCB007).

References


Westmaas, J. L., Fallon, E., McDonald, B. R., Driscoll, D., Richardson, K., Portier K., & Smith T. (2020). Investigating relationships among cancer survivors’ engagement in an online support community, social support perceptions, well-being, and moderating effects
of existing (offline) social support. *Supportive Care in Cancer, 28*(8), 3791–3799.


Anrong Fan, Qiao Wu, Xiaofei Yan, Xiaotong Lu, Yue Ma, Xue Xiao*

Research on Influencing Factors of Personal Information Disclosure Intention of Social Media in China

https://doi.org/10.2478/dim-2020-0038
received August 15, 2020; accepted September 15, 2020.

Abstract: The disclosure of personal information by users is very important for social media, in order to balance privacy protection and personalized service. This article probes into the factors influencing users’ disclosure intention. Based on the privacy calculus theory and theory of planned behavior, the study constructs an influencing factor model of social media personal information disclosure intention. Then an extensive survey of social media users is conducted through questionnaire, and the hypothetical model is verified using structural equation model, and finally the relationship between various influencing factors and personal information disclosure intentions is obtained. The results show that the perceived benefits and subjective norm are related to personal information disclosure intentions, and privacy view is associated with perceived risk. Finally, the study provides new ideas for social media services and user privacy protection, such as creating a secure social media environment, increasing valuable social services, reducing users’ risk perception and making information processing open and transparent.

Keywords: personal information disclosure intention, privacy calculus theory, theory of planned behavior, influencing factor model

1 Introduction

Personal information refers to all kinds of information recorded by electronic or otherwise that can be used to independently identify or be combined with other information to identify specific natural persons (National People’s Congress, 2020). It is always associated with the identity of a natural person in the sense of physical, physiological, genetic, mental, economic, cultural, or social (European Unit, 2018). From the above, it can be seen that personal information can be identified and closely linked with the information subject, which involves the field of privacy life. Once abused by a third party, it will cause great privacy infringement to the information subject.

Disclosure of personal information widely appears in various situations in human social life. In recent years, with the wide application of the mobile Internet, social media has become one of the important personal information disclosure scenarios. Posey, Lowry, Roberts and Ellis (2010) stated that the self-disclosure behavior of online community users refers to the behavior of users leaking personal information when registering or using mobile social networks. The term “social media” originated from What is Social Media written by Antony Mayfield in 2007; the author defines social media as “a new type of online media with the characteristics of participation, openness, conversation, community, connectedness, that gives users great participation space.” In the process of the gradual development and growth of the mobile Internet, the application scenarios and types of social media have further increased. The 2016 China Social Application User Behavior Research Report (CNNIC, 2017) divides social applications into three main types: instant messaging tools represented by WeChat and QQ, comprehensive social applications represented by Sina
Weibo, and vertical social applications represented by Douyin (overseas version: Tik Tok) and Zhihu. Although the boundaries of social media are becoming blurred, this classification still has some value.

In the process of using social media, users often need to actively accept the disclosure of personal information. Research by Nie and Luo (2013) pointed out that the disclosure of personal information on social media by users is an important issue that needs to be weighed. Although more disclosure of personal information on social media helps to make full use of the advantages of the platform to meet their own needs, increasing the disclosure of personal information will correspondingly increase the risk of personal privacy leakage. The deepening of this contradiction has further led to the paradox phenomenon of current social media users (Awad & Krishnan, 2006). “Although there are great concerns about privacy issues, they continue to disclose personal information in social media,” Li, Yu, Xu and Xie noted (2018). On one hand, through the disclosure of user’s personal information, social media can realize personalized service and directional recommendation content, and help users to establish a broader social relationship. But at the same time, the privacy security of users has been repeatedly in crisis, such as the public sale of user accounts, crime of face recognition, and so on. China is a major country of mobile Internet and social media applications, with a huge social media user base. In the 44th Statistical Report on China’s Internet Development Status released by China Internet Network Information Center (CNNIC, 2019), the number of instant messaging users reached 825 million, accounting for 96.5% of China’s total Internet users. Due to the large base and high proportion of social media users, the issue of personal information disclosure has become one of the common problems in our society.

Through the above background analysis, it can be seen that reasonable access to and use of user’s personal information in social media can alleviate their concerns in privacy disclosure paradox and improve their use experience. Therefore, it is very important to understand the factors that affect users’ intention to disclose information. Although many scholars have studied the issue of online privacy disclosure, the research perspectives are different, and the interaction between various factors is not clear. This study attempts to investigate various types of social media software, combined with multiple theoretical perspectives to establish an analysis framework. Structural equation modeling (SEM) is used to verify the hypothetical model of influencing factors of personal information disclosure intention to determine the specific relationship and net impact between each variable, to provide a theoretical basis for providing the relevant countermeasures and suggestion.

2 Literature Review

As early as 1992, Amidon put forward a discussion on privacy concerns related to electronic information media. With the rise of social media in 2002, a large number of information behavior research on users of different platforms emerged. In 2011, two literature reviews on Internet privacy were published in the MIS Quarterly, which made the disclosure of user information privacy a hot topic in this field (Smith, Dinev, & Xu, 2011; Pavlou, 2011). The research on personal information disclosure can be sorted out from the research perspective, research context, and influencing factors. In the early research, it focused on explaining the mechanism of personal information disclosure behavior by establishing interdisciplinary theory, and gradually expanded to different application contexts and their influencing factors.

2.1 Research Perspective of Personal Information Disclosure Intention

2.1.1 Privacy Calculus Theory

When studying the intention of personal information disclosure behavior, scholars generally use the existing behavior theory as a framework for further analysis. Among them, Privacy Calculus Theory is the most commonly used. It believes that users decide whether to disclose their personal privacy information in the balance of risks and benefits (Korzaan & Boswell, 2008). The perceived benefits and perceived risks are key considerations of personal information disclosure. User’s intention will be motivated, and their disclosure behavior will be generated, only when the perceived benefits are greater than the perceived risks (Li, Hong & Zhu, 2016). Metzger (2004) found that the privacy risks in the Internet environment have a negative effect on the disclosure of private information by users. As for privacy benefits, Hann, Hui, Lee and Png (2007) conducted an empirical research. It proved that economic compensation promotes user’s disclosure of privacy information (Hann et al., 2007) and even the vast majority of users are willing to sacrifice personal privacy in exchange for personalized online services (Chellappa & Sin, 2005).
2.1.2 Theory of Planned Behavior

Theory of Planned Behavior (TPB) believes that personal behavior intention is the direct factor that affects actual behavior, and personal behavior intention is affected by behavior attitude, subjective norm, and perceived behavioral control (Ajzen, 1991). This means that when a user has a positive attitude, strong outside support, and strong perceived control over his behavior, he would likely disclose his personal information. Based on this theory, Guo, Duan, & Wang (2018) verified the privacy information disclosure behavior of mobile learning users from the above three aspects. In addition, this theory is often used in conjunction with Privacy Calculus Theory. For example, some researchers have combined the two theories to establish models in their research (Li & Chen, 2010; Shi, 2011).

2.1.3 Trust Theory

Trust Theory is mainly applied in the fields of psychology, sociology, organizational behavior, and so on. Trust means that one party ignores the ability to monitor the other party and accepts the intention to be at a disadvantage based on the expectation of the other party and that the other party will perform a specific and important behavior for itself (Mayer, Davis, & Schoorman, 1995). In the Internet environment, users are the subject of personal information. Neither their amount of information possession nor their ability to protect information is sufficient to “confront” information service providers. Thus, users can be seen as a relatively disadvantaged party. At this time, trust can resolve the risks and anxiety perceived by users, thereby increasing users’ personal information disclosure intention (Dinev & Hart, 2006). Trust can be disassembled into two concepts: Social Networking Site (SNS) and SNS users, both, can influence personal information disclosure intention through the mediating variable “personal information disclosure attitude” (Li & Wang, 2015).

2.1.4 Other Theoretical Perspectives

Scholars have also analyzed personal information disclosure intention from other theoretical perspectives: the Uses and Gratifications Theory believes that users’ disclosure of private information has a certain purpose and motivation (Hollenbaugh & Ferris, 2014); Personality Theory believes that personality factors such as narcissism and social anxiety have an important impact on the disclosure of personal identity information by young social network users (Liu, Cong, Ang, Rebecca & Lwin, 2013). From the perspective of Justice Theory, Culnan and Armstrong (1999) proved that perceived fairness has a positive impact on users’ privacy information disclosure intention.

2.2 Context of Personal Information Disclosure Intention

2.2.1 E-commerce Context

Since the rise of e-commerce and other emerging business models in 2006, the privacy information leakage of Internet consumers is facing great risks. Scholars began to pay attention to the privacy leakage channels, privacy concern measurement scale, privacy protection, and other issues in personalized e-commerce services (Yang, Wang & Wang, 2008). Zhao, Lu, and Guptaet (2012) found that business preferences, privacy policies, and legal awareness had indirect effects on users’ intention to disclose location related information. Ouyang and Yuan (2016) constructed a model of consumer privacy concerns influencing behavior intention in the e-commerce environment, in which information importance, website reputation, and trust are the main influencing factors. Zhu, Liu, Chen, and Lu (2014) studied the privacy threat avoidance behavior of mobile commerce smart phone terminal consumers, which is triggered by avoidance motivation and affected by social impact, perceived threat, and perceived avoidance ability.

2.2.2 Social Network Context

In addition, with increasing of social network users, personal information disclosure has become very convenient and more prevalent. Bergström (2015) found that trust had different influences on the online privacy concern of different social network user groups, and then affected the users’ online privacy settings and information disclosure degree. Hou and Ren (2013) built a model of individual privacy perceived risk and protection behavior through interview with users in social environment as QQ and Sina Weibo. Wang and Li (2016) discussed that the impact mechanism of fear appeals on the information security of social network users by analyzing questionnaire survey data, and built a behavior model based on the Fear Appeals Theory, revealing how social network users adopt privacy protection measures.
2.3 Influencing Factors of Personal Information Disclosure Intention

Personal information disclosure intention is affected by many factors, including the external social environment and user’s internal reasons. The external factors mainly show as social norms, peer effects, social culture environment, and policy environment. The research of Livingstone (2008) showed that peer effects had great influence on the privacy disclosure intention of teenagers. Zlatolas, Welzer, Heričko and Hölbl (2015) stated that user information disclosure intention was affected by privacy policy, privacy concerns, and privacy social norms.

In addition, the incentive measures and privacy policies of online platforms often affect the extent to which users disclose personal information. Research by Wu, Huang, Yen, and Popova (2012) showed that privacy policy has a negative impact on privacy concerns, thereby enhancing disclosure. However, the empirical study conducted by Stutzman, Capra, and Thompson (2011) on Facebook users showed that people who read more privacy policies tended to disclose less information. User’s internal factors include demographic characteristics, user habits, privacy view, and personality traits. For example, Zhang, Chen, and Lee (2013) found in research that female and male employees paid different attention to information privacy. In addition, although factors such as education level and nationality cannot directly establish a connection with user’s intention to disclose personal information, they can also be used as control variables to influence the research results (Valk, 2015).

In sum, relevant researches on personal information disclosure intention are still in the exploratory stage, empirical studies are mostly based on a single application (App) context, and the verification of influencing factors is relatively one-sided. Therefore, it is necessary to investigate the applicability of existing theories in different types of social media, and to explore the internal relationship between the factors influencing the personal information disclosure intention.

3 Research Model and Hypotheses

3.1 Theoretical Grounding

3.1.1 Theory of Planned Behavior

Proposed by Icek Ajzen (1988, 1991), TPB is the successor of the Theory of Reasoned Action (TRA) proposed by Fishbein and Ajzen (1975, 1980). Holding the view that a well-thought out plan is on account of human behavior, and all factors that may affect behavior indirectly affect the behavior performance through behavior intentions, TPB is generally used to understand how people change their behavior patterns. The behavior intention is affected by three related factors: attitude (internal factor, which means that individual’s attitude toward a particular behavior); subjective norm (which affects individuals to take a certain behavior); and perceived behavioral control (external factor, including secondhand information that comes from acquaintance, friends, and so on) (Ajzen & Driver, 1991).

TPB is widely used and has applications in various research fields of behavioral intention. It is not only the basis of the prototype willingness model (Wang & Zheng, 2016), but is also related to TRA and social cognition theory, which makes it a unique advantage in the comprehensive application of related theories.

3.1.2 Privacy Calculus Theory

As early as 1973, Laufer and Wolfe began to pay attention to the difficulties faced by personal information management – the balance between interpersonal interaction and privacy. They introduced Social Exchange Theory of economic field into user research, and refer to the cost–benefit evaluation as “Privacy Calculus”; subsequent research further deepened the theoretical conception, pointing out that calculus behavior is affected by both technology use experience and information management capabilities (Laufer, Proshansky, & Wolfe, 1973). Privacy Calculus Theory is a typical extension of behavior model, which discusses people’s attitude and behavior when facing privacy issues. The theory not only considers the positive factors that affect behavioral intentions (perceived benefit), but also takes account of some negative factors called perceived risk (Laufer & Wolfe, 1977). Based on Privacy Calculus Theory, users will weigh the benefits and risks when making personal information disclosure decisions. If the perceived benefit exceeds the cost, people will choose to disclose personal information.

3.2 Conceptual Model

The study selects two of the main factors that affect the behavior intention from TPB, and combine them with the basic frame of Privacy Calculus Theory. Finally, it chooses four measurement dimensions: perceived benefit,
Research on Influencing Factors of Personal Information Disclosure Intention of Social Media in China

perceived risk, privacy view, and subjective norm. These variables are assumed to have a direct relationship with personal information disclosure intention. Besides, based on previous studies, the relationship between privacy view and perceived risk is tested. Figure 1 shows the proposed hypotheses.

3.2.1 Perceived Benefit

Perceived benefit is defined as the benefit that is brought to users when they use a product, and users are aware of that (Lu, Tan, & Hui, 2004; Culnan & Bies, 2003). The main purpose that users disclose personal information on social media is to maintain or establish more interpersonal relationships, in order to gain a sense of belonging to the organization, and to obtain more services (Shi, 2011). The usefulness of social media includes four dimensions: relationship value, communicational value, informational value, and instrumental value. Some researchers have pointed out that, as an Internet application that helps people build and maintain social networks, the function of social media first performs in the establishment and maintenance of interpersonal relationships, then manifests as communicating with like-minded people, which means that social media have communicational value (Ellison, Steinfield, & Lampe, 2007). For informational value, it means that users obtain or share information through social media to understand popular opinions in the society. Instrumental value means that users can find solutions to problems and obtain others’ help through continuous online interaction and communication in order to complete certain tasks (Nie & Luo, 2013).

In previous studies, the significant impact of the perceived benefits to social network privacy disclosure has been proved (Li & Chen, 2010; Li et al., 2016). Sharma and Crossler (2014) found that perceived benefits will increase personal information disclosure for social commerce customers. It is also proved to be negatively related to the active protection of privacy (Anic et al., 2019). The opposite influence of perceived benefit and privacy concern on privacy disclosure has also been mentioned (Al-Jabri, Eid, & Abed, 2019). Therefore, we establish the first hypothesis:

H1: Perceived benefit will positively affect personal information disclosure intention.

3.2.2 Perceived Risk

Perceived risk refers to the loss that may be caused by the users’ personal information disclosure behavior due to social media’s illegal or improper use of information, and it is the user’s prediction of the worst outcome (Xu, Dinev, Smith, & Hart, 2008). The risks perceived by users mainly come from inappropriate access and inappropriate use of information (Dinev, Xu, Smith, & Hart, 2013). Previous studies have confirmed that there is a negative correlation between perceived privacy risk and disclose intention online (Dinev & Hart, 2006). That is to say, the more risks a user perceives, the less intention there is, to disclose personal information.

The negative influence of perceived risk to personal information disclosure has been proved by many studies. Sharma and Crossler (2014) found the impact of perceived risk in the social commerce environment. It also has a significant effect on social network (Xu, Michael, & Chen, 2013; Li, Hong & Zhu, 2016), and mobile learning users (Guo, Ma, & Xu, 2019). But the influence of perceived risk has also been proved insignificant. Lan (2017) and Guo et al. (2019) have found that perceived risk does not have an apparent impact on mobile social network users. As a result of the contrary conclusions, the effect of perceived risk needs testing. Therefore, the second hypothesis is established:

H2: Perceived risk will negatively affect personal information disclosure intention.

3.2.3 Privacy View

Privacy view reflects the importance of an individual’s treatment of personal privacy, and is related to the individual’s personal characteristics, cultural background, and experience (Li & Wang, 2015). Information sensitivity is defined as the level of privacy concerns that users
perceive for certain types of information in a specific context (Li, Hong & Zhu, 2016). Privacy view affects disclosure intention through information sensitivity. The stronger the privacy view is, the more sensitive users are to perceive risks, and users may show more cautiousness when they are asked to disclose their personal information. Therefore, the following hypotheses are made:

H3: Privacy view will negatively affect personal information disclosure intention.
H4: Privacy view will positively affect perceived risk.

3.2.4 Subjective Norm

Subjective norm refers to the social pressure that users perceive when deciding whether to perform a certain behavior (Ajzen, 1991). Social pressure mainly comes from the social and cultural environment, and people (family members, friends, etc.) who have a certain influence on the user. Subjective norm affects the risk users perceive and their intention to disclose their personal information.

Subjective norm has been widely used in researches about personal information disclosure intention, but the relationship between subjective norm and personal information disclosure intention depends on how others tend to behave. Some researchers hold the opinion that subjective norm shows a negative impact on personal information disclosure intention through perceived risk (Li, Hong & Zhu, 2016). Meanwhile, other researchers explain subjective norm as others’ willingness to show personal information and have discovered a positive relationship between subjective norm and personal information disclosure intention (Heirman, Walrave, & Ponnet, 2013; Varnali & Toker, 2015; Jiao, 2019). Therefore, Hypothesis 5 is established:

H5: Subjective norm will negatively affect personal information disclosure intention.

4 Methodology

4.1 Questionnaire Design

The following five constructs were measured using multiple-item scales: perceived benefit, perceived risk, subject norm, privacy view, and personal information disclosure intention. All items use five-point Likert-type scales (strongly disagree = 1 to strongly agree = 5). Prevalidated items were used following a pretest to ensure content validity.

Prior to data collection, a pilot test was conducted with 33 social media users who have different demographic backgrounds. They were requested to review the items to evaluate the constructs, semantics, suitability, and format of the questionnaire. According to their feedback, the questionnaire was revised. Moreover, we conducted a pilot study to ensure reliability and validity of the scales. The result of data analysis shows that Cronbach’s alpha of all constructs were above 0.7, which implies a strong internal consistency of theory. An online version of survey was designed and its hyperlink was posted on Wenjuanxing, a famous online survey platform, and distributed via WeChat. The subjects who had experience of using social media were invited to fill out the questionnaire. The items are listed in Appendix.

4.2 Data Collection

We posted a survey hyperlink to the social media users from May 30 to the end of June and September 10 to 28, 2020. Convenience sampling and snowball sampling were used as sampling method for respondents who filled out the survey successful received small monetary rewards. Finally, a total of 580 questionnaires were collected. Based on the selection of question 25 (“Please select “very disagree” for this question”) and the response time of the questionnaire (less than 60 s) and the same answer to all questions, we eliminated invalid data, and a total of 517 responses were left. The description of demographic characteristics of the respondents is shown in Table 1. More than half of the respondents are female, accounting for 69.83%. The ages of the respondents are concentrated in the young and middle-aged (19–59 years old), which is in line with the overall characteristics of the social media user group. The educational background is principally undergraduate. The description of demographic characteristics of the respondents is shown in Table 1.

5 Data Analysis and Results

5.1 Non-response Bias

This study addressed the issue of non-response bias follow the procedure suggested by Armstrong and Overton (1977) to conduct Chi-Square test for the early and late respondents. A total of 410 respondents who completed the survey during the early stage were considered earlier respondents and 107 respondents completed the survey
Research on Influencing Factors of Personal Information Disclosure Intention of Social Media in China

5.2 Common Method Bias

We conducted three tests to address the potential concern for common method bias (CMB). One is Harman’s one-factor test (Podsakoff & Organ, 1986). Evidence of common method bias exists when a general construct accounts for the majority of covariance among all constructs. A principal component factor analysis was performed and the results excluded the potential threat of common methods bias (Shiau & Luo, 2012). The combined five factors accounted for 76.205% of total variance; the first (largest) factor accounted for 17.510% (the variances explained ranged from 13.056% to 17.510%) and no general factor accounted for more than 50% of variance, indicating that common method bias may not be a serious problem in the data set. The second followed Liang’s (2007) method. A common method factor was included in the PLS model and it determines all indicators of the principal constructs. The result is shown in Table 2, indicating that the ratio of average substantively explained variance of the indicator (0.750) to average common method-based variance (0.011) was about 69.61. In addition, most method factor loadings were not significant. The third one used VIF as an indicator to test whether common method bias exist or not (Kock, 2015). In this study, all factor-level VIFs were range from 1.419 to 2.032, lower than 3.3, so the model can be considered free of common method bias. Based on the above tests, we concluded that common method bias was not a major concern in this study.

5.3 Measurement Model Test

We used partial least squares (PLS) with SmartPLS 3.0 to test the measurement model, which evaluates the measurement and structural model at the same time (Gefen et al., 2000). The result is presented in Table 3. All factor loadings of measurement scales were above 0.7 and the average variance extracted (AVE) values of every construct ranged from 0.661 to 0.856, all exceeding 0.50, showing a satisfactory convergent validity. All Cronbach’s alpha values and the composite reliabilities (CR) exceeded 0.70, implying a good reliability of the scale.

The convergent and discriminant validity were assessed by checking whether the AVE (average variance extracted) of each construct is larger than its correlation with the other constructs, and whether each item had a higher loading on its assigned construct than on the other constructs (Fornell & Larcker, 1981; Gefen, Straub, & Boudreau, 2005). The results indicate that the discriminate validity was achieved, as shown in Table 3.

5.4 Structural Model Test

A bootstrap analysis with 5,000 resampling method was applied to determine the significance of the structural model paths. The path coefficient and significance of each hypothesis were examined and the explained variance ($R^2$) of each dependent construct was calculated. Figure 2 shows the results of the tests. Overall, the model explained 18.4% of the variance for personal information disclosure intention and 33.0% of the variances was explained in privacy view and perceived risk. Specifically, perceived benefit was significantly related to personal information disclosure intention ($b = 0.151, t = 3.443$), which supported H1. Privacy view was significantly related to perceived risk ($b = 0.575, t = 12.603$) and subjective norm was significantly

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic Characteristics of the Respondents ($N = 517$)</td>
</tr>
<tr>
<td><strong>Category</strong></td>
</tr>
<tr>
<td>Gender</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Education</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Occupation</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Table 2
Common Method Bias Analysis

<table>
<thead>
<tr>
<th>Construct</th>
<th>Indicator</th>
<th>Substantive factor loading($R_1^2$)</th>
<th>$R_1^2$</th>
<th>Method factor loading($R_2^2$)</th>
<th>$R_2^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived Benefit(PB)</td>
<td>PB1</td>
<td>0.81***</td>
<td>0.6561</td>
<td>-0.058</td>
<td>0.003364</td>
</tr>
<tr>
<td></td>
<td>PB2</td>
<td>0.882***</td>
<td>0.777924</td>
<td>-0.061</td>
<td>0.003721*</td>
</tr>
<tr>
<td></td>
<td>PB3</td>
<td>0.805***</td>
<td>0.648025</td>
<td>0.125</td>
<td>0.015625**</td>
</tr>
<tr>
<td>Perceived Risk(PR)</td>
<td>PR1</td>
<td>0.925***</td>
<td>0.855625</td>
<td>-0.03</td>
<td>0.0009</td>
</tr>
<tr>
<td></td>
<td>PR2</td>
<td>0.926***</td>
<td>0.857476</td>
<td>0.03</td>
<td>0.0009</td>
</tr>
<tr>
<td>Subjective Norm(SN)</td>
<td>SN1</td>
<td>0.906***</td>
<td>0.820836</td>
<td>0.062</td>
<td>0.003844</td>
</tr>
<tr>
<td></td>
<td>SN2</td>
<td>0.907***</td>
<td>0.822649</td>
<td>-0.062</td>
<td>0.003844</td>
</tr>
<tr>
<td>Privacy View(PV)</td>
<td>PV1</td>
<td>0.792***</td>
<td>0.627264</td>
<td>0.294</td>
<td>0.086436</td>
</tr>
<tr>
<td></td>
<td>PV2</td>
<td>0.866***</td>
<td>0.749956</td>
<td>-0.075</td>
<td>0.005625</td>
</tr>
<tr>
<td></td>
<td>PV3</td>
<td>0.786***</td>
<td>0.617796</td>
<td>0**</td>
<td>0**</td>
</tr>
<tr>
<td>Personal Information Disclosure Intention(PIDI)</td>
<td>PIDI1</td>
<td>0.887***</td>
<td>0.786769</td>
<td>0.05</td>
<td>0.0025*</td>
</tr>
<tr>
<td></td>
<td>PIDI2</td>
<td>0.885***</td>
<td>0.783225</td>
<td>-0.051</td>
<td>0.002601*</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td>0.86475</td>
<td>0.750304</td>
<td>-0.00025</td>
<td>0.01078</td>
</tr>
</tbody>
</table>

Note: *p<0.05; **p<0.01; ***p<0.001

Table 3
Results of Confirmatory Factor Analysis

<table>
<thead>
<tr>
<th>Factor</th>
<th>Item</th>
<th>Mean</th>
<th>SD</th>
<th>Factor Loading</th>
<th>Cronbach’s $\alpha$</th>
<th>CR</th>
<th>AVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived benefit</td>
<td>PB1</td>
<td>3.443</td>
<td>1.117</td>
<td>0.828</td>
<td>0.779</td>
<td>0.871</td>
<td>0.693</td>
</tr>
<tr>
<td></td>
<td>PB2</td>
<td>3.364</td>
<td>1.080</td>
<td>0.891</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PB3</td>
<td>3.807</td>
<td>1.003</td>
<td>0.775</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived risk</td>
<td>PR1</td>
<td>4.325</td>
<td>0.873</td>
<td>0.923</td>
<td>0.832</td>
<td>0.923</td>
<td>0.856</td>
</tr>
<tr>
<td></td>
<td>PR2</td>
<td>4.468</td>
<td>0.773</td>
<td>0.927</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subjective norm</td>
<td>SN1</td>
<td>3.221</td>
<td>1.059</td>
<td>0.904</td>
<td>0.783</td>
<td>0.902</td>
<td>0.822</td>
</tr>
<tr>
<td></td>
<td>SN2</td>
<td>2.642</td>
<td>1.150</td>
<td>0.909</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Privacy view</td>
<td>PV1</td>
<td>4.321</td>
<td>0.863</td>
<td>0.833</td>
<td>0.747</td>
<td>0.853</td>
<td>0.661</td>
</tr>
<tr>
<td></td>
<td>PV2</td>
<td>4.360</td>
<td>0.920</td>
<td>0.853</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PV3</td>
<td>4.667</td>
<td>0.686</td>
<td>0.749</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personal information disclosure intention</td>
<td>PIDI1</td>
<td>2.720</td>
<td>1.206</td>
<td>0.894</td>
<td>0.727</td>
<td>0.880</td>
<td>0.785</td>
</tr>
<tr>
<td></td>
<td>PIDI2</td>
<td>2.190</td>
<td>1.032</td>
<td>0.878</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
associated with personal information disclosure intention (b = 0.371, t = 8.857), which supported H4 and H5 respectively. H2 and H3 were not supported because they are not significantly related to personal information disclosure intention.

### 6 Discussion and Implications

#### 6.1 Discussion of the Preliminary Results

The results verified by previous studies show that perceived risk has a significant negative impact on personal information disclosure intention (Li & Wang, 2015), but the results of this study show that it has no effect. It is mentioned that if users do not provide certain personal information, they cannot use social media functions in the privacy policy to users made by social media. As a result, users have to accept the provision of personal information even when they perceive risks. It may be one of the possible reasons why the perceived risk does not have a significant impact. It is also possible that users feel that they are lurkers in social media and will not gain too much attention, and whether or not to disclose personal information is not so important. Users’ trust in social media brands may also affect users’ perceived risks.

With this study, we make two aspects of theoretical contributions to the theory and literature. First, our research uses social media with rich connotations and classifications, which means that our research objects extend to all social media rather than a specific social media, so it can provide some ideas for other specific research. Second, for the Privacy Calculus Theory, our preliminary findings imply that the social media’s restrictions on the user’s acceptance to use will affect the loss calculation in the user’s privacy calculation. The Privacy Calculus Theory needs to consider more adjustment factors.

Drawing on preliminary findings from this study, we also provide some guidelines for social media. First, social media should improve the privacy settings and enhance their privacy protection technology, which should be humanized and easy to use. Second, the more benefits users perceive, the stronger their intention to disclose personal information will be. Therefore, social media should further enhance their use experience and provide users with valuable social services. They can design the corresponding functions to meet users’ relationship and instrumental benefit needs, such as self-display, perceived usefulness of life or work, and so on. At the same time, while meeting the basic social needs of users, social media can also provide other personalized functions to ensure users’ diverse needs. Third, the proliferation of opportunism has caused various types of information of users’ to be stolen or sold. The personal information disclosed by users on social media is varied, which creates various risks. Social media should use content or function as a profitable method and reduce the investment in worthless advertising or unnecessary recommendations. And social media should pay attention to the protection of users’ sensitive information and balance the relationship between providing users with personalized services and limiting access to users’ personal information.

#### 6.2 Limitations and Future Research

There are also some limitations in this study. First, the main sample of this study was middle-aged and young people, while there is still space to optimize research methods and
research objects. Second, this research is only conducted on Chinese social media users, and is insufficient in terms of regional cultural diversity. Therefore, future research will expand the age range and geographical scope of the research objects. Otherwise, some scholars (Awad & Krishnan, 2006) believe that Privacy Calculus Theory is the most useful framework for analyzing user privacy issues; however, our preliminary findings imply that users cannot use some services without accepting personal information, which makes the calculation and balance of users' interests in privacy less obvious. The impact of this factor on the calculation of privacy interests will also be one of the considerations of future research.

Acknowledgments: This work was supported by grant from the Social Science Foundation of China (17BTQ079).

References


Research on Influencing Factors of Personal Information Disclosure Intention of Social Media in China


### Appendix

**Questionnaire Items**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Item</th>
<th>References</th>
<th>Item design</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Part one</strong></td>
<td><strong>Social Media Usage</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1. Frequently used social media names</td>
<td></td>
<td>Multiple choice</td>
</tr>
<tr>
<td></td>
<td>3. The average amount of time you spend on social media in a week.</td>
<td>Qiu (2019)</td>
<td>Single choice</td>
</tr>
<tr>
<td><strong>Part two</strong></td>
<td><strong>Perceived Risk</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. Disclosure of personal information to social media may lead to</td>
<td>Dinev et al. (2013)</td>
<td>Five-point Likert-type</td>
</tr>
<tr>
<td></td>
<td>disclosure of personal information.</td>
<td></td>
<td>response scale</td>
</tr>
<tr>
<td></td>
<td>5. In general, there are risks in disclosing personal information</td>
<td></td>
<td>Five-point Likert-type</td>
</tr>
<tr>
<td></td>
<td>to social media.</td>
<td></td>
<td>response scale</td>
</tr>
<tr>
<td><strong>Perceived Benefit</strong></td>
<td>6. Social media can help me build identity and a sense of belonging</td>
<td>Nie et al. (2013)</td>
<td>Five-point Likert-type</td>
</tr>
<tr>
<td></td>
<td>in a virtual community.</td>
<td></td>
<td>response scale</td>
</tr>
<tr>
<td></td>
<td>7. I use social media to help shape my opinion.</td>
<td></td>
<td>Five-point Likert-type</td>
</tr>
<tr>
<td></td>
<td>8. I use social media to communicate ideas with others.</td>
<td></td>
<td>response scale</td>
</tr>
<tr>
<td></td>
<td>10. I care about how others handle my personal information.</td>
<td></td>
<td>response scale</td>
</tr>
<tr>
<td></td>
<td>11. It is important for me to keep my privacy from being</td>
<td></td>
<td>Five-point Likert-type</td>
</tr>
<tr>
<td></td>
<td>violated by others.</td>
<td></td>
<td>response scale</td>
</tr>
<tr>
<td><strong>Subjective Norm</strong></td>
<td>12. Most of the people around me share real personal information on</td>
<td>Xu, F., Michael, K., &amp; Chen,</td>
<td>Five-point Likert-type</td>
</tr>
<tr>
<td></td>
<td>social media.</td>
<td>X. (2013)</td>
<td>response scale</td>
</tr>
<tr>
<td></td>
<td>13. Most of the people around me think I should pay attention</td>
<td>Xu, F., Michael, K., &amp; Chen,</td>
<td>Five-point Likert-type</td>
</tr>
<tr>
<td></td>
<td>to the protection of personal information.</td>
<td>X. (2013) &amp; Venkatesh &amp;</td>
<td>response scale</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Davis, 2000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>15. I will disclose more personal information on social media in</td>
<td></td>
<td>response scale</td>
</tr>
<tr>
<td></td>
<td>the future.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16. (lie detector)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Part three</strong></td>
<td><strong>Demographic Information</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>17. Your gender</td>
<td></td>
<td>Single choice</td>
</tr>
<tr>
<td></td>
<td>18. Your age</td>
<td></td>
<td>Single choice</td>
</tr>
<tr>
<td></td>
<td>19. Your education background</td>
<td></td>
<td>Single choice</td>
</tr>
</tbody>
</table>
An Empirical Study on the Cueing Effect of the Emotional Post Title in a Virtual Community

Guihua Li, Siyan Lin*, Weiping Yu, Sang Guo

Abstract: In a virtual community, the behavior of strengthening the emotion of a post title to draw attention of users is not uncommon, which can affect the overall performance of the information environment. This study focuses on exploring the influence of the emotional information of a post title on the users’ community perceived value in a virtual community. Based on the cue utilization theory, we propose a framework with several hypotheses. Data are collected using the experimental method from the college student sample in our study, and numerous tests are performed to analyze the data and verify the hypotheses. At the end of the study, it is found that the emotional information of the post title reduces the user community perceived financial value and it improves the user community perceived recreational value. The analysis of the mediating role reveals that emotional involvement facilitates the relationship between emotional information of post titles and user community perceived recreational value. This study adds a new dimension by discussing the user community value perception on post title expression and it reveals the conflict of interest between the manager of the virtual community and the producers of the post. Our findings may also provide guidelines and references for virtual community managers. Specifically, they should view the behavior of making post titles more emotional critically, and choose specific information management strategies based on the different value pursuit of community users.

Keywords: virtual community, emotional information, cueing effect, cue utilization theory

1 Introduction

The social networking environment has occupied an important space of people in seeking information, in which “making information noticed by users” has become the bottleneck of information delivery. Some post producers consider that strengthening the emotion of a post title is an important method of getting attention and promoting post hits. However, as these kinds of behaviors become rampant, virtual community managers have to consider whether the behaviors of making post titles more emotional should be managed and strategized to manage them.

Based on the economic studies, cues can reduce the marginal cost of information seeking or increase its marginal revenue, thus influencing users’ intention of seeking information and value perception (Stigler, 1961; Nelson, 1974). It is worth noting that virtual communities can be regarded as tridimensional information networks in which post titles are usually read from a list by users at first, which may contain a dozen or dozens of post titles. Normally, users generally browse the post titles of the first page or first few pages and then read selectively or leave the virtual community. Hence, the skimming and scanning part is a process through which users evaluate specific posts based on whether the titles of the posts match their demands and further, it’s also a process of forming an overall impression of the community and evaluating the community. Thereby, we propose that strong emotional post title lists, as users’ first impression of the virtual community may influence the users’ community perceived

*Corresponding author: Siyan Lin, School of Public Administration, Sichuan University, Chengdu, Sichuan, China. Email:342615923@qq.com

Guihua Li, School of Public Administration, Sichuan University, Chengdu, Sichuan, China
Weiping Yu, Business School, Sichuan University, Chengdu, Sichuan, China
Sang Guo, School Library, Wenzhou Medical University, Wenzhou, Zhejiang, China

Open Access. © 2021 Guihua Li et al., published by Sciendo. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 3.0 License.
value. If so, what is the specific effect of post titles’ emotional information on the user community perceived value? Thus, this study is necessary and can provide managerial insights for virtual community managers.

2 Theoretical Background

2.1 Virtual Community

A virtual community is a network community that exists in the internet environment, and it uses electronic tools as a medium where people conduct online communication and interaction based on certain interests or certain demands (Armstrong & Hagel, 1996). Virtual community is an open and free information field, which meet the users’ needs of purposeful information search and scanning information search, and so it has become an important part of users’ information life (Ellis, Oldridge, & Vasconcelos, 2004).

One of the key areas of virtual community research is virtual community performance research. Existing researches on virtual community performance are mainly carried out based on certain type of consumers’ community performance evaluation, such as posters or browsers. A series of studies have found that the perceived community value, which is a performing predictor of community performance, has a significant impact on consumer behavior. It is noted that the perceived value is the main predictive variable of a consumer’s purchasing behaviors for the service environment. The perceived value is defined as a consumer’s overall assessment on the utility of a product based on the perceptions of what benefit is received and what is given in the process of obtaining products or services (Zeithaml, 1988). Surachartkumtonkun and Patterson (2007) pointed out that the perceived value that users obtain from joining online brand communities can be divided into three dimensions, namely, emotional, social, and functional values, based on the theory of perceived value. Furthermore, Sicilia & Palazón (2008) analyzed the online strategies developed by Coca-Cola in Spain and concluded that the advantage of the Spanish strategy relies on the creation of a virtual community that provides consumers with functional, social, and experiential values. The empirical analysis conducted by Jin (2007) which considered the members of China’s network virtual brand community as the object showed that the value perception of virtual community chiefly consists of four dimensions: information value, financial value, entertainment value, and social value. However, current studies on value perception of the virtual community mainly consider an independent variable. Only fewer studies have asked: what exactly influences community value perception?

Although no research has explored the relationship between community content and community value perception, it is a fact that a good information content is the main factor that affects user participation in a virtual community (Hagel & Armstrong, 1997; Chang, Kannan, & Whinston, 1999), and if community managers take active measures to optimize the allocation of information resources, it would help users get high-quality information and promote the rapid development of the community (Porter and Donthu, 2008). However, the current research on information characteristics of virtual community only includes the content characteristics of information, and the formal characteristics of information are not at all included.

The balance of interests among community participants should be considered while managing the community content and making related policies. The research made by Tony, Constantino and Dobele (2018) shows that there are three key levers which keep the balance of demand between enterprises and community members: appealing, consulting, and affiliating actions. But in fact, generally community members and community managers are considered as the participants of virtual community and there may be conflicts of interests between them. Therefore, the study on these conflicts of interests is of more general significance, especially when members of the community have specific identities such as posters and browsers, and members of different identities have different demands. However, there is no study that explored the conflicts of interest between community posters and community managers specifically.

2.2 Emotion and Cognition

Early psychology considered emotion as a reflection of cognition. However, the current psychological research on emotion–cognition relationship has supported the idea increasingly that emotion has independent effects. For example, there is a finding that the cause of emotion or emotional arousal has nothing to do with cognitive evaluation (Wilke, 2001), and the effect of emotion on cognition is two-fold. That is, it affects both what people think (information effect) and how they think (processing effect) (Kahneman, 2003).

On the other hand, researches show that emotion has a tendency to approach or avoid a scenario, thus affecting the changing direction and aspect of human intelligence,
and this phenomenon, in turn, can inevitably affect the acquisition of knowledge (Piaget, 1981). Furthermore, Izard (1977) believes that emotion provides clues to cognitive activities and drive cognitive behaviors. More specifically, emotion changes the function of cognition not only in the aspect of processing speed and accuracy, but also in the categories and hierarchies, or emotion causes a quality change of blocking or interference in the course of information processing. Experiments show that, compared with non-emotional stimulation, stimulation with emotional implications is more likely to attract attention and occupy attention resources, leading to attentional bias (Lang, 1995).

However, researchers have found out that emotion can be one of the crucial factors that accelerate decision-making. Epstein’s theory of double processing (Epstein, 1994) states clearly that, the analysis process is undoubtedly important in a complex, uncertain, and dangerous situation for taking decisions, and it is a faster, simpler, and more effective way to make judgment and decisions based on feelings and emotion. Furthermore, Clore, Schwarz, and Conway (1994) put forward the concept of emotion–information equivalence, believing that emotion can be used as a kind of information clue directly influencing one’s judgment.

Linguistic signs are considered an important stimulation of emotion as the emotional information contained in a language may help people understand each other better and express their own thoughts and feelings, and the emotion can often be processed first (Ohman, Flykt, & Esteves, 2001). Hermans, De Houwer, and Eelen (2001) found that the emotional meaning of emotional words is captured earlier than their conceptual meaning. Furthermore, emotional information will have an impact on other information receptions, which is supported by Nielsen, Shapiro, and Mason (2010) on the affective titles of semantic advertising, showing that the affective semantic title can cause readers’ directional and preferential processing. That is, when readers pay attention to other different objects, the affective semantic title is more likely to capture their attention, and it may enable readers to have a more significant awareness of the brand among the corresponding advertisements.

More and more scholars have begun to explore the value of emotional information in the context of social media. For instance, Krishen, Berezan, and Raab (2019) has found that, in a social network, a sense of belonging and emotional connection are key hedonic elements and are more relevant for promotion-focused individuals. But, for prevention-focused individuals, the utilitarian and functional aspects of interactivity(influence) and innovativeness(risk) are more important factors in influencing satisfaction with a social network. A research also shows that virtual communities, like real ones, are connected not only because of utilitarian information exchange, but also because they serve the social need of having a friend and getting social support (Leng & Lim, 2013). However, these studies chiefly emphasized the emotional mechanism caused by interpersonal emotional interaction, and did not conduct any research with information affective expression characteristics as the antecedent object. At the same time, the existing researches on the emotional value of virtual community are generally based on rational behavior model and its derived technical acceptance model, and they have been carried out according to the logic of cognition-emotion-behavior, without taking into account the independent role of emotional information.

### 2.3 Cue Utilization Theory

Cox (1967) suggested that cues serve as important information for consumers in the quality perception process, and quality cues can be described and categorized in two dimensions, i.e., cue predictive value (PV) and cue confidence value (CV). PV is defined as the degree at which customers associate the cue with the quality of products or services. That is, PV shows the reliability of the cue and the accuracy of consumers of using the cue to judge the quality of products or services. CV is defined as the extent to which consumers are confident in their abilities to judge and use the cue in the quality perception process. PV and CV are both necessary, and when the cue has both high PV and CV, it will be more likely to be and effectively used to judge the quality of products, which has a stronger effect on the final judgment.

Based on the above points, Olson and Jacoby (1972) further added the third variable-cue internality and externality (I/e) to cox’s PV and CV dimensions. The internal cues refer to the internal information of the product. If the physical characteristics of the product itself are not changed, they cannot be changed or controlled, such as the taste and shape of the product. External cues are attributes that are related to the product but not part of the physical product, such as brand name, packaging, price, etc.. They also pointed out that the quality perception process is considered to have two stages. During the first stage, consumers acquire and carefully choose the quality cues from a series of product-related attributes, in which the product scenario and consumer scenario will have an impact on these cognitive activities,
thus influencing the results of cue utilization. Meanwhile, intrinsic and extrinsic information also play a role in this process as well as the fact that cues with high PV and CV are more easily noticed and put into use. At the second stage, consumers integrate the cues derived from the first stage to form their impressions of the object and establish their judgments based on various attributes of the object, including the differentiation and sorting of these attributes. Olson and Jacoby regarded this stage as the evaluation stage of cues, and this stage completes the task of cue utilization.

The cue utilization theory has been widely used in the field of service marketing. Some researchers pointed out that consumers pay more attention to design-oriented, social, and ambient environmental cues when evaluating products because they believe that these cues (e.g., music, smell, and price) provide reliable information about product-related attributes such as quality, price, and shopping experience (Baumgarten & Hensel, 1987; Ward, Bittner, & Barnes, 1992; Greenland & McGoldrick, 1994; Wakefield & Baker, 1998). However, the application of the cue utilization theory chiefly focused on the physical environment, and only few studies applied the cue utilization theory to the network environment.

3 Hypotheses Development

In the contemporary e-commerce environment, consumers rely more on all kinds of information cues for service prediction. Grounded on the two-stage cue utilization process mentioned above, we postulate that the cueing effect of post titles' emotional cues on user community perception also involves two stages. According to postulation, strong emotional cues compete with other cues in the first stage. According to the cue utilization theory, post titles with strong emotional cues are more likely to draw users' attention, thus more easily acquired and processed by users, and in the meantime their emotional involvement increased simultaneously. Therefore, other post titles with relatively weak emotional cues are paid less attention to and the information are processed to a lesser extent as most of the users' attention is devoted to post titles with strong emotional cues. In addition, in the second stage, the emotional information influences users' judgment on the attributes of the virtual community. That is, the attributes of the community relevant to strong emotional cues are more likely to be highly rated; however, other attributes of the community that need more support from other information resources are probably given a lower rating.

We posted a series of hypotheses to test the above inferences, that is, the cueing effect of post titles' emotional cues on the user community perceived value.

3.1 Effects of Post Title Lists' Emotional Cues on User Community Perceived Value

Based on the comprehensive analysis on existing researches, it is observed that studies about the dimensions of the community perceived value have not reached a consensus, but they are still very close. We adopted Jin's point of view and divided perceived community value into four categories. Meanwhile, we think that the name of perceived information value is too vague, and that functional value is a more appropriate name according to its specific definition and test items, so we change the name of perceived information value to perceived functional value. Therefore, we subdivide perceived community value into functional value (information value included), social value (or psychological value), recreational value, and financial value. Consequently, we employ these four dimensions to measure the perceived value of virtual communities in this study.

According to the above discussion, we postulate that the emotional information on post titles in virtual communities influences the users' community perceived value. That is, the strength of emotion differs depending on the emotional cues on post titles, leading to a significant difference in users' perceived value of virtual communities. However, owing to the fact that different dimensions of perceived value, for example, functional value, and social value are different in origin, the strength of cues may have varied influences on the different dimensions of the perceived value.

The perceived recreational value of virtual communities is closely related to users' motivations in satisfying curiosity and seeking thrill and adventure. In fact, these motivations are described as emotional motivations, so an emotional stimulus is more likely to satisfy these motivations easily. Moreover, it has been shown that an emotional stimulus (including words, pictures, cartoons, and stick figures) is more likely to capture and take up users' attention, which causes attentional bias when compared to a non-emotional stimulus, (Lang, 1995). Accordingly, we presume that, emotional post titles will get more information processing from users as emotional post titles take up more attentional resources than non-emotional ones. Furthermore, it is
also postulated that emotional post titles containing more emotional stimuli meet users’ recreational demand better, thus affecting the judgment of the virtual community, i.e., improving the user’s perceived recreational value. Hence, the following hypothesis is proposed:

\( H_{1a} \): Compared to the post title lists with weak emotional cues, post title lists with strong emotional cues lead to a higher community perceived recreational value.

It was pointed out that the social value correlated with the relatedness needs (Jin, 2007). That is, virtual communities assisted users to build relationships online, exchange information, and express emotions, in which strong emotional post titles can play a promoting role. Therefore, we propose the following hypothesis:

\( H_{1b} \): Compared to post title lists with weak emotional cues, post title lists with strong emotional cues lead to a higher community perceived social value.

Emotional information is evocative. Lang (1995) suggested that there is a strong correlation between the emotional intensity of words and the arousal of individuals, while Eysenck (1982) has found out that arousal leads to narrowing of attention. More studies have found that the arousal effect of emotional stimulation leads to significant attention attraction and attention bias. Further, Gorn, Pham, and Sin (2001) demonstrated that affective cues are registered more rapidly than cognitive assessments; the relative accessibility of affective cues also increases with their volume and evaluative clarity or intensity. The research conducted by Yiend (2010) and Lang (1995) shows that when compared to the stimulus without emotional color (or emotional meaning), the stimulus with emotional color (or emotional meaning) is more capable of attracting attention, occupying attention resources, and causing attention bias. Moreover, the research made by Nielsen et al. (2010) also proves that emotional advertisement titles are more likely to attract readers’ attention easily and lead to the directional diversion of their attention.

Thus, we can postulate that, as information processing of post titles with strong emotional cues gains more attention, the processing resource of other post titles will be inadequate if the post titles’ content is ignored.

The community perceived functional value refers to the functional benefits that users get from joining virtual communities. Particularly, these functional benefits can be information, efficiency, or convenience (Yang & Liu, 2010). Furthermore, the community perceived financial value is related to transaction demand, which is the financial benefits that are obtained by the users by joining virtual communities and taking part in activities, mainly consisting of discounts users can get after becoming members (Jin, 2007). The judgment of the perceived functional value and perceived financial value is usually based on material information. However, when post titles with strong emotional cues occupy more attention, it results in less elaborate processing of other post titles, thus supporting that these two perceived values may be inadequate. On the basis of these arguments, we propose the following hypotheses:

\( H_{2a} \): Compared to post title lists with weak emotional cues, post title lists with strong emotional cues lead to a lower community perceived functional value.

\( H_{2b} \): Compared to post title lists with weak emotional cues, post title lists with strong emotional cues lead to a lower community perceived financial value.

### 3.2 Mediating Effect of Emotional Involvement

Involvement is defined as an individual’s internal state of arousal with intensity, direction, and persistence properties (Andrew, Durvasula, & Akhter, 1990). Researchers mostly agree that involvement has two dimensions, namely, cognitive involvement and emotional involvement (Zaichkowsky, 1994). Cognitive involvement specifically stresses an individual’s information processing activities and achievements on idealization states (McGuire, 1974), usually measured in terms of relevant, important, needed, and valuable, whereas emotional involvement stresses a person’s feelings and achievements of certain emotional states (McGuire, 1974), which are measured as in terms of interesting, appealing, fascinating, exciting, and involving. That is, cognitive involvement corresponds to the cognitive effects of the information value itself, whereas emotional involvement corresponds to the affective effects of the information’s emotional cues.

In addition, studies on physical stores and online stores found out that consumer emotion mediates or facilitates the relationship between the products’ external characteristics, shopping environment, and consumer behaviors (Eroglu, Machleit, & Davis, 2003; Menon & Kahn, 2002; Sherman, Mathur, & Smith, 1997; Spies, Hesse, & Loesch, 1997). Therefore, we postulate that emotional post titles lead to a higher level of users’ emotional involvement, further resulting in attention focusing and deep information processing, and finally influence the user community perceived value. On the basis of the above analysis, we infer the following:

\( H_{3} \): Emotional involvement mediates the relationship between the strength of the emotional cues of post titles and community perceived value.
According to the cue utilization theory, cues have both PV and CV. Moreover, when intrinsic cues have low credibility and PV, extrinsic cues can be especially valuable (Cox, 1967). As an example, when consumers have a problem explaining how well a product will perform, how safe the product is, and how socially acceptable it may be, they tend to enhance the confidence of predicting product performance by product price, warranty information, or other extrinsic cues (Bearden & Shimp, 1982).

Value judgment of service usually lacks intrinsic and objective criteria, especially as the virtual community has a free and open communication environment and particularly the answers to questions whether the users of the virtual community are true consumers, whether the information of the post is trustworthy, and whether the service management of the community are reliable due to lack direct evidence. Therefore, the post titles’ strength of emotional cues is especially valuable when we lack other cues. That is, post titles with strong emotional cues enhance users’ confidence in predicting the community value.

Therefore, consistent with the above analysis, the following hypothesis is proposed:

$H_3$: Compared to post title lists with weak emotional cues, post title lists with strong emotional cues enhance users’ confidence in predicting the community perceived value.

On the basis of what has been analyzed and hypothesized, we present our research model in Figure 1.

### 4 Research Design and Methodology

#### 4.1 Overview of the Study

The experiment was conducted using a between-subjects design to analyze the emotional strength of post titles. More specifically, this experiment considered gender as a random variable; knowledge of post titles, affective valence, cognitive involvement, and age as control variables; and community perceived value as the dependent variable.

#### 4.2 Subjects

Subjects were 118 undergraduate business students at Sichuan University in China. There were 52 males and 64 females, and the average age was 20 years old. They were invited to participate in this experiment.

#### 4.3 Materials

The experiment employed a virtual brand community as an experimental environment. The virtual brand community comes from a cluster of consumers who are interested in a particular brand on the Internet. In the experiment, post title lists served as the main stimulus, and another relevant experimental environment is designed as follows:

##### 4.3.1 Product Category Selection

In our experiment, the selection of the product category is chiefly based on the degree of involvement because only when the product has a high level of consumer involvement the consumers will have a tendency to search information related to that product. Furthermore, choosing a product category with a higher level of involvement is beneficial for subjects to be more involved in the experiment, thus increasing the reliability of the experiment.

As the subjects of this study were undergraduate students, we separately interviewed 20 undergraduate students (10 females and 10 males) in order to select a product category with a high level of involvement. Specifically, they were asked to name the most important product category that comes to mind first in purchase decision making. Two of the most frequently mentioned categories were laptops and mobile phones. Therefore, we chose laptops as a product category of the experiment.

##### 4.3.2 Product Brand Selection

According to ZOL.com (http://nb.zol.com.cn), a professional IT website, the forum of ThinkPad had the largest number of members among all laptop forums, which indicated that a variety of consumers were familiar
with and had a high level of involvement with this brand. Thus, we chose ThinkPad as a product brand in our experiment.

### 4.3.3 Web Design of the Virtual Brand Community

We employed Douban.com (https://www.douban.com/group/thinkpadclub/), which is a famous social networking platform in China, as our template for the web design of a virtual brand community in the experiment. According to the layout of the group webpage on Douban.com, there were 25 post titles per screen, that is, the post title list always contains 25 post titles.

### 4.3.4 Production of Target Post Titles in the Post Title List

The webpage in our study contained 12 target post titles and 13 non-target post titles. The target post titles in two experimental groups are presented in Table 1, and the extraction method of 12 target posts of two experimental groups was as follows:

First, we selected 30 post titles with strong emotional cues in ThinkPad’s community on Douban.com and asked 10 subjects to rate their emotional strength. Based on the rating, we chose four positive posts with strong emotional cues, four negative posts with strong emotional cues, and four neutral posts with strong emotional cues. Therefore, the post titles of these 12 posts chosen served as the target post titles for the strong emotional cue group. Then, we processed the content of these 12 target posts after removing the emotional expression component of these posts. Further, we chose 10 subjects to rate the emotional strength of these posts to test whether these posts were weak emotional after processed and found that their emotional strength has decreased dramatically. Thus, the post titles of these posts processed served as part of the weak emotional cue group.

### 4.3.5 Production of Non-target Post Titles in the Post Title List

The other 13 post titles in the post title list are non-target post titles. First, we used convenient sampling to select the page of the post title list in ThinkPad’s community on Douban.com. Then, we adopted the method of quota sampling to select 13 post titles by employing the subject of post titles as standard.

### 4.3.6 Sorting of the Post Title

We adopted the method of perforated sorting to get the distribution of strong emotional post titles more natural. That is, 12 target post titles were evenly distributed in the positions of odd numbers. Further, we did dispersive processing for the target post titles with different affective valence.

---

**Table 1**

*Comparison of Target Post Titles in Two Experimental Groups*

<table>
<thead>
<tr>
<th>Post Titles with Strong Emotional Cues</th>
<th>Post Titles with Weak Emotional Cues</th>
</tr>
</thead>
<tbody>
<tr>
<td>I always think that the ThinkPad’s sound is low, does anyone agree with me?</td>
<td>The ThinkPad’s sound is low.</td>
</tr>
<tr>
<td>The x200 fan!!! I wanna cry!!!</td>
<td>My x200 fan is broken.</td>
</tr>
<tr>
<td>Save me!!! My 9000rmb E40 laptop is driving nuts!!!!!</td>
<td>9000rmb E40 laptop</td>
</tr>
<tr>
<td>My ThinkPad’s battery is gone...What to do, what to do?</td>
<td>My ThinkPad’s battery is gone</td>
</tr>
<tr>
<td>There is a problem with my W510 microphone!!!!</td>
<td>There is a problem with my W510 microphone</td>
</tr>
<tr>
<td>Oh my goodness! What kind of keyboard does the ThinkPad have!?</td>
<td>The ThinkPad has bad keyboard</td>
</tr>
<tr>
<td>Someone help me, my computer won’t start:'(‘</td>
<td>Help, my computer won’t start</td>
</tr>
<tr>
<td>Can anyone save my T60 laptop!?</td>
<td>My T60 laptop is broken</td>
</tr>
<tr>
<td>The T34P’s ultra-sharp screen is pretty awesome!!</td>
<td>The T34P’s ultra-sharp screen is not bad</td>
</tr>
<tr>
<td>So depressing, sold my laptop and just thought about coming here:'(‘</td>
<td>Sold my laptop and just thought about coming here</td>
</tr>
<tr>
<td>Anyways, how does everyone clean stains!?</td>
<td>How does everyone clean stains?</td>
</tr>
<tr>
<td>Urgent!!!When is the student version of the ThinkPad coming out?</td>
<td>When is the student version of the ThinkPad coming out?</td>
</tr>
</tbody>
</table>
4.3.7 Measurement Development

Existing scales were adopted to measure the research construct of this study and Table 2 presents where the items of scale come from. The community perceived value and perceived functional value were adopted from Korgaonkar and Wolin (1999), Cha and Wang (2006), and Jin (2007), and the perceived recreational value was adopted from Torkzadeh and Dhillon (2002), Cha and Wang (2006), and Jin (2007). The perceived social value and perceived financial value were measured by adopting items from Jin (2007). In addition, the personal involvement inventory (PII) was used to measure emotional involvement.

We measured predictive confidence by adopting items from Yaveroglu (2002), which was measured using a seven-point Likert scale. Specifically, for different dimensions of perceived value, we set the question about predictive confidence just after questions about community perceived value for each group, and the related statement was “I am confident in my judgment of community perceived value in the above four aspects.”

Table 2

<table>
<thead>
<tr>
<th>Scale</th>
<th>Item</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotional involvement</td>
<td>Interesting</td>
<td>The Personal Involvement Inventory</td>
</tr>
<tr>
<td></td>
<td>Appealing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fascinating</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Exciting</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Involving</td>
<td></td>
</tr>
<tr>
<td>Community perceived value</td>
<td>Perceived functional value</td>
<td>Korgaonkar, Wolin, 1999; Cha, 2006; Jin, 2007</td>
</tr>
<tr>
<td></td>
<td>Perceived recreational value</td>
<td>Torkzadeh, Dhillon, 2002; Cha, 2006; Jin, 2007</td>
</tr>
<tr>
<td></td>
<td>Perceived social value</td>
<td>Jin, 2007</td>
</tr>
<tr>
<td></td>
<td>Perceived financial value</td>
<td>Jin, 2007</td>
</tr>
<tr>
<td>Predictive Confidence</td>
<td>Confident in evaluation</td>
<td>Yaveroglu, 2002</td>
</tr>
</tbody>
</table>

4.4 Procedure

We did a pretest to check whether the emotional involvement of a strong emotional post title list is significantly higher than that of the weak emotional post title list \( (F = 3.664, S = 0.069, r < 0.1) \), but the cognitive involvement of two groups did not show a significant difference \( (F = 1.168, S = 0.298, r < 0.1) \). Furthermore, most subjects in the pretest only needed 40 s to finish reading the post title list and hence, we changed the reading time of the experiment to 40 s.

The procedure of the formal experiment is as follows: First, subjects are told to read the instruction which explained the purpose and anonymity of the study.

Then, the subjects are presented with background information as given below:

As a college student, it is important for you to get a computer to help you with your schoolwork and you find it inevitable to get one. Looking around, you notice that some people are using ThinkPad. What is this computer like? You want to know more about it. Thus, you look up the information about this computer on the Internet and you come across a group about ThinkPad on Douban.com. You enter the group and browse the front page of the community. Please answer the questions below according to what you see in these pictures.

Thirdly, we presented the front page of this community to the subjects and asked them to browse the webpage carefully. Subjects are divided into two groups by their student numbers. Both post title lists of the webpages of this community are shown to the two groups, but the difference was that students whose student number was an odd number were shown a strong emotional post title list whereas students whose student number was an even number were given a weak emotional post title list. After 40 s, the students were asked to stop browsing.
Finally, the students were required to complete or fill up the involvement scale data and scale of community perceived value and to fill in the demographic options.

5 Data Analysis and Results

5.1 Manipulation Check and Demographic Characteristics

The subjects of the formal experiment were 118 undergraduate business students at Sichuan University in China. There were 52 males and 64 females, accounting for 44.8% and 55.2% of the sample, respectively. We conducted a chi-square test on gender and group and found no significant correlation between the two \((p > 0.1)\), which means that the male and female students were evenly divided into two experimental groups. Additionally, there was no significant correlation between gender and brand familiarity \((S = 0.471)\) according to the crosstabs of brand familiarity and gender. Therefore, the interaction between the two has no relation to the data results, and the main effect of the variables was a valid one.

5.2 Reliability and Validity Analysis

We tested the reliability of three scales using Cronbach’s alpha. As shown in Table 3, personal involvement, community perceived value, and predictive confidence were found to have good reliability, with Cronbach’s alpha values between 0.854 and 0.918.

We conducted Kaiser–Meyer–Olkin (KMO) and Bartlett tests on personal involvement scale. It was found out that the KMO value was 0.869, and the Bartlett value was less than 0.01, which indicated that the data of this scale were suitable for factor analysis. Thus, we conducted a factor analysis for the personal involvement scale. After rotation, the factor analysis extracted two factors, which accounted for 70% of the total variance, and the details are given in Table 4. Specifically, the first factor mostly depicted the respondents’ emotional involvement, while the second factor showed the respondents’ cognitive involvement. Then, the emotional involvement factor and cognitive involvement factor were extracted by analyzing the personal involvement scale, consistent with other researcher’s study, which also indicated that the scale had good construct validity.

As for as the scale of community perceived value is concerned, the results of the KMO and Bartlett tests showed that the data of this scale were also appropriate for factor analysis, with the KMO value above 0.7 \((=0.789)\) and Bartlett test value below 0.01. Based on the factor analysis after rotation, we extracted four factors, which accounted for 67.8% of the total variance and the details are given in Table 5. More specifically, the first factor mainly described the respondents’ perceived functional value, and the second factor mostly showed the respondents’ perceived recreational value. Furthermore, the third factor depicted the perceived financial value of the respondents, and the fourth factor revealed the perceived social value of the respondents. So, based on the factor analysis, we extracted four factors, namely, community perceived functional value, community perceived recreational value, community perceived financial value, and community perceived social value. These values are consistent with other researcher’s study indicating that this scale has good construct validity.
Hypothesis Testing

After examining the validity and reliability of the study, we tested the proposed hypotheses using SPSS 17.0.

5.3.1 Tests of Main Effects

5.3.1.1 Effects of the Strength of Emotional Cues of the Post Title List on the Community Perceived Value

We performed one-way ANOVA using the emotional strength of the post title list as the independent variable and four dimensions of community perceived value as dependent variables and the results are shown in Table 6. According to the results of the one-way ANOVA, the strong emotional cue group and weak emotional cue group showed significant differences in community perceived recreational values ($F = 4.659, p < 0.1$) and community perceived financial values ($F = 2.848, p < 0.1$). Furthermore, in terms of the community perceived recreational value, the mean of the strong emotional cue group was significantly higher than that of the weak emotional cue group, whereas for the community perceived financial value, the mean of the strong emotional cue group was significantly lower than that of the weak emotional cue group. Additionally, when the community perceived functional value and community perceived social value are considered, two groups show no significant differences. Therefore, it is observed that $H_{1a}$ and $H_{1d}$ were supported while $H_{1b}$ and $H_{1c}$ were not supported.

5.3.1.2 Effects of the Emotional Strength of the Post Title List on Predictive Confidence

We performed the one-way ANOVA using the emotional strength of the post title list as the independent variable and predictive confidence (that is, the scores of the community perceived value of the four dimensions) as the dependent variable. The results (Table 7) indicated that the strong emotional cue group and weak emotional cue group showed significant differences in terms of the community perceived functional value ($F = 4.930, p < 0.1$), community perceived recreational value ($F = 4.733, p < 0.1$), and community perceived financial value ($F = 3.258, p < 0.1$). More importantly, when compared to the weak emotional cue group, the strong emotional group had higher predictive confidence on these three aspects, whereas these two groups showed no significant difference in the aspect of community perceived social value. In summary, strong emotional cues deepen the users’ predictive confidence of the community perceived functional value, community perceived recreational value, and community perceived financial value. Thus, $H_3$ was supported.
5.3.2 Test of Mediating Effect

As mentioned above, this study proposed that emotional involvement mediated the relationship between the emotional strength of the post title list and community perceived value (H$_2$). Based on the analysis of the main effect, it is found out that the emotional strength of the post title list is mostly influenced by the community perceived recreational value and community perceived financial value and hence, we performed a test of mediating effect for these two dimensions.

We conducted a mediation analysis recommended by Baron and Kenny (1986) to test the mediated role of emotional involvement and the results are shown in Table 8. When the strength of emotional cues is considered as the independent variable and the perceived recreational value as the dependent variable, the regression analysis indicated that the strength of emotional cues positively influenced the perceived recreational value (b = 0.196**). Furthermore, when we used both the strength of emotional cues and emotional involvement as the independent variables and perceived recreational value as a dependent variable, the positive impact of the strength of emotional cues was not significant (b = 0.117), and the positive influence of emotional involvement was still of great significance (b = 0.506**), which indicated that emotional involvement completely mediated the relationship between the strength of emotional cues and perceived recreational value, which was depicted in Figure 2.

In addition, we conducted a similar test to test the mediating role of emotional involvement between the strength of emotional cues and perceived financial value. First, we employed the strength of emotional cues as the independent variable and perceived financial value as the dependent variable, and found that the former had a significant negative impact on the latter (B = −0.155*). Then, we used emotional involvement as an independent variable and perceived financial value as a dependent variable to do regression analysis, but the result indicated that the two did not have a significant relationship (B = 0.005), which means that the emotional involvement did not mediate the relationship between the strength of emotional cues and perceived financial value.

6 Conclusions and Discussion

We explored the influence of strength of post titles on community perceived value in a virtual community based on the cue utilization theory. We found out that users regarded the emotional information of post titles as cues of the community perceived value when they are entering into virtual communities. The strong emotional information of the post title has both negative and positive effects on the users’ community perceived value, that is, increasing the
community perceived recreational value while decreasing the community perceived financial value. It is very important to note that emotional involvement mediated the relationship between the emotional information of post titles and community perceived recreational value.

This paper contains three theoretical values: first, this study found that the emotional intensity of post title list can coordinate or stimulate users’ perception of virtual community value by adding a new dimension namely post title emotional expression, for users’ perception cause of virtual community value. Different from the previous studies that focused on the content of community posts as the antecedent, this study focused on the relation between the emotional expression of posting titles and the community visiting behavior with a gradual and in-depth process. Second, this study discusses the need for the virtual community to balance the interests of community managers and users of posts, and proves that there is indeed a specific conflict of interests in the expression of post titles. Third, based on the theory of cue utilization, this study proves that emotional expression of post title can act independently on consumers as an independent cue, expanding the application field of cue utilization theory, which is highly consistent with the emotion–information equivalence theory of Clore et al. (1994).

However, the conclusions above lead to more thoughts. First, existing researches on information characteristics in a virtual environment mostly focus on the quality of information content (Lin, 2009; Dholakia, Bagozzi, & Pearo, 2004) based on rationalism theoretically. However, the results of this study demonstrate that text emotion, an information factor, influences the consumers’ value judgment of virtual communities. According to the results of this study, when consumers first enter the virtual community and face or view high emotional post title lists, the traditional “cognition–emotion–behavior” consumer behavior route fails. Possible factors explaining the above phenomenon are threefold. First, people are highly uncertain about the value of the virtual brand community, which is similar to the Faraji-Rad and Pham’s (2016) study that showed uncertainty increase consumers’ reliance on affective in-puts in judgments and decisions. This uncertainty makes consumers more susceptible to emotional information and gives rise to deviation in the evaluation of virtual community value. Second, people have higher requirements or factors for rapid decision making under the network environment, which makes the title of the post emotional expression as a key clue to determine whether to further explore the virtual community. Third, widespread text symbols lead to a relative lack of attention, thus causing emotional cues to exclude other useful information.

Moreover, existing studies on virtual communities involve only community managers and community users, ignoring the important role of information content producers. However, this study shows that, once the interest of the information content producers is not similar or at odds with that of community managers and users, the information content producers may use the cueing effect of information to meet their own interest first, which may indirectly harm the interest of community managers and users. It was also noted by Mcalexander, Schouten, and Koenig (2002) that there are four kinds of relationship in virtual brand communities that should be paid attention to, which include customers and enterprises, customers and products, customers and brands, and customers and customers. Moreover, as the types of customers in virtual communities are becoming more diverse and so the classification of customers should be considered and information content producers, a new type of customer, should be paid attention in particular. But there are only few researches involving this subject at present. Further, when visiting a virtual community for the first time, users tend to rely more on information cues for indirect reasoning to judge the value of the community, whereas the information content producers only care about personal information visits and build the post titles more emotional to attract users, which may mislead users leading to damaging the interest of both the virtual community managers and users at the same time.

Third, the cue utilization theory has been applied to the business field under the physical environment only, but has not yet been introduced to study the network environment. The results of this study demonstrate that...
emotional texts also have a cueing effect under the network environment, and the phenomenon of making the post titles more emotional to get attention is a manifestation of the cueing effect, which may lead to information deterioration under the network environment. Individuals need to select, accept, and digest more symbols in the network environment, so they can conduct selective perception, in which emotional symbols have the priority to be processed (Flykt & Esteves, 2001), which provide a basis for emotional symbols to play the cueing effect. According to symbolic interactionism, it is reasonable for information producers to use the above rule as individuals will always instinctively adjust themselves according to the reaction of the other individual, which, in turn, becomes a new stimulus for the other individual to change his or her own behaviors (Mead, 1992). Therefore, the behavior of making the post titles more emotional is actually the result of the symbolic interaction between information content producers and information content users, and continuous development of this interaction may result in sentiment polarization of the whole community. In summary, virtual community managers should induce or compel the information content producers to take the community’s performance into consideration and form a win–win–win relationship, and to achieve that a certain mechanism needs to be established to generate positive and consistent symbol interactions between information content producers and community managers, community managers and community users, and information content producers and community users.

7 Practical Implication

This study has the following implications for information management in virtual communities:

First, virtual community managers need to be aware that they should take both information content and information expression as management objects. Nowadays, online expression of emotion is becoming increasingly extreme, and the phenomenon of making the post titles more emotional disturbing the community order is becoming more and more serious, and so it is imperative for virtual community managers to put the emotional expression of community information into the management scope to ensure the healthy growth of the community.

Second, emotional titles have two sides; hence, community managers should choose specific information management strategies based on the different value pursuit of community users. As the emotional cues of the post title have different effects on the perceived value of the community at different latitudes, virtual community managers should choose specific information governance policies according to different value quest of the users in the community. Moreover, in a virtual community where users mainly pursue entertainment value, there is no need to have many constraints on the behaviors of strengthening emotion of the post titles. While in virtual communities where users mainly pursue financial value and functional value, behaviors of using strong emotional titles to attract the visitors’ attention should be strictly controlled.

Third, managers of virtual communities should optimize the allocation of information resources and upgrade the formation mechanism of post title lists to avoid the negative effect of emotional titles on community information processing and to make full use of the positive effect. As the functional rules of the emotional information of the post titles have been mastered nowadays, they can be used as one of the quality evaluation indexes of the post titles by using a sorting algorithm. Moreover, users’ preferences for emotional information can be assessed based on community and user characteristics, and a sorting algorithm can also be built based on the quality of the community’s existing posts and data of emotional information strength. Meanwhile, for posts of high content quality, the emotional information strength of post titles should be regarded as a value-added factor when making a quality calculation. However, the emotional information strength of post titles should be regarded as a value-lessen factor for low-content-quality posts; therefore, the misleading effect of emotional information in the processing of cues can be decreased, and its positive value can be highly increased in the formation stage of post title lists.

8 Limitations and Future Research

This study has some limitations that warrant further research. First, since we explored the cueing effect of post titles’ emotional strength chiefly, other emotional information attributes like color emotion were not discussed, which may also influence the perceived value of users. Second, the specific context that we chose to carry on in this study is the context of the computer brand community; so, the generality of the findings may be limited. Future research should consider exploring the inherent factors causing the cueing effect of emotional post
titles. It is necessary to have future research conducted in various kinds of virtual communities. Furthermore, future studies should further explore the cueing effect of color emotion and even the interaction between emotional strength and color emotion.

Acknowledgements: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References


Appendix

*Questionnaire Construct and Items*

There are two groups of questionnaires in which guide language, background information, brand familiarity scale, perceived community value scale, and population attribute scale are the same. There are two different versions of the title list pages (strong textual emotional cues and weak textual emotional cues). The contents include an introduction, background, brand familiarity scale, post title list pages, personal involvement scale, perceived community value scale, and demographic attribute scale.

1  Introduction

Dear Sir and Madam,
It is widely recognized that the Internet has been an inseparable part of our life, which brings a variety of convenience to us. To enhance people’s understanding of the Internet environment and consumer behavior, we are committed to doing research related to the virtual community. Please read related materials carefully and tick the answers according to what you really feel.
This is an anonymous survey, and the data will only be applicable to this study.
Thank you for your participation.

College of Business at Sichuan University

2  Background

As a college student, it is important for you to get a computer to help you with your schoolwork, and you find it inevitable to get one.

Looking around, you notice that some people are using ThinkPad. What is this computer like? You want to know about it. Thus, you look up the information about this computer on the Internet, and you come across a group about ThinkPad on Douban.com. You enter the group and browse the front page of the community. (The pictures are as follows.) Please answer the questions below according to what you see in these pictures.

3  Brand Familiarity

Please answer the questions below before having a look at these pictures.
Do you know much about ThinkPad? (Choose a number below)
1. Don’t know it at all. 2. Don’t know it much. 3. Know a little bit about it. 4. Know it pretty well. 5. Know it very well.

4  Post Title List Pages

Please have a look at the pictures and read the text carefully.
### ThinkPad爱好者小组

<table>
<thead>
<tr>
<th>贴图</th>
<th>作者</th>
<th>回复</th>
<th>最后回复</th>
</tr>
</thead>
<tbody>
<tr>
<td>我终于把ThinkPad的声音特别小，有木有！</td>
<td>arrowsTime</td>
<td>1</td>
<td>10-18 23:28</td>
</tr>
<tr>
<td>那啥……大家都是如何清理键盘的呢？</td>
<td>物京</td>
<td>1</td>
<td>10-18 23:09</td>
</tr>
<tr>
<td>想买一个X230，请推荐一个卖家</td>
<td>爱爱丝</td>
<td>5</td>
<td>10-18 14:21</td>
</tr>
<tr>
<td>T60的性价比如何</td>
<td>eyes on book</td>
<td>3</td>
<td>10-18 12:44</td>
</tr>
<tr>
<td>高分屏的T43P真给力</td>
<td>odlews25</td>
<td>10</td>
<td>10-03 13:45</td>
</tr>
<tr>
<td>真郁闷，出了笔记本才想起这里</td>
<td>柚子</td>
<td>5</td>
<td>10-09 13:44</td>
</tr>
<tr>
<td>漫骂流泪ThinkPad</td>
<td>going on</td>
<td></td>
<td>10-09 09:12</td>
</tr>
<tr>
<td>麻烦大家对比这个E系列</td>
<td>zhouchedad</td>
<td>10</td>
<td>10-08 22:40</td>
</tr>
<tr>
<td>小黑thinkpad t40终于上市啦啦啦啦啦啦啦啦啦啦啦啦啦啦啦啦啦</td>
<td>Deepthinker</td>
<td>10</td>
<td>10-05 18:30</td>
</tr>
<tr>
<td>thinkpad小知识</td>
<td>Mathilda</td>
<td>1</td>
<td>10-04 06:55</td>
</tr>
<tr>
<td>我感冒了，想知道R系列已经发了吗？</td>
<td>rakingroll</td>
<td>24</td>
<td>10-02 12:24</td>
</tr>
<tr>
<td>求购一台全新的thinkpad X201</td>
<td>小草出</td>
<td>4</td>
<td>10-01 09:59</td>
</tr>
<tr>
<td>我的小黑漏电了……怎么办怎么办怎么办</td>
<td>青蛙花花螃蟹</td>
<td>2</td>
<td>10-01 20:55</td>
</tr>
<tr>
<td>我还用X30，左组里水平</td>
<td>肖像</td>
<td>3</td>
<td>10-01 17:40</td>
</tr>
<tr>
<td>有对1、3小黑情有独钟吗</td>
<td>Charfes</td>
<td>7</td>
<td>09-28 10:58</td>
</tr>
<tr>
<td>新入手新到货T500保存素材T4550</td>
<td>zhouchedad</td>
<td>11</td>
<td>09-26 08:15</td>
</tr>
<tr>
<td>带鼠标不稀的还是普通小黑？？纠结中，求指导</td>
<td>Terry</td>
<td>2</td>
<td>09-24 18:28</td>
</tr>
<tr>
<td>ThinkPad一般多久该清理灰尘？</td>
<td>虚幻四维之鸟</td>
<td>1</td>
<td>09-24 13:58</td>
</tr>
<tr>
<td>出自用SL500</td>
<td>Laputa小灣</td>
<td>9</td>
<td>09-23 17:49</td>
</tr>
<tr>
<td>求解！！！9000大洋E40让我无限解烦恼啊！！！</td>
<td>林林西</td>
<td>4</td>
<td>09-22 20:29</td>
</tr>
<tr>
<td>想入X230，求买配置</td>
<td>孙甲出</td>
<td>6</td>
<td>09-22 20:27</td>
</tr>
<tr>
<td>急！！！THINKPAD新生机究竟什么时间能出来啊。。</td>
<td>必定不惑</td>
<td>1</td>
<td>09-21 11:20</td>
</tr>
<tr>
<td>我是很依赖大家意见的</td>
<td>哈哈哈</td>
<td>7</td>
<td>09-21 02:13</td>
</tr>
<tr>
<td>各位！！！X1250c怎么样啊！！！</td>
<td>陰陽魚</td>
<td>7</td>
<td>09-21 02:11</td>
</tr>
<tr>
<td>T430使用了日立7m硬盘</td>
<td>Wayne</td>
<td>2</td>
<td>09-21 00:47</td>
</tr>
</tbody>
</table>
Version II

ThinkPad爱好者小组

<table>
<thead>
<tr>
<th>讨论</th>
<th>作者</th>
<th>回应</th>
<th>最后回应</th>
</tr>
</thead>
<tbody>
<tr>
<td>thinkpad 的声音特别小</td>
<td>arrowtime</td>
<td>1</td>
<td>10-10 23:20</td>
</tr>
<tr>
<td>大家都是如何清理灰尘的呢?</td>
<td>！！！</td>
<td>1</td>
<td>10-10 23:09</td>
</tr>
<tr>
<td>想买一个 X230, 请推荐一个卖家</td>
<td>爱丽丝</td>
<td>5</td>
<td>10-10 14:21</td>
</tr>
<tr>
<td>T40 的性价比如何</td>
<td>eyes on book</td>
<td>3</td>
<td>10-10 12:44</td>
</tr>
<tr>
<td>高分屏的 T43p 很不错</td>
<td>sdlee1925</td>
<td>10</td>
<td>10-09 13:49</td>
</tr>
<tr>
<td>出了笔记本才想起这里</td>
<td>南风</td>
<td>5</td>
<td>10-09 13:44</td>
</tr>
<tr>
<td>沉鱼落雁 thinkpad</td>
<td>going on</td>
<td>1</td>
<td>10-09 02:12</td>
</tr>
<tr>
<td>大家对比这个 E 系列</td>
<td>zhousoledad</td>
<td>1</td>
<td>10-08 22:40</td>
</tr>
<tr>
<td>thinkpad t50 上市</td>
<td>Deep thinker</td>
<td>1</td>
<td>10-05 18:30</td>
</tr>
<tr>
<td>thinkpad 小知识</td>
<td>Mathilda</td>
<td>1</td>
<td>10-04 05:55</td>
</tr>
<tr>
<td>R系列已经没了?</td>
<td>rackingroll</td>
<td>24</td>
<td>10-02 12:24</td>
</tr>
<tr>
<td>求购一台全新的 thinkpad X201</td>
<td>马肚子</td>
<td>4</td>
<td>10-01 09:59</td>
</tr>
<tr>
<td>我的 thinkpad 漏电了</td>
<td>竹竿绕开青苔</td>
<td>1</td>
<td>09-30 23:55</td>
</tr>
<tr>
<td>我还在用 X40，在组里什么水平</td>
<td>安眠酒</td>
<td>3</td>
<td>09-29 17:40</td>
</tr>
<tr>
<td>你认为 4; 5 thinkpad 怎么样</td>
<td>Charles</td>
<td>7</td>
<td>09-28 10:50</td>
</tr>
<tr>
<td>助人最新到货 T500 库存新机 T650</td>
<td>zhousoledad</td>
<td>11</td>
<td>09-25 08:15</td>
</tr>
<tr>
<td>选带翻转笔触的还是普通 thinkpad?</td>
<td>Terry</td>
<td>1</td>
<td>09-24 18:28</td>
</tr>
<tr>
<td>thinkpad 一般用多久该清理灰尘?</td>
<td>黄鹂因材之分身</td>
<td>9</td>
<td>09-24 13:58</td>
</tr>
<tr>
<td>出自用 T500</td>
<td>Laputa今非昔比</td>
<td>9</td>
<td>09-23 17:49</td>
</tr>
<tr>
<td>9000 天洋 E50</td>
<td>神来之</td>
<td>4</td>
<td>09-22 20:20</td>
</tr>
<tr>
<td>想入 T220，求推荐配置</td>
<td>马来西亚</td>
<td>5</td>
<td>09-22 20:27</td>
</tr>
<tr>
<td>THINKPAD 学生机什么时候能出来?</td>
<td>心在不感</td>
<td>1</td>
<td>09-21 11:20</td>
</tr>
<tr>
<td>我是被依赖大家意见的</td>
<td>你</td>
<td>7</td>
<td>09-21 02:13</td>
</tr>
<tr>
<td>X120e 怎么样</td>
<td>陆地乐</td>
<td>7</td>
<td>09-21 02:11</td>
</tr>
<tr>
<td>T420 使用了日立 7MM 硬盘</td>
<td>Wayne</td>
<td>2</td>
<td>09-21 00:47</td>
</tr>
</tbody>
</table>
5 Personal Involvement Scale

There are several sets of statements. Please choose the corresponding figure according to your degree of agreement on the statements (tick the number).

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly disagree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>This group is important for me to fulfill my task.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>This group is highly relevant to my needs.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>This group is valuable.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>I am in great need of this group.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>This group is meaningful to me.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>This group is interesting.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>This group is attractive.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>This group is exciting.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>This group can easily get people focused.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
</tbody>
</table>

6 Perceived Community Value Scale

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly disagree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>The group provides ample and novel information.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>Readers can be benefited a lot from this group.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>People can get a great source of knowledge in this group.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>Consumers can get the information and materials they need.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>I am confident with the answers to the four questions above.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>Its articles, activities, and communication meetings are enjoyable.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>It is interesting to look through the group.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>It is pleasant to look through the group.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>It cheers people up if they log in the group and take part in the group activities when they are boring.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>I am confident with the answers to the four questions above.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
</tbody>
</table>
An Empirical Study on the Cueing Effect of the Emotional Post Title in a Virtual Community

It can help people to make some friends with the same interests.

Strongly disagree 1 2 3 4 5 6 7

Strongly agree

It can enrich the network of interpersonal relationships.

Strongly disagree 1 2 3 4 5 6 7

Strongly agree

It is an important platform to get contact with fans of other computer brands and to exchange experience.

Strongly disagree 1 2 3 4 5 6 7

Strongly agree

Its related activities can enrich the social life of its members.

Strongly disagree 1 2 3 4 5 6 7

Strongly agree

I am confident with the answers to the four questions above.

Strongly disagree 1 2 3 4 5 6 7

Strongly agree

Being a member of it helps me get some economic benefits to some degree.

Strongly disagree 1 2 3 4 5 6 7

Strongly agree

The members can get some economic benefits by joining the group.

Strongly disagree 1 2 3 4 5 6 7

Strongly agree

Members of the group will enjoy some special offers in terms of consumption.

Strongly disagree 1 2 3 4 5 6 7

Strongly agree

The group can help consumers save some money.

Strongly disagree 1 2 3 4 5 6 7

Strongly agree

I am confident with the answers to the four questions above.

Strongly disagree 1 2 3 4 5 6 7

Strongly agree

7 Demographic Attribute Scale and Acknowledgment

Your basic information
Your gender: O Male O Female
Your age:__________

Thank you so much for your support!
Information Search Trail Recommendation Based on Markov Chain Model and Case-based Reasoning

Abstract: An information search trail recommendation method based on the Markov chain model and case-based reasoning is proposed. A laboratory user experiment was designed to evaluate the proposed method. The experimental results demonstrated that novice searchers have a positive attitude toward the search trail recommendation and a willingness to use the recommendation. Importantly, this study found that the search trail recommendation could effectively improve novice searchers’ search performance. This finding is mainly reflected in the diversity of information sources and the integrity of the information content of the search results. The proposed search trail recommendation method extends the application scope of information recommendations and provides insights to improve the organization and management of online information resources.

Keywords: Web search, search process, search trail recommendation, Markov chain, case-based reasoning

1 Introduction

Using Web search engines, people can obtain a piece of information or navigate to a target website quickly. However, frequently, many information search tasks are complex or exploratory, such as acquiring knowledge items of a particular subject or writing a course thesis. The information needs of these tasks include multiple aspects or steps, and the search processes often require access to different types of websites. However, in practice, many novice searchers find that locating the appropriate websites efficiently is difficult. As a result, many specialist websites fail to establish connections with target users. Fortunately, expert searchers with certain domain knowledge, search experience, or skills can conduct efficient searches and locate the appropriate websites or website sequences to find what they are looking for (Tabatabai & Shore, 2005). White, Dumais, and Teevan (2009) recommend that query suggestions and website recommendations generated through domain experts’ search history could be provided to novice searchers to help them gain expertise. They even envisaged developing such Web search support services in future research.

Some previous studies have attempted to support new searchers by providing webpage recommendations (Hendahewa & Shah, 2017; White, Bilenko, & Cucerzan, 2007), optimizing webpage rankings (Ziegler, McNee, Konstan, & Lausen, 2005), and query expansion (Smith, Gwizdka, & Field, 2016; Huang, Wang, Zhang, & Liu, 2020). However, such supports are often insufficient to meet complex information needs. Novice searchers may need support that alerts them to the steps, or a webpage sequence, or websites required for complex task completion.

Previous studies have shown that trails or tours consisting of filtered documents or webpages can reveal the value of user search processes, and trail-based search recommendations could improve new searchers’ overall search performance. For example, White and Huang (2010) demonstrated that following search trails provides significant additional benefits to searchers in terms of coverage, diversity, novelty, and utility over origin pages.
and destination pages. Hendahewa and Shah (2017) found that, with the help of search trail recommendations, searchers can find more information across multiple facets and dig deeper into the detail associated with certain facets.

However, most search trails mined from logs often only contain pages visited after a single query rather than all pages or websites visited during the full search process of a search task. Besides, to date, proposed methods for finding trails often focus on trails consisting of webpages that may be sensitive to Web dynamism (e.g., dead links or changing content), rather than a more general level of abstraction (e.g., website categories) that may be more widely applicable. Moreover, showing the trails to searchers directly on the search engine result page (SERP) is also an unaddressed challenge (Hassan & White, 2012). Addressing the shortcomings of previous studies, our primary goal is to present and evaluate a method to create search trails that can help novice searchers perform complex search tasks. We focus on (i) how to model and recommend search-task relevant trails of expert searchers and (ii) whether the search trail recommendation is useful for novice searchers.

The remainder of this paper is organized as follows. An extensive literature review of search trail recommendation studies, the utility of the Markov chain model for Web path analysis, and the case-based reasoning (CBR) approach are presented in Section 2. In Section 3, we introduce our proposed information search trail recommendation method based on the Markov chain model and the CBR approach. In Section 4, we describe a user study designed to evaluate the search trail recommendation method. The user study was conducted in a laboratory environment, and the results were analyzed to determine how novice searchers evaluate the effectiveness of search trail recommendations. The implications and limitations of this research and suggestions for future research are summarized in Section 5.

## 2 Related Work

Several research areas are relevant to the current study: (i) search trail modeling and recommendations, (ii) the Markov chain model, and (iii) CBR searching. In this section, we describe relevant studies in each area in more detail and discuss how the current research could extend them.

### 2.1 Search Trails Modeling and Recommendation

Some researchers defined the search trail as a search path that comprises query and post-query pages visited to carry out relevant studies (Bilenko & White, 2008; Singla, White, & Huang, 2010; Hendahewa & Shah, 2017). Recently, Capra and Arguello (2019) defined a search trail as an interactive visualization of how a searcher performs a search task. The visualization may include queries posted, sites or pages visited, and annotations made. It is evident that the latter definition has a larger scope and greater applicability. Thus, we adopted the second definition to carry out our research.

Berrypicking, orienteering, and information foraging are three well-known information seeking models related to search trails. The berrypicking model describes the movement between information sources associated with dynamic information needs (Bates, 1989). The orienteering analogy was proposed to understand searchers’ information seeking strategies (O’Day & Jeffries, 1993). Information foraging emphasizes how information searchers use clues left by previous visitors to find information (Pirolli & Card, 1999).

Interaction logging has implicitly made us all trailblazers of search trails (White & Huang, 2010). Trails or tours created by previous searchers form links between stored information resources that can help other searchers make better decisions about information source selection during the search process. Trigg (1988) proposed a guided tour consisting of a sequence of hypertext pages to alleviate disorientation for new searchers. Wheeldon and Levene (2003) proposed an algorithm to generate trails as trees to assist searchers in Web navigation. The results showed that participants found these trails useful for navigation. Singla, White, and Huang (2010) proposed trailfinding methods to support Web searches by identifying query-relevant trails from logs that could be shown to complement or replace traditional search result lists. Recently, Hendahewa and Shah (2017) demonstrated that recommending search trails of each query to struggling users in exploratory search tasks could better assist them to find the information they were seeking. Moreover, the results showed that the order of the recommended search trails plays an important role. Capra and Arguello (2019) discovered that task determinability is a significant factor that affects whether to recommend search trails. Besides, they found that the system should provide trails with the same scope as the searcher’s task. Search trail recommendation systems, such as WebWatcher (Joachims, Freitag, & Mitchell, 1997), Footprints (Wexelblat & Maes,
Markov chains are useful tools for Web link sequence modeling and path analysis. Liu, Huang, and An (2007) proposed a mixture of Markov models to cluster searchers, capture the sequential relationships in searchers’ access histories, and provide searchers with personalized recommendations.

In this paper, we apply the Markov chain model to analyze the search trail of site nodes. The state transition matrix of the Markov chain model can be considered as a “weighted traversal” representation of the user’s model of the Web space, and further analysis can be performed on this matrix, such as link relationships between different categories of websites.

2.3 CBR Searching

The principle of the CBR approach is analogical reasoning, and its basic idea is that new problems can be solved with the help of the solutions to similar past problems (Gentner, 1983; Hüllermeier, 2007). The knowledge base of the CBR system consists of a collection of cases and a set of search criteria used to retrieve cases similar to the target problem (Althuizen & Wierenga, 2014). A historical case in the case base is represented as

\[ c = (\text{Specification}, \text{Solution}) \]

where Specification is the description of the problem consisting of n features, and Solution provides the solution to the problem. The CBR approach was originally applied in the field of artificial intelligence (Aamodt & Plaza, 1994). And now, the CBR approach has since been applied in many other fields, such as business (Gavetti, & Rivkin, 2005; Goldstein, 2001; Gregan-Paxton & Cote, 2000), medical diagnosis (Bichindaritz & Marling, 2006), information seeking (He, Erdelez, Wang, & Shyu, 2008; Alptekin & Büyükozkan, 2011), engineering (Shokouhi, Skalle, & Aamodt, 2014), architecture, and law (Bridge, Göker, McGinty, & Smyth, 2006; Hüllermeier, 2007). For example, doctors may benefit from using a CBR system that accesses the case of a previously treated patient with symptoms similar to those of a new patient (Bichindaritz & Marling, 2006).

Although many studies have focused on search trail recommendations, few studies consider CBR searching. Moore, Erdelez, and He (2006) conducted a controlled experiment that demonstrated the difference between traditional keyword searching and CBR searching. However, the authors did not explain why the difference occurs. He and Tian (2017) conducted an 8-year longitudinal analysis of the query logs of a Web-based case library system. They found those return users employed CBR searching much more frequently than one-time users.
In addition, return users tended to use more query terms to find information.

We assume that a recommendation based on the CBR approach might recommend search strategies that reflect how experts conducted search tasks (e.g., search trails at the websites level) to novice searchers who perform the same or similar tasks, which will extend the scope of search recommendation. Therefore, in this study, the CBR method was adopted by storing expert searchers’ search trails of site nodes in the case base and generating recommendations for novice searchers who need to solve the same or a similar task.

3 Search Trail Recommendation Method Based on Markov Chain Model and CBR Approach

In this section, we describe the search trail recommendation procedure. First, the Markov chain model was used to extract the search trails of the user information search process. The full search trail, that is, the sequence of Internet information sources that experts visited, and one-step trail, that is, the transfer probability of Internet information sources categories, are identified and stored in the form of a case base that serves as the data source of information search trail recommendations for novices. Then, based on the CBR approach, the expert search trail is re-used as recommended content to improve a novice’s information search experience and information resource utilization efficiency.

3.1 Search-Task Relevant Search Trails Generation

The search task is the original driving force of a searcher’s information seeking behavior. The search task shapes the searcher’s interaction with various information sources during the search process (Li & Belkin, 2010). Task attributes, such as type, complexity, and goal, greatly influence how a searcher formulates their search strategy and their search behavior (Kim, 2009; Li, 2009; Li & Belkin, 2010).

First, the sequence in which searchers access Internet sources during a search task must be identified. A sequence of Internet sources accessed by a searcher when performing a search task is described in Equation (1) as follows:

$$SrcSeq = \left\{ Src_{1}, Src_{2}, \ldots, Src_{i}, \ldots, Src_{t} \right\}$$

To handle the large volume of Internet sources, we propose using Internet source categories to reduce the state space of Markov models as follows:

$$SrcCatSeq = \left\{ SrcCat_{1}, SrcCat_{2}, \ldots, SrcCat_{i}, \ldots, SrcCat_{t} \right\}$$

In Equation (2), $SrcCat_{i} \in Cat = \{ cat_{1}, cat_{2}, \ldots, cat_{n} \}$, $n$ represents the number of Internet sources categories and indicates that there are $n$ states in the Markov chain model.

Then, deploying the CBR approach, the Markov chain model of search-task relevant search trails can be expressed as follows:

$$tmc = (T, Cat, S, SC, A)$$

Here, $T$ is a set of properties that describe the search task and also represents a set of criteria for retrieving cases that are similar to the target search task. $S$ denotes collections of Internet source sequences and $SC$ represents information source category sequences generated by expert searchers when performing certain information search tasks. $A$ is the transition probability matrix of the search trail Markov model of certain information search tasks. The transition probability matrix can be trained using historical data. Without loss of generality, this study uses the principle of maximum likelihood to estimate $A$. Mathematical representations of $T$, $S$, $SC$, and $A$ are given as follows:

$$T = (a_{1}, \ldots, a_{m})$$

$$S = \left\{ SrcSeq_{1}, SrcSeq_{2}, \ldots, SrcSeq_{k} \right\}$$

$$SC = \left\{ SrcCatSeq_{1}, SrcCatSeq_{2}, \ldots, SrcCatSeq_{k} \right\}$$

$$A = \left( p_{ij} \right) = \begin{bmatrix}
    p_{11} & p_{12} & \cdots & p_{1j} & \cdots & p_{1n} \\
    p_{21} & p_{22} & \cdots & p_{2j} & \cdots & p_{2n} \\
    \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
    p_{nj} & p_{j2} & \cdots & p_{jj} & \cdots & p_{jn} \\
    \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
    p_{n1} & p_{n2} & \cdots & p_{nj} & \cdots & p_{nn}
\end{bmatrix}$$
Then, based on the Markov chain model, the search-task relevant search trails case base, which we refer to as TMC, is formally expressed as follows.

\[
TMC = \{tmc_i | tmc_i = (T_{mc_i}, Cat_{mc_i}, S_{mc_i}, SC_{mc_i}, A_{mc_i}), i = 1, \ldots, N \}
\]  

(9)

In Equation (9), \(tmc_i\) represents the search trail recommendation case of a certain type of search task, and \(N\) represents the number of cases in the case base.

### 3.2 Search Trail Recommendation Procedure

#### 3.2.1 Retrieving Similar Relevant Cases

To retrieve the expert search trails for the recommendation, the similarity of task characteristic attributes between the current search task (\(SrchTsk\)) and the task in the search trails case base should be calculated. The greater the similarity value, the more relevant the case. When the similarity value is higher than a certain threshold, the case will be added to the expert search trail recommendation case set, \(ET\), which is expressed as follows:

\[
ET(\text{SrchTsk}) = \{tmc_i | \text{sim}(\text{SrchTsk}, T_{mc_i}) \geq q, \forall tmc_i \in TMC\}
\]  

(10)

#### 3.2.2 Search Trail Recommendation During Search Process

The initial interface of the search trail recommendation system is shown in Figure 1. After the searcher inputs the description of the search task, the search trail recommendations list ranked by similarity will be generated in a timely manner. The left part of the interface presents the default search engine’s results list, and the right part presents the search trail recommendations. The upper half of the right part presents the S and SC lists, that is, the full search trail list. The lower half displays the Internet source categories that expert searchers are most likely to investigate next under the current Internet source, that is, the one-step trail recommendations.

If the vector \(s(t + 1)\) denotes the probability vector for all states at time \(t + 1\). Given the history of “Internet information sources” of the novice searcher \(s(t - k), s(t - k + 1), \ldots, s(t)\), each Internet information source could be represented as a vector with probability 1 at
that state for that time (denoted $i(t-k), i(t-k+1), ..., i(t)$). The Markov chain model estimation of the probability of being in a given state at time $t+1$ is shown in Equation (11).

$$\hat{s}(t+1) = \hat{i}(t) A$$ (11)

Then, the one-step search trail recommendations list ranked in descending probability value in vector $s(t+1)$ would be provided. The interface for a specific Internet source is represented in Figure 2. The Internet source cases under each category are given, as shown in Figures 1 and 2.

4 Evaluation

In this section, we describe a user study that was designed to evaluate the effectiveness of the search trail recommendation method proposed in this paper. First, the search trails generation method proposed in Section 3 is used to identify and organize the search trails generated by expert searchers when performing some search tasks. Then, novice searchers are organized to perform the same or similar tasks, and the relevant search trail recommendations are provided during the search process. Finally, based on novice searchers’ attitudes (e.g., perceived usefulness, satisfaction, and acceptance) toward the search trail recommendations and their search results performance, the recommendation method proposed in this study can be evaluated. In addition, the characteristics of tasks and searchers that may influence a searcher’s willingness to engage with search trails, and their ability to benefit from these search trails could be studied deeply. If the results of the user study show that novice searchers benefit from search trail recommendations, likely post-task trails could be considered in real search system design and even as units of retrieval in practice.

4.1 User Experiment Design

4.1.1 Data Collection

Questionnaire surveys, screen recordings, think-aloud guidelines, and interviews were used to collect data.

Before the experiment, participants were asked to complete an entry questionnaire that was used to collect user demographic data, including search experience and habits, as well as their attitude toward sharing search trails or accepting the search trail recommendations. A software tool recorded all participant activities during the experiment. A think-aloud guideline was designed to elicit participants’ cognitive activities during their search performances. After the search experiment, a post-search questionnaire queried the novice searchers’ perceptions of their search results, search performance, difficulties, and the search trail recommendations provided during the search process. In addition, we conducted follow-up interviews to explore the novice searchers’ comments and suggestions about the search trail recommendations.

4.1.2 Participants

In this experiment, participants were recruited from an iSchool at a national major university in China. Twelve participants, all of whom had experience searching on the Internet, were recruited. Each participant was paid RMB 50 Yuan as compensation.

One senior undergraduate and four postgraduates were recruited as expert searchers (one male, four females, denoted expert $[i], i=1, ..., 5$). All expert searchers had considerable information literacy training before the experiment, and all were willing to share search trail experience with others. Seven freshmen were recruited as novice searchers (two males, five females, denoted novice $[j], j=1, ..., 7$). None of the novice searchers had information literacy training experience prior to the experiment.

4.1.3 Materials and Apparatus

4.1.3.1 Tasks

During the experiment, the subjects were asked to solve two information problems: one low complexity task and one high complexity task, as shown in Table 1. In this experiment, the task context refers to an environment that involves a combination of various factors and conditions, such as task product and search time. Participants could search for information on the Internet without any restrictions while performing these search tasks. Each task requires the submission of a document containing the required information collected and organized by the searcher.

4.1.3.2 Setting and Equipment

The information search experiment was conducted in a laboratory. EV software was installed on the computers and the entire experiment was recorded. Both screen
capture and audio recording modes were used to record the entire task performance process and the participants’ verbalized thoughts. The questionnaire survey was launched using wjx.cn, a Web-based survey tool.

4.1.4 Measurements

4.1.4.1 Internet Information Source Categories

In this experiment, considering the tasks, we classified Internet information sources into the categories listed in Table 2.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Comprehensive search engine</td>
<td>baidu.com, cn.bing.com</td>
</tr>
<tr>
<td>B Academic search engine</td>
<td>xueshu.baidu.com, academic.microsoft.com</td>
</tr>
<tr>
<td>C E-journal database</td>
<td>cnki.net, jstor.org, link.springer.com</td>
</tr>
<tr>
<td>D Electronic library</td>
<td>sslibrary.com, brill.com, books.google.com</td>
</tr>
<tr>
<td>E Document sharing platform</td>
<td>wenku.baidu.com, doc88.com</td>
</tr>
<tr>
<td>F Online encyclopedia</td>
<td>wiki.tw.lvukuji.com, baike.baidu.com</td>
</tr>
<tr>
<td>G Online community and social Q&amp;A platform</td>
<td>researchgate.net, bbs.pinggu.org, zhihu.com, zhidao.baidu.com</td>
</tr>
<tr>
<td>H Blog platform</td>
<td>blog.csdn.net, blog.sciencenet.cn</td>
</tr>
<tr>
<td>I Other websites</td>
<td>Such as news websites and government websites</td>
</tr>
</tbody>
</table>

4.1.4.2 Attitude Measurements

This study adopted and modified some attitude measurements (Bhattacherjee, 2001; Lin & Wang, 2012) to investigate novice searchers’ perceived usefulness, satisfaction, and acceptance attitudes toward the search trail recommendations provided in this experiment. The measurement items are listed in Table 3. Each evaluation dimension score is the average of all measurement items, where 1 was the lowest score and 7 was the highest score.

4.1.5 Procedure

Prior to participating in the experiment, the participants were given a description of the experiment and completed the entry questionnaire. After reading the search task assignment, they were asked to read the “think-aloud guideline” carefully. In addition, participants labeled as novice searchers received a list of recommended search trails generated by participants labeled as expert searchers performing similar tasks, and they were asked to use these recommendations as much as possible. After finishing the search, all participants were asked to complete a post-search questionnaire. Besides, we conducted an exit interview with each participant.
Table 4
Source Categories Transition Probability Matrix of Class Presentation Task

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>Σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>7%</td>
<td></td>
<td>7%</td>
<td>29%</td>
<td>21%</td>
<td>7%</td>
<td>29%</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>50%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>33%</td>
<td></td>
<td></td>
<td></td>
<td>33%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>50%</td>
<td>50%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>50%</td>
<td></td>
<td></td>
<td></td>
<td>25%</td>
<td>25%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>17%</td>
<td>17%</td>
<td></td>
<td></td>
<td></td>
<td>17%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>42%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>58%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 5
Source Categories Transition Probability Matrix of a Course Thesis Task

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>Σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>13%</td>
<td></td>
<td>13%</td>
<td></td>
<td>13%</td>
<td></td>
<td>63%</td>
<td></td>
<td></td>
<td>100%</td>
</tr>
<tr>
<td>B</td>
<td>25%</td>
<td>50%</td>
<td>25%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>13%</td>
<td>13%</td>
<td>50%</td>
<td>13%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>20%</td>
<td>20%</td>
<td></td>
<td></td>
<td></td>
<td>20%</td>
<td></td>
<td></td>
<td></td>
<td>40%</td>
</tr>
<tr>
<td>F</td>
<td>100%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>100%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td></td>
<td>50%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>8%</td>
<td>8%</td>
<td>17%</td>
<td>8%</td>
<td>8%</td>
<td>8%</td>
<td>42%</td>
<td></td>
<td></td>
<td>100%</td>
</tr>
</tbody>
</table>

4.2 User Experiment Results

4.2.1 Expert’s Search Trails Generation

According to the search-task relevant search trails generation method proposed in Section 3.1, the expert search trail case base for this experiment was generated and represented as follows:

\[ T_{mc_1} = \{\text{tmc}_1, \text{tmc}_2\} \]

\[ \text{tmc}_1 = (T_{mc_1}, \text{Cat}_{mc_1}, S_{mc_1}, \text{SC}_{mc_1}, A_{mc_1}) \]

\[ \text{tmc}_2 = (T_{mc_2}, \text{Cat}_{mc_2}, S_{mc_2}, \text{SC}_{mc_2}, A_{mc_2}) \]

\[ T_{mc} = (\text{Class presentation, Low task complexity, Knowledge about "cloud computing"}) \]

\[ T_{mc_2} = (\text{Course thesis, High task complexity, Artificial intelligence in education or healthcare}) \]

\[ \text{Cat}_{mc_1} = \text{Cat}_{mc_2} = \{A, B, C, D, E, F, G, H, I\}, \] the categories of Internet information sources classified in Section 4.1.4.

\[ S_{mc_1} \] and \[ SC_{mc_1} \], the full search trails of the low complexity search task, were generated after expert searchers executed the task. For example, the Internet information sources accessed by expert [5] were baidu.com, zhihu.com, xueshu.baidu.com, wanfangdata.com.cn, baidu.com, wiki.tw.lvfukeji.com, baidu.com, zhihu.com, cloud.idcquan.com, cloud.it168.com. Then, the Internet sources sequence \[ \text{SrcCatSeq}_{\text{expert}[5]} \] was stored in the \[ S_{mc_1} \] set, and the corresponding \[ \text{SrcCatSeq}_{\text{expert}[5]} = \{F, G, B, C, A, F, I, I\} \] was stored in the \[ SC_{mc_1} \] set. The transition probability matrix of the search trail Markov
model of the class presentation task was estimated, as shown in Table 4.

Similarly, $S_{\text{inc}_2}$ and $SC_{\text{inc}_2}$, the full search trails of the high complexity search task, were generated after expert searchers performed the task. For example, the Internet information sources accessed by expert [1] were baidu.com, tech.163.com, 360.doc.com, gov.cn, con.com.cn, ex.cssn.cn, useit.com.cn, xueshu.baidu.com, cnki.net, xueshu.baidu.com, baidu.com, blog.csdn.net, useit.com.cn, app.webofknowledge.com. Then, the Internet sources sequence $SrcCatSeq_{\exp [1]}$ was stored in the $S_{\text{inc}_2}$ set, and the corresponding $SrcCatSeq_{\exp [1]} = (A,I,E,I,I,E,B,C,B,A,H,E,C)$ was stored in the $SC_{\text{inc}_2}$ set. The transition probability matrix of the search trail Markov model of a course thesis task was estimated, as shown in Table 5.

Since the two types of search tasks in this study are highly differentiated, it is easy to get the expert search trails recommendation set of these search tasks as follows:

ET(Class presentation, Low task complexity, Knowledge about “cloud computing”) = $\{tmc_1\}$

ET(Course thesis, High task complexity, Artificial intelligence in education) = $\{tmc_2\}$

ET(Course thesis, High task complexity, Artificial intelligence in healthcare) = $\{tmc_3\}$
4.2.2 Novice Searchers’ Evaluation of the Search Trail Recommendation

From Table 6, it is evident that the average scores of novice searchers’ perceptions (i.e., perceived usefulness (PU), satisfaction (SA), and acceptance (AC)) on the search trail recommendations are all higher than 4. This result shows that novice searchers could benefit from search trail recommendations and are willing to accept this search recommendation during the search process.

The extremely large score values for each measurement item indicate that different novice searchers may have different perceptions of the search trail recommendation. Therefore, we used the k-means clustering algorithm to cluster the scores (i.e., perceived usefulness, satisfaction, and acceptance). Here, the number of categories was set to 3. The clustering results are shown in Table 7.

According to the ratings, clusters 1, 2, and 3 are classified as high, medium, and low rating groups. Among the novice searchers, there were two participants in the high rating group, four participants in the medium rating group, and one participant in the low rating group. Then, we performed an in-depth analysis of the different rating groups. In this analysis, we considered the participants’ background information, search behavior data, post-experiment questionnaire results, and data generated from interviews.

4.2.2.1 High-Rating Group

Novice [3] and novice [7] assigned a score greater than 5 to each evaluation item, indicating that they may have benefited significantly from the search trail recommendation during the search process. These novice participants said that the search trail recommendation was useful, enriched the diversity of information sources, and potentially, improved their search skills. They indicated that, before this experiment, they only used Baidu, the largest Chinese search engine, to search for information. They did not know that they could search for information in databases or professional forums.

The background survey showed that novice [3] only searched for academic information online once a week. In this experiment, novice [3] demonstrated that, during the search process, she had difficulty selecting information sources and forming queries.

Novice [7] stated that she only searched for academic information once a month, and she reported that her search skills were poor. She has difficulty expressing her information needs clearly. Novice [7] explained that “After reading the task requirements, I don’t know how to make a search strategy and don’t know which information source to search from at first. Fortunately, with the help of the search trail recommendation, I learned about many types of online information sources. In this experiment, the diversity of information sources and content in the search results is good, with greatly improved compared to before. I felt that my search skills had also improved.”

4.2.2.2 Medium-Rating Group

The perception of usefulness, satisfaction, and acceptance scores for the search trail recommendation given by novice [2], novice [4], novice [5], and novice [6] were all higher than 4. These novices indicated that the search trail recommendation helped them to increase the diversity of their searches.

The background survey showed that these students only searched for academic information online once every 3 days or once a month. Novice [5] felt that her information search skills were poor and that she had difficulty expressing her information needs clearly. However, she did not want to use a complex search strategy or change her search habits. Novice [4] and novice [6] also indicated low confidence in their information search abilities. However, novice [4] expressed willingness to use a complex search approach, and novice [6] expressed willingness to change their search strategy to suit different search tasks. Compared with other users, search results of novice [5] included relatively few information sources and low diversity of content. Participants in this cluster indicated that they usually obtained information from the Baidu search engine, Online encyclopedia, and E-journal databases. They also acknowledged that they had difficulty choosing information sources during the search process. For example, novice [2] said, “I don’t know whether to use a search engine or go to a database;” novice [4] said, “I feel that the content I find on different pages is the same;” and novice [6] said, “I don’t know which information source can find more information.” These problems infer that these participants have no idea how to get diverse information.

4.2.2.3 Low-Rating Group

Novice [1] assigned a low score to perceptions of usefulness and satisfaction for both the full trail and the one-step trail recommendations for search tasks with high or low complexity. He assigned a score of 4 to the perception of acceptance, which indicated that he did not reject such recommendations. In the exit interview, novice [1] stated that the one-step trail recommendation could prompt
him to obtain information from other information sources after obtaining information from some Internet sources. However, the recommendation of five full trails with different sequence patterns confused him and he did not know which one to choose.

The background survey showed that novice [1] searched for academic information online every day. He believes that he can clearly express his information needs and find the information easily. Novice [1] used various Internet sources in both search tasks, and his search results showed diversity. The difficulties he encountered in the search process were primarily related to queries, for example, “When searching for information about unfamiliar professional words, I could not find relevant pages in SERP by directly pasting its abbreviation in the search engine.”

### 5 Discussion and Conclusion

In this paper, we have proposed a method to learn the behavior of expert searchers to support novice searchers engaged in the same or similar complex tasks. Developing the proposed method involves identifying the sequence of internet information sources explored by expert searchers using Markov chains and organizing these search experiences using a search-task relevant case base. We employ the CBR approach to generate search-task relevant search trail recommendations for novice searchers to assist them in identifying necessary steps or information sources to achieve task completion. We demonstrate through a user study that our task-relevant search trail recommendations can help improve novice searchers’ search performance. The novice searchers who participated in the user study were satisfied with the search trail recommendations and were willing to use them in the future.

#### 5.1 Discussion

The user study results showed that novice searchers have positive attitudes toward the search trail recommendation. Most novice searchers stated that the recommendation helped improve the diversity of information sources and the integrity of the search results’ information content. They also indicated that they were satisfied with the search trail recommendation provided during the search process. All novice searchers were willing to accept the search trail recommendation.

An in-depth analysis of novice searchers’ evaluations of the search trail recommendation found that searchers with different evaluation levels have different information search experience and habits (Table 8), encountered different problems in the search process, and have different task performance results (Table 9).

#### Participants with a high rating for search trail recommendations stated that the recommendation could help them to develop search strategies and select the category and quantity of information sources. Typically, they search online for academic information infrequently and are not confident in their search skills. During the information search process, they encountered difficulty understanding task requirements, developing an information search strategy, selecting information sources, and formulating queries. Participants in the medium rating group stated that the search trail recommendations could complement their source categories and increase the number of relevant sources during the search process.

### Table 8

**Information Search Experience and Habits of Different Rating Clusters**

<table>
<thead>
<tr>
<th>Rating</th>
<th>No.</th>
<th>Information search experience and habits</th>
<th>Search frequency</th>
<th>Complex search intention</th>
<th>Search confidence</th>
<th>Change search strategy intention</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>3</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>2</td>
<td>Medium</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Low</td>
<td>High</td>
<td>Medium</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>1</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
<td></td>
</tr>
</tbody>
</table>

- **Table 8**: Information Search Experience and Habits of Different Rating Clusters
They searched for online academic information every few days or once a month, with low search confidence. For this group, the difficulties encountered in the search process were primarily associated with information source selection and evaluation of the retrieved result. One novice searcher gave the search trail recommendation a low evaluation, which indicates that this recommendation did not significantly improve his information source diversity. This novice searcher searched for online information frequently, had a high degree of search confidence, and was willing to try a complex search. The only difficulties he encountered during the search process related to queries, and his search results were sufficient.

This experiment found that, related to the search trail recommendation, novice searchers who were willing to conduct complex search activities or to change search habits obtained information search results that demonstrated good information source diversity and content integrity. However, even though novice [5] was provided with search trail recommendations, her search results were relatively poor. This may have occurred because she has poor search skills and is unwilling to try complex searches or change her search habits.

Some novice searchers suggested that it would be more beneficial to include expert searchers’ evaluation of each Internet source category in the search trail. This would allow them to know what information sources the expert searchers used and why the experts used these sources. Some novice searchers expect search engines to automatically separate aspects of an information search task. Besides, search engine result pages could be layered to display corresponding information source results, which would greatly reduce the searcher’s workload and improve the effectiveness of search results.

5.2 Implications

The research presented in this paper could extend previous studies in several ways. First, the search-task relevant search trail recommendations proposed in this study expanded the scope of traditional trail recommendations, which primarily focus on query-relevant trail origins, sub-trails, and destination recommendations (White & Huang, 2010). Our findings suggest that task-level search trail recommendations will provide useful guidance to novice searchers. Second, we proposed a method for search trail generation using Markov chains and evaluated the method experimentally. The results suggested that Markov chains were useful tools for Internet information source sequence modeling and search trail analysis. Third, the most innovative feature of this study is that it recommended search trails based on CBR searching. Employing CBR searching represents a new research idea that may benefit research on Internet information recommendations for complex or exploratory search tasks.

The results of this study are also important for search engines and search assistance service designers. Our findings suggested that search trail recommendations incorporated into SERPs and certain webpage can help novice searchers. Most novice searchers reported good experiences interacting with the trails provided, and some even indicated that the search trail recommendations potentially improved their search skills. The information search trail recommendation method proposed in this study can provide a reference to improve network information organization and management, information recommendation services, and improve the utilization efficiency of professional or domain Internet sources.
5.3 Limitations and Future Work

Despite the theoretical and practical value of this research, we should acknowledge some limitations. First, the Markov chain model currently used to identify search trails is limited in terms of the amount of training data required and dimensionality with Internet sources categories classified in this study. In the future, more work needs to be done to extend the method to all Internet information sources.

Second, the use of a CBR approach in the search trail recommendation may also have adverse effects. The knowledge structures activated in the searcher’s mind by the provided case may hinder their access to other areas of the solution space (Althuizen & Wierenga, 2014). In addition, this study covered a limited number of task features. Considering the various features that might exist in the actual CBR process, we will enrich features more comprehensively in the next study.

We have shown the important promise of our approach in supporting some important dimensions of search performance (i.e., develop a search strategy, select the type, and number of sources). However, we also need to evaluate the effectiveness of these recommendations with different evaluation criteria, such as relevance, topic coverage, and topic diversity of the search trail recommendations.

What we can infer from the search interaction log data is limited; however, our approach has provided insight on how to organize expert searchers’ trails experience and how to generate and represent search-task relevant search trail recommendations on the SERPs and certain webpages. To provide more accurate and efficient information search trail recommendations, in the future, we intend to conduct large-scale user studies that consider and evaluate task and searcher characteristics, as well as Internet sources and queries.

Acknowledgments: This work is supported by the National Natural Science Foundation of China under Grant No. 71874130 and the Ministry of Education of the People’s Republic of China under Grant No. 18YJC870026.

References


Abstract: This paper presents the findings of a study exploring the information practices of members of a religious organization. Its focus is the “Mahamevnawa Buddhist Monastery.” Particularly, this paper focuses on the study’s findings in relation to participants’ information practices in constructing their understanding of “the Temple.” The study is informed by an information practices theoretical perspective, drawing on work from a variety of disciplines, including Castells’ space of flows, and Fisher’s information grounds. Data was gathered from participant observation, interviews with both monks and devotees and email follow-ups, and analysis of the online presence of the temple through its website. Five social constructs for the temple appear frequently in the interviews: Virtual space; Physical/geographical place; Virtual space; Symbol; Process and practices; and Organization. Participants’ information practices are not only limited to spiritual purposes but also are linked to various social practices, activities, and interests. The study’s findings suggest that constructions of place play a hitherto underexplored role in the multi-layered relationship between people and information.

Keywords: information practices, information flow, networked society, practice theory, information behavior, Castells

1 Introduction

This paper presents the findings of a study exploring the information practices of members of a religious organization. Its focus is the Mahamevnawa Buddhist Monastery (generally referred to by participants in the study as the Temple), established in Sri Lanka in 1999 which is now operating with branches in many countries, including Australia.

In adopting an information practices perspective (Savolainen, 2007), the study seeks to move beyond the implicit limitations that are prevailing in information behavior approaches to focus on the social, collective, non-purposive aspects of the ways people deal with the information and their interactions, deviating from the conceptual tradition of information behavior research, with its focus on the active information seeking behavior of individuals.

The social relationships observed in the Temple are varied. Gherardi (2008, p. 517) pointed out the importance of knowledge in sustaining complex social relationships: To know is to be capable of participating with the requisite knowledge competence in the complex web of relationship among people material artifacts and activities. Few studies have attempted to untangle this complex web of relationships by any means, but our study has attempted to do it using an information practices approach. Further, the conceptual starting point of Castells’s (2004) notion of information flows has provided a path for directing focus towards information practices in a setting comprising of multi-layered social relationships, by bringing together the people, their communications, and the means of communication as well as the outcomes of interactions. Sassen’s (2006) concept of globalized organization shed light on the institutional perspectives of the Mahamevnawa Temple and its workings. Fisher’s (2005) theory of Information Grounds provided a basis for interpreting face to face interactions in specific locations.
Up to this time, these conceptual frameworks have not been linked in terms of information practice research. Analyzing them in depth can bring out their similarities, differences, and overlapping ideas, which will shed the light on Gherardi’s complex web of relationships.

The study has shown that participants are attracted to the temple and connected to the temple for various reasons. The information practices that emerge out of these reasons are not only limited to dharma—that is, spiritual purposes—but also are linked to various social practices, activities, and interests such as food, music, celebrations, culture, careers, business connections, and friendships. These practices emerge from a range of events such as blood donations, working bees, and children’s programs as well as from prayers and other religious activities, engagement with the voluntary activities of donations, and the commercial activities of producing and buying religious texts and recordings. The experiences of monks and devotees not only link individuals with the physical social group of the Temple, but also demonstrate the existence of information networks, social networks, and media networks of various interests and expertise. The findings of the study demonstrate that the information not only flows between individuals, but also within the various social groups and even between global institutions (branches of the temple and other collaborative institutions) through varied information networks. These practices are part of the everyday lives of participants, who consider what they do in this shared community to be unremarkable (Olsson, 2013) and they describe what they do in a taken-for-granted way. As a consequence, these practices are rendered “invisible” to the participants (Savolainen, 2008, p. 3), which presents a challenge for the researcher who must elaborate them in detail.

These social experiences are considered based on shared activities and to some extent a shared sense of identity has laid a good foundation for investigating the information flows of the temple with its emphasis on social context.

Thus, it is important to emphasize the significance of the setting within which these social interactions take place and the contexts within which information flows. Moreover, Moring and Lloyd (2013) argued that particular social practices are produced through social interactions within this specific setting and that it is possible to identify social conditions that fundamentally shape this setting.

The setting for this study, the MA Temple, is extremely complex. The findings indicate that the Temple exists as a permanent building, but it can be created temporarily only; it is a place for spiritual development, for friendship, and for the maintenance of ethnic culture; it is a global business and a place for quiet contemplation; it is sustained through the skills of the monks and through information and communication technologies; its devotees see themselves as individuals and as members of overlapping networks as do the monks; the interactions of those engaged in this social setting lead not only to spiritual capital as expected from a religious setting, but also to economic capital and informational capital. Further, exercising of power can also be determined within this setting.

The paper specifically focuses on the study’s findings related to participants’ information practices in constructing their understanding of the Temple: the variety of complex and sometimes contradictory shared discursive practices that members of the community, both monks and devotees, used to describe and make sense of in terms of spiritual, social, financial, physical and virtual entities.

1.1 Background to the Temple

Mahamevnawa is both an organization and a network of temples, as well as the name of a Buddhist temple which is presently located on the outskirts of Sydney, Australia. Since the first temple was opened at Polgahawela, Sri Lanka in August 1999, the organization has expanded in all these years to 50 branches in Sri Lanka and worldwide including nine for nuns. Branches of the temple are found in many countries including Canada, France, Germany, India, Italy, the Netherlands, New Zealand, South Korea, the United Kingdom, and the USA as well as Australia. Wherever possible, the temples have been established in the form of substantial buildings, created with the intention of maintaining ties to the extensive heritage of spiritual art in Buddhism. The founder of the temple, the Venerable Kiribathgoda Gnanananda Thero, often referred to as the Guru, was particularly determined to ensure that the Dhamma, the teachings of the Buddha, were available to people not in Pāli, the ancient language of the religion, but in language that people would find easy to understand. Thus, he has involved himself in a large translation project, apart from a publishing program through which he has written more than 80 books based on the teachings of the Buddha and compiled more than 100 books based on his own ideas and preaching. These appeal to younger people particularly. The temple organization makes great use of information and communication technologies, namely CDs and DVDs, websites, a television station, and a YouTube channel among the many approaches
2 Conceptual Framework

The present study should be seen as part of the “information practice” umbrella discourse which was described by Savolainen (2007) as an emerging critical alternative to the prevailing “information behavior” discourse in the first decade of the twenty-first century. Savolainen follows Talja in suggesting that the key characteristic of this new discourse is that it represents “a more sociologically and contextually oriented line of research” which...shifts the focus away from the behavior, action, motives and skills of monological individuals. Instead the main attention is directed to them as members of various groups and communities that constitute the context of their mundane activities. (Savolainen, 2007, p. 120)

The study is informed by an information practices theoretical perspective grounded on the understanding that information practices research:

... requires us to understand how shared, practical understanding is derived from becoming or being embodied in context (in situ). Consequently, to know is to be capable of participating with the requisite competence in the complex web of relationships among people, material artifacts and activities ... information practices are context specific, and are entwined with a range of modalities (social, corporeal and epistemic/instrumental) (Olsson & Lloyd, 2017).

As is common in information practices research, the study’s conceptual framework is informed by work in a variety of disciplines, including Castells’ (2010) “space of flows,” and Fisher’s “information grounds” (2002) frameworks. Theories drawn from these authors have contributed to how the study understands the social, corporeal, and epistemic/instrumental modalities described by Olsson and Lloyd (2017).

Castells’ influential work, The Network Society, published over several years at the end of the 1990s, proposed a reconsideration of the concept of space, leading to the development of the idea of the “space of flows.” His fundamental understanding of the space of flows was stated as “the material organization of time-sharing social practices that work through flows” (1996, p. 204) and although he has modified some of his understandings over time, this has not been changed; the space of flows allows people who are in different physical spaces to interact together at the same time.

The space of flows has three key dimensions to it, according to Castells; the first is the infrastructure that allows for communication; the second is the hubs and nodes where messages cluster and social interactions take place; and the third is the people who have the knowledge, skills, and power to take part in the exchanges in the network. The flows concern all aspects of life that can be exchanged in digital space, with an emphasis on economic and informational flows. Thus, the space of flows, the space of the Informational Society, is complex, because of the existence of a multiplicity of elements in it and the wide range of interactions among and between them.

On the other hand, the space of places is a physical location where experiences take place and this space represents the geographic spaces of everyday life. Stalder (2006) argues that Castells’ original notion that there was a sharp distinction between the space of flows and the space of places is no longer a valid one. It is no longer possible to claim that the space of flows is a place where dominant power in a society is exercised since only elites are present in the space of flows, while physical locales are the places where people with less power gathered. Castells himself (1999) acknowledges that the new society will arise not from “the separation between places and flows, but out of the interface between places and flows and between cultures and social interests, both in the space of flows and in the space of places.” Castells’ work provides one conceptual lens through which the study has explored the Mahamevnawa Buddhist Monastery: an organization and community that exists both in physical space (in Sydney, in Sri Lanka, and in regional temples around the world) but whose success is also strongly rooted in its extensive virtual presence and the online community of devotees it has developed.

Information grounds are clearly physical places, but they are more than the geographic spaces of everyday life that mark the space of places. An information ground emerges, temporarily, when people come together for a common reason and as a result create social interactions, thus leading to the sharing of information. This sharing of information is supplementary to the main purpose of being in the requisite place (Fisher, Landry, & Naumer, 2006). Conceptually, information grounds are extremely complex and the factors that comprise them are grouped into three categories: people related, place related, and information related. More importantly, the feelings experienced by different people participating in the same information ground at the same time are different.
3 Methodology

Approval for the study was granted by the Human Research Ethics Committee of the University. As one of the authors is a member of the temple community, participant recruitment was carried out at arm’s length by exhibiting the required notices in meeting places and posting the notices online, inviting participation in the study. Participants in the study are referred throughout as monks or devotees, the terms that they use to refer to themselves and each other in the context of the temple.

The focus of this research is to bring out the social reality. Ethnography is a good theoretical fit for this research for understanding various social worlds from the perspectives of the community around the temple and its branches. It will provide a micro level perspective on the effects of global transformation processes of the MA temple as well as systematic insights into perceptions and strategies of local actors and their information flows. Being a member of the temple community, the interviewer was able to build better rapport with the interview participants which resulted in a better insight into the temple. This helped to look at the scenarios outside the academic boundaries and uncover hidden agendas, ideas, motives, and relationships that the temple participants have with the temple and analyze them from their own experiences. This leads to an understanding of the real meaning of those motives.

Furthermore, the research was undertaken by interviewing participants only in Australia. However, the participants are members of a diasporic community and therefore have multiple, intersecting cultural identities. The temple being a globalized organization, the monks also have served and traveled to branches all over the world.

There were 8 monks and 13 devotees among the participants and each was interviewed for approximately one hour. In addition to interviews, participant observation gave insight into the different social practices existing within the temple.

Data have been gathered through participant observation, interviews with monks and devotees, email follow-ups, and analysis of the online presence of the temple through its website. Questions for the monks focused on their work in the globalized context of the temple, their contribution to the temple as an organization, and their engagement with other monks and with devotees in the sharing of information and knowledge. Similarly, questions for devotees are designed in such a way that they explore their engagement with the temple and their contribution to it.

The questions were open-ended ones. The interview began with the primary question: “Tell me the story about how you came to this temple?”

We have secondary questions based on their answers such as:

- From Devotees
  How did you hear about the temple?
  What does temple mean to you?
  What is your contribution to the temple, and your relationship with temple community?

- From Monks,
  What factors have encouraged you to be a monk on this temple?
  Your contribution as a monk serving in the temple
  How do you work with Monks and other participants?
  In what ways do you work with devotees and what benefits do they get from the temple?

The information regarding the conduct of interviews was advertised on the notice board of the temple. They were conducted in various places such as temple premises, temporary gatherings, the places where their events may take place temporarily, routinely, or permanently.

Interviews were conducted predominantly in English, but as the researcher conducting the interviews and the devotees had Sinhalese as their common language, there were times, especially during the discussion of spiritual practices, when Sinhalese words were used rather than its customary English translation. The interviews were audio-recorded and then transcribed before analysis. During transcription, obvious errors in grammar have been corrected. The analysis of the website of the main temple organization and of the Sydney organization used the English-language version and focused on the information-based activities of the temple organization, including the publishing program, the TV station, and YouTube channel, as well as the use of technologies such as Soundcloud to communicate with devotees.

Preliminary analysis of the data was performed using both a priori coding drawn from Castells’ and Fisher’s work and emergent coding (Bryman, 2008). Emergent coding helped the coding framework to develop significantly during the analysis process. The analysis has revealed the complexity of the organization and the relationships of both monks and devotees to the organization and to each other, in relation to the importance of information technologies and physical places at the heart of these relationships.
4 Findings

Five social constructs for the temple appear frequently during interviews with monks and devotees:

- **Virtual space** — many participants described online virtual presence and active online communities as one of the defining features of the Temple setting, how it was distinct from more traditional Buddhist institutions and how effective this was in affording its geographically dispersed diaspora community a sense of belonging.

- **Physical/geographical place** — participants would refer to the Temple when talking about it as a physical place, for example when referring to the Sydney temple or the foundation temple in Polgahawela, Sri Lanka.

- **Symbol** — participants would also use the Temple in a symbolic way—a way of referring to the community as a whole and also a body of people with a shared set of religious beliefs.

- **Process and practices** — participants would refer to the Temple as a place where religious ceremonies are performed, a place where devotees can engage in and develop their understanding of spiritual practices such as meditation.

- **Organization** — many participants also considered the Temple as an organization in the context of managerial insight; for example, it is understood to be a financial entity with expenses, cash flow, and maintenance costs. On this level, in contrast to its more spiritual constructions, certain contexts would require the Temple to be perceived as an organizational entity, little different from a corporation or a university.

It is clear that the work of the Temple and the interactions among and between the monks and the devotees would not be possible without the internet. The temple websites indicate significant use of information and communication technologies (ICTs) in carrying out the work of the temple and in communicating with monks and devotees. Use of the mobile phone to contact the monks is suggested on the website and both monks and devotees made comments on this means of communication. Technologies used in spreading the message of Buddha include the sale of digital and printed books, online webcasts, and sermons stored in Soundcloud, YouTube, and Google+, and the use of Blogspot, Facebook, Skype, and Twitter for spreading messages about temple programs as well as in the solicitation of donations.

The temple has a strong focus on money-related matters to certain extent. A number of devotees commented on making contributions to the Temple, including through the website: *I even contributed to the temple through the internet because they have their account numbers displayed online*. Others commented on the range of books and other merchandise available for sale:

They sell Buddhist books in front of the place [when] they conduct the Sill program. So I think those funds go to improvement of fundamental facilities of the temple. So I bought plenty of books which I think would directly contribute and go to the improvement of the facilities.

These activities are in addition to the television channel and the extensive use of YouTube, as mentioned above. Devotees welcome these technologies and use them alongside real world interactions as they engage in their spiritual activities:

First we went to Live Sessions in the monastery conducted by the Ven Kiribathgoda Gnanananda Thero, our Teacher. They have books and also they have a library so we joined the library and we bought some books and CDs; subsequently, we heard he was doing his preaching on the TV, Radio most of the time in Live sessions and [recording on]Audio Cassettes, MP3s.

I got to know those programs through emails and through my friends like the community and some through leaflets.

Devotees indicated that they use a range of technologies to communicate with each other and with the monks, such as mobile phones, Skype, email, and Viber, and in the process, build friendships with other devotees. Similarly, as expected, the monks also use these technologies to communicate with monks in other locations and to create links in new communities:

We did preaching through Skype to Korea and Germany. In Germany, kids’ programs are done through Skype. We hold discussions too over the phone.

The use of technology that monks described was not focused on social interaction as it was used for the devotees, but at one level was used to resolve conflicts and find answers. At an organizational level, it was used to promote the work of the temple, support decentralized decision-making, and sustain the governance processes of the temple, as it is seen as a globalized organization. *Monks are able to attend meetings through online Technology and thereby support the sales and distribution of the publications produced by the Guru*.

At the same time, place is extremely important both to the devotees and to the monks. The ways of discussing place is much greater than the ways in which the use of technology and the flows of information are discussed. It is clear that participants view the temple through different lenses because they meet in different settings,
which depend on the reasons and purposes for which such meetings are conducted. Even when they meet in the same physical location and for the same program or activity, their reasons will differ and this creates a different construct.

When monks refer to the temple, they tend to use the formal name of Mahamevnawa, which is now the formal name of the building on the outskirts of Sydney; now they use it to represent the temple as an organization, rather than the physical building. Although one devotee specifically talks in organizational terms: I’ve been to the head office when I was in Sri Lanka, devotees tend to have a much more physical sense of place:

I go to the Dharma program in Baulkham Hills [a suburb in north-western Sydney] which will be on the first Saturday of each month.

They had these functions regularly in different places close to where I lived. Some of them were in [northern suburbs of Sydney] Cherrybrook, Pennant Hills, Castle Hill, Seven Hills, Blacktown.

The temple is also conceptualized as a symbol by most devotees. On one hand, it is a representation of beliefs: Temple is a place that represents Buddha, Dharma, and Sangha; and a place where I can gather merits. On the other hand, it represents a sense of peace and tranquility: and it is a place that promotes peace of mind. A female devotee said: as the name implies, it was an Asapuwa to me.

This is an interesting and complex reference that harks back to traditional notions of a hermitage, a place independent from the everyday world, for example, as a forest retreat and which at the same time carries with it a vastly contemporary allusion to luxurious looking temples which are wholly different from the traditional ones. This idea of luxury is apparent in other responses from devotees, for example: if they are building a bathroom, I mean a bath room with cement for the floor; you can put beautiful tiles on the floor.

It is a place which helps to build trust among friends. It is also a place for action, whether religious or community based, as is evident in the following statement: Today we had a working bee program. We enjoyed it a lot. It is very nice to do such things at the temple and other places we gather.

Finally, it is a place where devotees acknowledge that information on many topics is shared. Many devotees and the monks place an emphasis on the exchange of religious ideas, and the understanding of dhamma and its practices: If [devotees] practice Buddhism, they can experience those stages of mind. Therefore, the Devotion is essential. However, most devotees acknowledge that place functions as an information ground:

So we get to know about other things and information through them, which is not only related to Buddhism but also something that could be helpful for your career. The communication is very broad in that place.

This view of the place and the processes happening there is different from the view held by at least one of the monks: we do not focus or regard it as a meeting and greeting place. In spite of this view from one of the monks, the devotees describe a number of information grounds, including working bees, community events such as blood donation, cultural events and celebrations, and children’s programs, where they use the time to socialize and to share information.

The study’s findings show that even a physical place is more than simply a geographic location and its meaning and significance to members of the temple community can shift depending on the context of its use. A close analysis of the way the devotees and monks talk about the temple as a place indicates that another conceptualization of place exists for them, which is related to the permanence of the relationship between the devotees and the place where spiritual practices are held. Within this conceptualization, there are three categories namely temporary, routine, and permanent.

Temporary places are places where people meet for different events from time to time. It could be where they have social activities such as blood donations, working bee programs, and youth programs; often, however, these places are used for Dhamma preaching:

A few devotees got together and invited the monk and we first had a meeting at a house in Granville. That was where my temple was. One of my friends, who lives in Canberra has organized a one-day program and invited all his friends and neighbors to his place.

The place where monks lived was a bit too far and I could not visit the place, but they had everything in different places and managed to create the same temple atmosphere such as in community centers, clubs, and even individual devotee’s houses. They used to bring huge statues and everything from the temple to these places all over the NSW and gave a marvelous, unthinkable, and unforgettable experience to the people.

Routine places are places that exist in the community and are used for other purposes most of the time. However, at a given frequency, say once a month, that geographic space becomes a part of the temple. One of the monks explained how spiritual practices are developed and established in these routine places:
In countries like this, we do it via the people. We select one person to initiate and organize Dharma programs close to their residence. He/she knows the area well and they also know the people who live close by. So, we carry out programs at the selected areas every month.

Permanent places are places that are part of the temple and utilized in the temple: I visit the monastery on Wednesday and Friday as there is a Dharma Sermon after the chanting in the evening. I also visit the temple whenever I’m free.

The study has given rise to the observation that participants are attracted to the temple and connected to it for a variety of reasons. The information practices that emerge from these reasons are not only limited to dharma — namely spiritual purposes — but are also linked to various social practices, activities, and interests such as food, music, celebrations, culture, careers, business connections, and friendships. These practices emerge from a range of events such as blood donations, working bees, and children’s programs, as well as through prayers and other religious activities and engagement in the voluntary activities of donations and the commercial activities of producing and purchasing religious texts and recordings. These activities of monks and devotees not only link individuals with the physical social group of the temple, but also demonstrate the existence of information networks, social networks, and media networks consisting of various interests and aspects of expertise. The findings of the study demonstrate that information not only flows between individuals but also within various social groups having many interests and even between global institutions (branches of the temple and other collaborative institutions) through varied information networks.

These practices are part of the everyday lives of participants, who take what they do in this shared community for granted (Olsson, 2013), describing what they do in a taken-for-granted way. As a consequence, these practices are rendered “invisible” (Savolainen, 2008, p. 3) and this presents a challenge for the researcher who must elaborate them in detail. Extracting information regarding these social experiences based on shared activities, and a shared sense of identity, have to some extent laid a good foundation for investigating the information flows of the temple with its emphasis on social context.

Thus, it is important to emphasize the significance of the setting within which these social interactions take place and the contexts within which information flows. As Moring and Lloyd (2013) argue, particular social practices are produced through social interactions within this specific setting and it is possible to identify social conditions that fundamentally shape this setting. The setting for this study, the MA Temple, is multi-layered. The findings indicate that it exists as a permanent building and it can be created temporarily; it is a place for spiritual development, friendship, and the maintenance of ethnic culture; it is a global business and a place for quiet contemplation; it is sustained through the skills of the monks and through information and communication technologies; its devotees see themselves as individuals and as members of overlapping networks as do the monks; the interactions of those engaged in this social setting lead not only to spiritual capital as expected from a religious setting, but also to economic capital and informational capital. Within this setting, exercises of power can also be determined.

Gherardi pointed out the importance of knowledge in sustaining complex social relationships:

> To know is to be capable of participating with the requisite knowledge competence in the complex web of relationship among people material artifacts and activities (Gherardi 2008, p. 517)

Few studies have attempted to untangle this complex web of relationships, but our study is attempting to do so using an information practices perspective. The conceptual starting point of Castells’ notion of information flows has provided a way to focus on the information practices in this setting comprising of multi-layered social relationships, by bringing together the people, their communications, and the means of communication as well as the outcomes of interactions. Fisher’s notion of Information Grounds provided a basis for interpreting face to face interactions in fixed locations. Till now, these conceptual frameworks have not been linked in terms of information practices research. Analyzing them in depth can bring to light their similarities, differences, and overlapping ideas to impart clarity on Gherardi’s complex web of relationships.

First, Fisher’s (2005) research was not only limited to physical places but also incorporated online settings of information grounds such as social media settings. Chatman’s “small world”(1991) and Oldenburg’s (1999) “the Great Good Place” have also explored the idea of information grounds in different contexts, although without using the phrase. In terms of the temple, casual interactions and conversations take place in different social spaces of the temple e.g., the Library, Meditation halls, outdoor activities, and so on.

On the other hand, Castells investigated space and place in his work “Network Society.” He states that the place is unimportant, and “networked society” has no boundaries and thus can be global (p.2737), marked by “timeless time “and “space of flows.” The concept
of “space of flows” has molded the social settings interconnecting different places through different people and communities locally and globally. According to him, the “networked organization” should have, open[ed] up unlimited horizons of creativity and communication inviting us to the exploration of new domains of experience (p1) and they all are interconnected with physically disjointed positions held by social actors (p.412) in terms of information practices. Davies and McKenzie (2004) also argued that information practices are based on social interactions, activities in multiple locations and times, and Nicolini (2009, p. 213) stated that relationships [enacted] in space and time are important. However, Castells space of flows theory mostly is related to the commercial aspect of information flows. It is evident from our study that some of these online practices are not only informational, but also economic, cultural, and even spiritual.

The temple can be seen as exemplifying Castells’ concept of the space of flows: a virtual network whose nodes connect with one another online. The temple uses a broad range of technologies resulting in high-end production values to communicate the message of the Buddha and the special approach of this temple lies in focusing on modernity and aesthetic appeal. Many devotees first become part of the temple community online and many participants describe the temple’s virtual presence as one of the most important ways through which they engage with it. They enthusiastically use communication technologies to communicate with the monks and with each other, and value the YouTube channel and recordings of the sermons as ways to continue their spiritual development when attending a physical practice is not possible.

In addition, the temple has made sophisticated use of information technology to develop a decentralized structure where many functions are devolved to its various international branches. Monks describe the “meetings” in which they interact with monks in other temples around the world and with the main temple in Sri Lanka. They also speak of the ways in which they and their fellow monks have used the technology, especially Skype, to extend their spiritual programs into places where otherwise there would be no program.

At the same time, the temple is a specific, physical place where monks and devotees come together and where information is exchanged, both formally and informally. It acts as an information ground since this information exchange is not exclusively confined to religious practices, since the participants describe the temple as an important source of information relating to social and business opportunities. It is clear that the devotees carry out a number of social roles when they engage in the information grounds, as Fisher, Landry, and Naumer (2006) indicated and expected. For example, the devotees perform many social roles beyond their devotional roles, including functioning as a donor (of blood and money), consumer, learner, volunteer, logistics coordinator, and parent. The devotees clearly value their engagement with the temple and the various opportunities for activities created through the information grounds. However, it would appear that some of the monks are less enthusiastic about the social practices that have evolved around some of the community-based programs. This might be of evidence that the monks are seen as members of elite found in the space of flows as argued by Castells (1996); it is the transactions prompted through the work of the temple as an organization and the monks as individuals that are important.

Castells, in his later work (2004), draws attention to the importance of considering the interface between flows and places, acknowledging that it has become very easy for people to move between place and the space of flows because of easy universal access to the internet. Here, we draw attention to the separation of place into three categories by the devotees: temporary, routine, and permanent, with a particular focus on temporary places. Castells argues for the existence of timeless time in the space of flows, where the structure of events is disrupted. In the context of the temple, for example, the posting of the sermons in the Soundcloud platform means that they can be listened to at any time or whenever needed and the opportunity to donate online means that money can flow to the temple at any time. The idea of a temporary place is also disruptive of time, as it exists today, but not tomorrow. At the same time, the use of statues and decorations convey a sense of permanence by bringing in a sense of familiarity in the context of the temple, perhaps representing in a way an infrastructure that is needed for the temple; in the same way the space of flows requires an infrastructure. Temporary uses of urban space can be experimental and innovative; they can subvert the power of authority, especially planning authority, and now people in a locality assert their power using their needs, culture, and aesthetics to create the atmosphere that successfully reflects their wishes (Henneberry, 2017, p. 6-8). When considering the case of the temple, which often creates temporary temple spaces in devotees’ homes or community halls, they are clearly innovative, and often implemented by an individual, although equally supported by the organization of the temple. Yet, as Henneberry has pointed out, it is the users of the spaces and the atmosphere they create which render a temporary
space authentic; further, without the necessary flows of information, the social practices and the culture of the temple as a spiritual organization, the temporary places of the temple would not be possible. The temporary places created by devotees of the temple are clearly disruptive of the urban environment in which they briefly exist.

Another point of tension exists within the workings of the temple itself. Whilst the purpose of the temple is grounded on spirituality, it has necessarily become involved in the creation and dissemination of both printed and online information products in order to disseminate its spiritual message. Thus, from Castells’ perspective, it is an example of both informational capital and economic capital at the same time. This particular place of the temple within Buddhism is different from others due to the Guru’s emphasis on using modern language and the vernacular of the countries where the temple is located. The books, CDs, online recordings, photos and so on have all become an integral part to the support and development of this particular approach to spread the messages of Buddha. Both the monks and devotees describe the importance of the wealth of information and knowledge contained in the communications of the temple and the important spiritual learning to be gained through engagement with this material. This clearly involves the creation of informational capital. At the same time, devotees, in particular, are aware of the ways in which the sale of books contributes to the economic capital of the temple, thus helping to support the monks and the work they do. Neither monks nor devotees comment or discuss on/about the grand buildings, luxurious appointments, and the ways the economic capital allow their development, although one monk discussed the cost of airline tickets involved in traveling around the branches of the temple or visiting the main temple in Sri Lanka. Castells allows for interaction between these two forms of capital. Qian and Kong (2018, p.161) assert that social changes which are attributable to modernity will bring “the secular logics of market (and) economy” into theological domains. Moreover, Asad (2003 quoted in Qian and Kong (2018)) suggests that to “make a rigid division between the sacred and the secular is surely to impoverish both.” In this brief article, we can do no more than to flag this friction between the spiritual and the economic as an area that requires further elaboration.

Further, a point of contention emerges in the development of the discussion. This arises due to the differences in the cultural background and experiences of the authors. It raises questions about the usage of concepts of one culture in understanding the practices of another culture as the authors acknowledge that there will be different interpretations resulting from the cultural assumptions they make. Castells’ concepts are complex and developed nearly twenty years ago, and applying them in a Western context requires a degree of interpretation based on rigorous analysis because of the significant changes which have emerged in the relationships between flows and places in everyday life. Using these same concepts in an environment that is not Western and not secular makes the process of rigorous analysis and interpretation more challenging. The authors have adopted a reflective practice approach to the analysis process, acknowledging, discussing, and at times challenging the dimension of the differing cultural assumptions and expertise they have brought to the research.

5 Conclusion

The study demonstrates how an information practices perspective can provide a useful conceptual lens that affords information researchers new opportunities to gain insights into aspects of people’s relationship with information that more traditional research discourses have not considered till date. This has allowed researchers to analyze participants’ sense-making of the temple as a complex web of information practices engaging with an array of social, corporeal, and epistemic/instrumental modalities, according to Olsson & Lloyd (2017).

By bringing together an information practices perspective and the work of Manuel Castells, it has identified various anomalies in Castells’ concepts of space of flows and space of places that go beyond the simple disparity existing in the hierarchy of power as discussed in the early elaborations of the concepts involved. The findings of this study suggest that the space of flows has been disruptive to social practices in the same way, and so the concept of “temporary places” is also disruptive to social practices in the urban environment, destabilizing both ideas of time and permitted use of space.

The present study has also indicated that the concept of information grounds can complement the concept of space of places through its capacity in analyzing the people to the context and the information engaged in any information ground. Although people are part of Castells’ concept of space of places, they are largely undifferentiated actors with no clearly specified purpose. The work of Fisher, Landry, and Naumer (2006) provides an analytical frame that identifies the range of social roles which devotees and monks play. Future works will demonstrate how use of this frame can shed light on the
relationships between the spiritual, informational, and economic capital, as well as secular social practices.

Finally, the study has demonstrated the value an information research can bring to the discipline as a whole based on non-Western contexts. Adapting Western theoretical and methodological approaches to studying the MA Temple, the researchers have forced other researchers to question and rethink the dimension of differing cultural assumptions and expertise they brought to the research. It is our hope that other information researchers around the globe will continue to expand the cultural, conceptual, and methodological boundaries of the field.

Acknowledgments: The authors wish to thank the anonymous reviewers for their feedback.

References


Oldenburg, R. (1999). The great good place: Cafés, coffee shops, bookstores, bars, hair salons, and other hangouts at the heart of a community. New York: Marlowe.


