ChinaRxiv

AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-202308.00329

Postprint: A Support Vector Machine-Based Study on Factors Influencing University Students' Serendipitous Online Information Encounter

Authors: Tian Mei, Zhu Xuefang

Date: 2023-08-26T00:00:00+00:00

Abstract

[Purpose/Significance] This study investigates the sensitive influencing factors of information encountering among college student populations in online environments, aiming to guide students in improving their information encountering capabilities and subsequently enhancing their information literacy. [Method/Process] Information gain is employed to analyze the correlation between each influencing factor and the frequency of information encountering occurrences, constructing a sensitive influencing factor model, and further utilizing Support Vector Machine (SVM) to establish a predictive model for information encountering frequency. [Results/Conclusion] The 10 influencing factors most relevant to information encountering are distributed across four dimensions: information user, encountered information, network environment, and contextual factors; the model's classification prediction accuracy reaches 82.96%, indicating that SVM demonstrates good effectiveness in predicting information encountering frequency.

Full Text

Preamble

Vol. 62 No. 8, April 2018

Research on Influencing Factors of University Students' Online Information Encountering Based on Support Vector Machine

Tian Mei[1],[2], Zhu Xuefang[2]
[1] School of Management, Xinxiang Medical University; Research Center of Health Information Resources, Xinxiang Medical University, Xinxiang 453003
[2] School of Information Management, Nanjing University, Nanjing 210023

Abstract

[**Purpose/Significance**] This study investigates the sensitive influencing factors of information encountering among university students in network environments, aiming to guide this population in improving their information encountering capabilities and subsequently enhancing their information literacy. [**Method/Process**] Information gain is employed to analyze the correlation between various influencing factors and the frequency of information encountering, constructing a sensitive influencing factor model. Furthermore, Support Vector Machine (SVM) is utilized to establish a predictive model for information encountering frequency. [**Results/Conclusion**] The ten most relevant influencing factors for information encountering are distributed across four dimensions: information user, encountered information, network environment, and situational factors. The model achieves a classification prediction accuracy of 82.96%, demonstrating that SVM yields favorable results in predicting information encountering frequency.

Keywords: information encountering; information behavior; support vector machine; influencing factors; information gain
Classification Number: G250
DOI: 10.13266/j.issn.0252-3116.2018.08.011

Introduction

During information searching, web browsing, and information interaction, we often serendipitously acquire information that we find "interesting" or "useful." This phenomenon of unintentionally obtaining needed information is known as information encountering. Although information encountering represents an "unexpected" and "unanticipated" gain, it not only expands individual knowledge but also fosters new problem-solving approaches by providing more useful or interesting information. As a passive information acquisition method, information encountering has received increasing attention, with numerous studies demonstrating its important role in individual work, life, learning, and research.

In the current Web 2.0 network environment, the widespread application of mobile internet has made information encountering more likely to occur. University students represent one of the primary populations using mobile internet and various new media applications. For this group, information encountering holds significant meaning for information acquisition, innovative problem-solving, and improving autonomous learning capabilities. However, substantial differences exist in information encountering experiences and frequencies among different individuals. How to develop targeted strategies based on the characteristics of university students' information encountering, combined with sensitive influencing factors, to stimulate and facilitate information encountering presents a worthwhile question.

Literature review reveals that many scholars have conducted research on the influencing factors of information encountering occurrence and frequency from

perspectives including personal factors, information factors, and network environment. However, few studies have focused on sensitive influencing factors of information encountering specifically among university student populations. Most existing research employs interviews, questionnaires, critical incident techniques, and experimental methods to collect data and construct influencing factor models, analyzing correlations among factors but lacking quantitative analysis and practical application research. From an application perspective, how to effectively utilize sensitive factors of information encountering to construct corresponding decision models for predictive analysis of new, unknown behavioral data represents the key to practical implementation of information encountering research. According to literature investigation, no existing literature has been found addressing quantitative analysis of information encountering frequency, and research on applying various information encountering influencing factors to predict unknown behavioral data remains scarce.

As an important branch of artificial intelligence, machine learning is a data analysis technique that has received widespread attention in recent years, currently applied extensively in natural language processing, computer vision, and other fields. Machine learning emphasizes extracting experience and domain knowledge from existing data to improve system performance and ultimately apply it to new, unknown data. If empirical data can be combined with machine learning to construct predictive models of information encountering behavior from actual collected data, it can effectively promote the practical application of information encountering research and make new attempts to expand research methods in information behavior. Based on this, this study addresses the prediction problem of information encountering behavior among university students in network environments, introducing the representative machine learning algorithm Support Vector Machine (SVM). Information gain is first used to quantitatively analyze corresponding sensitive influencing factors, and subsequently, an SVM-based classification prediction model for information encountering frequency is constructed and tested using survey data collected from university student populations to verify the model's validity and rationality.

2 Literature Review and Research Questions

Over the past two decades, numerous scholars have conducted research on the concept, process models, and influencing factors of information encountering.

2.1 Concept of Information Encountering

S. Erdelez first formally proposed the term "Information Encountering" in her 1995 doctoral dissertation, defining it as "the phenomenon of individuals accidentally obtaining information of interest or that can solve problems in unexpected situations" []. Since then, many scholars have proposed related concepts and definitions. For example, in 1996, K. Williamson et al. [] introduced the concept of "incidental information acquisition," defining it as "unexpectedly obtaining information while engaged in other activities." Scholar J. Heinström []

further defined it as "obtaining useful or interesting information without specifically searching for it." In 2000, S. Erdelez et al. [] proposed the concept of "information source encountering," suggesting that when using the internet to find information, many users have an opportunistic expectation of acquiring information when encountering an unfamiliar but potentially useful information resource. Also in 2000, E. G. Toms [] explored and discussed information encountering in browsing contexts, arguing that during information browsing across different topics, users focus on interesting and useful information they accidentally discover, and proposed the concept of "serendipitous information retrieval."

Although definitions and elaborations vary among scholars, they all emphasize two essential characteristics of information encountering: users' "low involvement" and "low expectation," as well as the prominent feature that encountered information is related to user interests or problems []. Based on domestic and international research, this paper defines information encountering as "an information acquisition behavior, specifically referring to the phenomenon where users accidentally obtain information they find interesting or useful when conducting various information activities using network terminals without specific purposes and with low expectations."

2.2 Research on Information Encountering Process Models

M. P. E. Cunha [] proposed a framework model for understanding the information encountering process from an organizational management perspective, consisting of four components: "precipitating conditions," "search for problem A," "bisociation," and "unexpected solution for problem B." L. McCay-Peet and E. G. Toms [] analyzed interview data from ten historians regarding "information seeking processes" and proposed an information encountering process model in knowledge work based on M. P. E. Cunha's model. V. L. Rubin et al. [] studied information encountering in daily life contexts, elaborated on all elements included in information encountering, and constructed an information encountering process element model. The above model research mostly focuses on conceptual feature elements in the information encountering process, while many scholars have also explored information encountering process models from a structured process perspective.

S. Erdelez proposed an information encountering process model in information search contexts, including five functional elements: noticing, stopping, examining, capturing, and returning []. Kuramura 伦久 revised this model, further emphasizing the utilization aspect of encountered information []. J. Lawley and P. Tompkins [] proposed an individual perception-based information encountering process model, dividing the information encountering process into six stages. S. Makri et al. [], based on empirical evidence, proposed an information encountering process model emphasizing that the core element of the model is establishing "new awareness connections."

2.3 Research on Influencing Factors of Information Encountering

S. Erdelez [] and K. Williamson [] identified three basic elements of information encountering behavior: information user, information environment, and encountered information. Current research on influencing factors of information encountering mostly revolves around these three elements.

2.3.1 Information User Perspective Research on how personal characteristics of information users affect information encountering includes aspects such as personal traits, information need motivation, information literacy, and information encountering experience. J. Heinström [] believes that user emotions, personality, and search style are important influencing factors of information encountering, with curious, extroverted, and studious users being more likely to experience information encountering. Taiwanese scholar Cai Yixin et al. [] found that curiosity and thirst for knowledge often stimulate information encountering; personal interests and specific information need motivations facilitate information encountering; users who frequently use the internet, can easily handle information-related situations, and have relatively high information sensitivity are prone to information encountering. Tian Lizhong and Yu Biyong [] argue that people who keep information confusion in mind, are not easily satisfied with search results, enjoy searching, have extensive browsing habits, and possess a curious mentality are more likely to have encountering experiences, while those with strong purposefulness and strategic thinking are less likely to have such experiences.

Guo Haixia [] believes that individuals with intrinsic motivation for information needs are more likely to gain information encountering experience than those with extrinsic motivation. Yuan Hong and Wang Zhipeng [] found in their study of information encountering phenomena in digital library usage that personal factors have a greater impact on information encountering than information factors, and that users' information literacy is a catalyst for obtaining more encountered information. S. Erdelez categorized information encounterers into four types: non-encounterers, occasional encounterers, encounterers, and super-encounterers []. Among them, super-encounterers experience information encountering at very high frequencies and treat information encountering as a method of information seeking. S. Erdelez believes that existing information encountering experience is an important factor affecting information encountering occurrence.

2.3.2 Information Environment Perspective The information environment can be understood as different contexts in which information encountering occurs. Network browsing, information retrieval, and information interaction are all common contexts. S. Erdelez [] found that scholars and researchers are prone to information encountering in network browsing states. K. Williamson [] believes that users often experience information encountering in information interactions with family and friends. Pan Shuguang's research suggests that

in network information retrieval, users' foreground problems and background problems can be interchanged, thereby stimulating information encountering; factors such as "low familiarity" with content and "goal-less browsing" in network browsing have important impacts on information encountering occurrence []. Tian Lizhong and Yu Biyong [] found that information encountering is more likely to occur in browsing contexts with low time pressure, weak purposefulness, and systematic feedback. Guo Haixia [] believes that without time pressure, information encountering experiences will increase and continue to emerge unstoppably. Du Xue and Liu Chunmao [] found that users in working states are more likely to experience information encountering.

2.3.3 Encountered Information Perspective S. Erdelez [] divides encountered information into two categories: problem-related and interest-related. Problem-related includes current, past, and future information needs. Information encountered related to past problems, although no longer needed by users and unable to generate direct value, can arouse users' interest in information sources and guide their future information behavior. Information encountered related to current and future problems saves users time and energy in obtaining information []. V. L. Rubin et al. [] classified users' information encountering outcomes from "very abstract" to "very concrete" based on the types of benefits users obtain from information encountering. The second category describes benefits users obtain by utilizing information related to past and current problems, namely obtaining solutions to previous problems or matters of concern. For problem-related encountered information, users know its specific use at the time of encountering. Unlike problem-related encountered information, interest-related encountered information mostly consists of surprising news, information fragments, or potentially useful information [], serving for entertainment without specific use, but it can broaden users' horizons, increase their knowledge reserves, potentially help them solve problems in the future, and help them identify problems []. Cai Yixin et al. [] believe that encountered information can be obtained by oneself or through others (directly, indirectly, or through public sharing); sources of encountered information are diverse, including various internet pages, BBS, email, search engines, and various network social tools and platforms. Tian Lizhong and Yu Biyong [] believe that information characteristics such as prominent external form and location, high source quality, interesting naming and content, and easy accessibility can increase the probability of information encountering. Guo Haixia [] believes that personally encountered information mainly prepares for individuals' possible future information needs and secondarily satisfies curiosity about a wide range of topics.

Based on the above analysis, existing research mostly focuses on theoretical foundations of information encountering, concentrating on positioning and relationship analysis within information behavior frameworks, and lacks empirical research on university students' information encountering behavior. Although some studies [] have analyzed the important significance of information encountering frequency, they have not conducted targeted quantitative analysis. Based

on existing results and combining the research object and objectives, this study proposes the following research questions: (1) What are the influencing factors of university students' online information encountering? (2) Among numerous influencing factors, which are the sensitive factors affecting information encountering? (3) How to construct a model of sensitive influencing factors of information encountering with predictive effectiveness for unknown behavioral data?

3 Research Design and Data Collection

3.1 Research Methods and Approach

The objective of this study is to identify sensitive influencing factors from numerous information encountering influencing factors. The key lies in how to quantitatively analyze sensitive influencing factors of information encountering from a data perspective and construct a classification prediction model. Information gain is an effective method for measuring feature importance, representing the difference in information quantity of a model with and without a particular feature. Larger values indicate that the feature can bring higher information quantity to the prediction model, meaning the feature is more important and has more obvious discriminative power. Compared with chi-square tests and other methods, information gain can comprehensively provide the importance of features for prediction models from an information theory perspective, making it effective for feature selection [] and attribute reduction [].

SVM is a representative machine learning method currently available. By seeking structural risk minimization, SVM achieves good generalization ability on limited samples; by introducing kernel functions, SVM provides effective solutions for nonlinear problems []. This aligns with the small sample size and the nonlinearity and intersection characteristics of information encountering behavior at the data analysis level. Therefore, this study introduces SVM and information gain as primary data analysis tools, where information gain is used to quantitatively analyze the correlation between influencing factors and information encountering frequency [], and SVM is used to establish a prediction model for information encountering frequency under limited samples.

The specific research methods and approach are shown in Figure 1 [Figure 1: see original paper] and can be summarized in the following steps: (1) Through literature review and meta-analysis, collect, organize, and categorize influencing factors of information encountering from existing research, analyzing and classifying them by dimension; (2) Based on discussion results regarding information encountering influencing factors from existing research, design interview outlines to further obtain more comprehensive influencing factor indicators; (3) Design and distribute questionnaires based on influencing factors collected through literature and interviews; (4) Use information gain to analyze correlations between each influencing factor and information encountering frequency, identifying sensitive factors and constructing a sensitive influencing factor model; (5) Use SVM to predict characteristics of information encountering frequency in test set data

and conduct error prediction and analysis of the sensitive influencing factor model.

3.2 Data Collection

Through literature investigation and interviews, relevant influencing factors of information encountering were obtained, statistically organized, and categorized, ultimately forming four dimensions: "information user," "encountered information," "information environment," and "situation," covering 12 aspects including "basic information," "personality traits," and "learning motivation," with a total of 32 specific influencing factor variables.

3.2.1 Variables of Information Encountering Influencing Factors (1) Information User Dimension Variables. These include basic personal information of information encountering users (Aa gender, Ab major, Ac grade), personality traits (Ad personality), learning motivation (Ae curiosity, Af thirst for knowledge), information literacy (information awareness and capability: Ba information needs expression, Bb information source evaluation capability, Bc capability to use common network tools, Bd proficient search engine usage; search style: C fast search style, D extensive browsing search style, E deep mining search style), and personal experience (X information encountering experience), totaling 14 specific variables across 5 aspects.

(2) Encountered Information Dimension Variables. These include information popularity of encountered information (F information popularity, G information novelty, H interest in information title), information quality (I information quality authority and scientific nature), and information content (J encountered information is very useful, K encountered information is related to current problems, L encountered information is related to past problems, M encountered information is related to future problems, N encountered information is interest-related), totaling 9 specific variables across 3 aspects.

(3) Situation Dimension Variables. These include information behavior situations when information encountering occurs (O information browsing, P information searching, Q information communication) and task situations (R task situation, S entertainment and leisure, V sufficient time), totaling 6 specific variables across 2 aspects.

(4) Information Environment Dimension Variables. These include network environment where information encountering occurs (T mobile internet, U computer internet) and system design (W system feedback), totaling 3 specific variables across 2 aspects.

3.2.2 Questionnaire Distribution and Recovery Based on the above 32 specific influencing factor variables and the final prediction indicator "information encountering frequency," a questionnaire with 33 questions was designed.

The questionnaire was published on the Wenjuanxing website, with survey subjects involving students from 3 majors (medicine, management, psychology) across 5 grades at Xinxiang Medical University. A total of 194 valid questionnaires were recovered. Using SPSS 22.0 to analyze the questionnaire's reliability, the Cronbach's alpha coefficient was 0.846, greater than 0.8, indicating acceptable reliability. The questionnaire used a 5-point Likert scale, with respondents selecting from "strongly disagree," "disagree," "neutral," "agree," and "strongly agree" based on their agreement with each statement, assigned corresponding weights of 1, 2, 3, 4, and 5.

4 Data Organization and Analysis

The 32 specific factors across four dimensions—information user, encountered information, information environment, and situation—were used as independent variables, with information encountering frequency as the dependent variable, to calculate information gain and analyze correlations, and to construct an information encountering influencing factor model. Based on the value of the prediction target "information encountering frequency," samples were divided into two groups: those selecting "agree" and "strongly agree" (values 4 and 5) were grouped as high information encountering frequency, while those selecting "strongly disagree," "disagree," and "neutral" (values 1, 2, and 3) were grouped as low information encountering frequency. Accordingly, 194 samples were constructed with input sample dimension of 32 and output sample dimension of 2 (high frequency and low frequency). First, information gain between each influencing factor and "information encountering frequency" was calculated, and sensitive factors were selected based on gain values. Then, the SVM model was introduced to construct an information encountering prediction model for predicting information encountering frequency.

4.1 Information Gain Calculation

In information theory and probability statistics, entropy represents the uncertainty measure of random variables, while information gain describes the information quantity that specific influencing factors bring to these random variables. By definition, the entropy of information encountering frequency is first calculated, as shown in Formula (1):

$$H(X) = -\sum p_i \log p_i \tag{1}$$

where p_i represents the probability of the ith random variable occurring, and X is the specific influencing factor. Larger entropy values indicate greater uncertainty in information encountering frequency. Second, the conditional entropy of information encountering frequency (i.e., Y) under the condition of known specific influencing factors (i.e., X) is calculated, as shown in Formula (2):

$$H(Y|X) = \sum P(X = x_i)H(Y|X = x_i) \tag{2}$$

Finally, the conditional entropy of information encountering frequency (Formula (2)) is subtracted from the entropy of information encountering frequency (Formula (1)) to obtain the information gain of specific influencing factors, as shown in Formula (3):

$$g(Y, X) = H(Y) - H(Y|X) \tag{3}$$

Based on Formula (3), the information gain of each feature can be calculated. The feature with the largest information gain value is selected as the optimal feature [].

4.2 Support Vector Machine Modeling

Support Vector Machine is a binary classification algorithm suitable for small samples. Based on statistical learning theory, SVM achieves good generalization ability on limited samples by seeking structural risk minimization and avoids overfitting. By introducing kernel functions, SVM provides effective solutions for nonlinear problems []. This aligns with the small sample size and the nonlinearity and intersection characteristics of information encountering behavior at the data analysis level.

The basic idea of SVM is to find a hyperplane between two types of data that maximizes the classification margin between positive and negative classes. In this study, among the data of "low information encountering frequency" and "high information encountering frequency," multiple classification models exist. The SVM model that maximizes the classification margin is the final prediction model for information encountering frequency, as shown in Figure 2 [Figure 2: see original paper]. Note that for drawing simplicity, Figure 2 only uses a linearly separable problem as an example; nonlinear classification and linearly inseparable problems can be extended through SVM's kernel functions and penalty functions.

In Figure 2, two-dimensional sample points of different shapes represent different information encountering frequencies as two distinct categories. Figure 2(a) shows that multiple classification models can completely separate the two types of data, but which model is optimal cannot be determined. In Figure 2(b), star-shaped points on the left and circular points on the right represent two types of samples. The solid line between classes is the classification line, and the dashed lines on both sides are parallel lines passing through samples closest to the classification line. The distance between them is the classification margin. According to SVM's basic principle, the classification line that correctly distinguishes the two types of samples and has the maximum classification margin is the optimal classification model. Sample points on the dashed lines are

called support vectors. Based on this principle, by constructing a minimization model and introducing Lagrange multiplier methods for solution, the general SVM classification model can be obtained: $f(x) = \text{sign}\,(\sum \alpha_i y_i K(x, x_i) + b)$, where α is the optimal Lagrange multiplier, $K(\cdot)$ is the kernel function for handling nonlinear classification, and b is the bias term. Specific theoretical details of SVM can be found in reference [].

5 Results and Analysis

5.1 Correlation Results Between Dependent and Independent Variables

The information gain values were sorted by magnitude, with results shown in Table 1 . For intuitive description, Figure 3 [Figure 3: see original paper] provides the ranking of information gain for each attribute. According to Formula (4), information gain value I should be greater than or equal to zero, and larger I values indicate stronger correlation. Table 2 and Figure 3 show that the ten independent variables most correlated with information encountering frequency are: O information browsing, P information searching, T mobile internet, X information encountering experience, N encountered information is interest-related, L information is related to past problems, V sufficient time, Bd proficient search engine usage, J encountered information is very useful, and M information is related to future problems.

5.2 Model Construction and Accuracy Analysis

5.2.1 Construction of Sensitive Influencing Factor Model for Information Encountering Based on the above results, the sensitive influencing factor model for information encountering was established, as shown in Figure 4 [Figure 4: see original paper].

5.2.2 Model Accuracy Analysis Based on the obtained sensitive influencing factors of information encountering, this study used Support Vector Machine as a classifier to predict information encountering frequency. First, the top 10 sensitive factors were extracted from each of the 194 questionnaires, and samples were constructed according to corresponding information encountering frequency levels. Second, 135 randomly selected samples were used as the training set and fed into the Support Vector Machine for training to construct an information encountering frequency prediction model. Finally, the remaining 59 samples were used as the test set and fed into the prediction model for evaluation. In the experiment, all samples underwent normalization preprocessing. SVM used the RBF kernel function $k(x, y) = \exp(-\sigma\|x - y\|^2)$, where σ is the kernel parameter. The LibSVM toolbox was used, with grid search and 5-fold cross-validation for model selection to determine the optimal combination of regularization parameter C and kernel parameter σ. The parameter optimization process is shown in Figure 5 [Figure 5: see original paper]. Prediction

effectiveness was evaluated using classification accuracy on the test set, i.e., the proportion of correctly predicted samples in the entire test set.

Figure 5 shows the parameter combination sequence number on the horizontal axis, automatically searched by LibSVM, with corresponding classification accuracy values shown in Table 2 . Table 3 shows that through SVM parameter optimization selection, the final classification accuracy on test data reached 82.96%, achieving favorable prediction results. This indicates that SVM is suitable for user information encountering behavior prediction and analysis research, while also proving the effectiveness of the 10 sensitive factors identified in Figure 2 and indirectly demonstrating the rationality of the proposed method. However, Table 3 also shows that the model did not achieve very high classification prediction accuracy. Possible reasons include the limitation of survey subjects to one university, resulting in homogeneity in behavioral characteristics and information literacy, as well as high similarity among individual questionnaire items leading to low data discriminability.

5.3 Analysis of Results

The above results show that this study extracted as many influencing factor indicators of information encountering as possible from multiple channels and identified the 10 most relevant sensitive factors from the perspective of constructing an information encountering frequency prediction model. As shown in Figure 2, these ten factors are basically consistent with existing research results, distributed across four dimensions: information user, encountered information, network environment, and situational factors. This not only corroborates previous research from another perspective but also provides a new application method for effectively obtaining sensitive influencing factors.

5.3.1 Situational Factors Dimension Among the ten most relevant factors, "O information browsing" and "P information searching" in the situational factors dimension rank first and second in information gain values, indicating that information encountering mostly occurs during information browsing and information searching. According to Yu Biyong's perspective [], an information encountering experience goes through a process of "information behavior - information acquisition - information need," where information is unexpectedly acquired in specific information behavior situations (information browsing, information searching, information communication), thereby triggering new information needs. In information browsing situations, users are in a state of "weak purposefulness" or even "no purpose," with "low expectation" or even "no expectation" for results—behavioral characteristics that exactly match the features of information encountering. Additionally, information browsing generally has the characteristic of low time pressure []. According to this study's results, "V sufficient time" is also a relatively relevant factor for information encountering occurrence ($I = 0.1100$, ranking 7th). Therefore, information encountering is more likely to occur during information browsing.

Information searching often occurs in certain task situations with relatively clear information needs and relatively insufficient time. Existing research mostly believes that information encountering is unlikely to occur in this context. However, during information searching, the massive, high-density network resources organized through hyperlinks promote information encountering. Coupled with personalized settings and recommendations from various search engines, the probability of information encountering undoubtedly increases. It is worth noting that in the information age, university students' information needs show a rapid growth trend, with obvious dependence on various external and internet information for learning, research, employment, and daily activities. Information browsing and information searching are the main ways for university student groups to obtain information []. Cultivating information encountering ability in these two contexts is of great significance for improving university students' information literacy.

5.3.2 Network Environment Dimension This study designed two questions regarding network environment: the impact of "mobile internet" and "computer internet" on information encountering occurrence. Among them, mobile internet represents the mobile internet environment, while computer internet represents the traditional internet environment. The results show that "T mobile internet" is more relevant to information encountering occurrence, ranking third ($I = 0.1416$). In the mobile internet environment, users' network usage frequency increases significantly. According to the 2012 China Internet Network Information Center (CNNIC) survey "Characteristics of Chinese Mobile Internet Users' Online Behavior" [], 72.2% of mobile internet users access the internet via mobile phones at least once a day, with nearly 60% using mobile phones to access the internet multiple times daily. The "2014 China Mobile Internet User Behavior Research Report" by iResearch Consulting shows that 67% of mobile users report "using multiple times per day" []. University students are one of the main user groups of mobile internet []. On the one hand, during frequent network usage, dynamic news, push information, and communication information continuously provide users with the latest information, which itself can stimulate information encountering. On the other hand, in the mobile internet environment, users' fragmented behaviors of "no purpose" and strong information communication willingness also facilitate information encountering.

5.3.3 Information User Dimension The results show that among the many personal-related factors, the two variables "X information encountering experience" and "Bd proficient search engine usage" have relatively high correlation with information encountering occurrence. Among them, "X information encountering experience" refers to personal past information encountering experiences having a strong impact on information encountering occurrence ($I = 0.1223$, ranking 4th). According to A. E. Foster et al.'s research, information encountering can not only strengthen individuals' understanding of problems and revise users' understanding of initial problems but also lead users in

a new direction to find new problem-solving approaches []. This turn in the information journey can produce positive action effects and positive emotional experiences [], positively influencing the next information encountering occurrence.

"Bd proficient search engine usage" represents users' higher personal information literacy, with strong information awareness and capability. Generally, users with relatively high information sensitivity who frequently use networks and are proficient in common tools can easily handle information-related situations and are more prone to information encountering []. On the other hand, proficient search engine usage makes it easier to master usage skills, utilize personalized recommendation functions to obtain large amounts of information, thereby stimulating information encountering. Like active information acquisition, information encountering capability can also be cultivated and improved. Using current information literacy practice models, information encountering awareness and capability can be guided and cultivated in various practice stages [].

5.3.4 Encountered Information Dimension Among the top ten relevant factors, four variables belong to the encountered information dimension: "N encountered information is interest-related," "L information is related to past problems," "J encountered information is very useful," and "M information is related to future problems." Among them, variable N represents that encountered information is interest-related, and variable J represents that encountered information is problem-related. According to the research results, "N encountered information is interest-related" is more relevant to information encountering occurrence ($I = 0.1130$, ranking 5th). Interest-related information encountering mostly occurs in task-free browsing situations, while problem-related information encountering mostly occurs in task-based searching situations []. In the current Web 2.0 environment, university student groups are major users of various social media applications. Using fragmented time to frequently access networks, they passively obtain "fragmented" information pushed by various platforms or obtained through communication. In this situation, surveyed university student groups believe they encounter more "interest-related" than "problem-related" information.

In the study, problem-related factors were further divided into "related to past problems," "related to current problems," and "related to future problems," corresponding to variables K, L, and M respectively. The results show that information related to "past problems" and "future problems" is more likely to stimulate information encountering. Among them, "L information is related to past problems" ranks 6th ($I = 0.1118$). In learning and life, there are often problems that cannot be solved at the time, and these problems are stored in users' subconscious as "background problems" []. When passively obtained information associates with "background problems," information encountering occurs.

This study focuses on university student groups, addressing three core issues:

"influencing factors of university students' online information encountering," "sensitive factors affecting information encountering," and "predicting characteristics of information encountering frequency for unknown behavioral data." From a machine learning perspective, sensitive influencing factor models and behavior prediction models for information encountering were constructed, and results were theoretically explained and analyzed.

In the mobile internet era, information encountering capability is an important component of university students' information literacy. The ultimate purpose of information encountering theory research is application. Based on this understanding, this study introduces machine learning into information encountering frequency prediction, representing a beneficial attempt at practical application. On the one hand, it can guide targeted improvement of university students' information literacy and autonomous learning capability. On the other hand, the results can be transferred to characteristic analysis and prediction of information behavior in different populations, such as analysis of how university researchers' information encountering capability enhances research work. Simultaneously, this study can provide innovative basis and theoretical guidance for various mobile terminal application service providers and software developers. For example, predicting users' information encountering behavior when using social media can recommend higher-quality push information, which is beneficial for further improving product design and service quality.

References

[1] ERDEEZ S. Information encountering: an exploration beyond information seeking[D]. Syracuse: Syracuse University, 1995.

[2] WILLIAMSON K. Discovered by chance: the role of incidental information acquisition in an ecological model of information use[J]. Library & information science research, 1998, 20(1): 23-40.

[3] HEINSTROM J. Broad exploration or precise specificity: two basic information seeking patterns among students[J]. Journal of the American Society for Information Science and Technology, 2006, 57(11): 1440-1450.

[4] ERDELEZ S, TOMS EG, RIOUX K, et al. Opportunistic acquisition of information: the new frontier for information user studies[J]. Proceedings of the American Society for Information Science and Technology, 2002, 39(1): 521-522.

[5] TOMS EG. Serendipitous information retrieval[EB/OL]. [2018-02-18]. http://www.doc88.com/p-9723328638065.html.

[6] PAN Shuguang. Research on information encountering[D]. Chongqing: Southwest University, 2010.

[7] ZHANG Qian, DENG Xiaozhao. Literature review on information encountering utilization research[J]. Library and Information Service, 2014, 58(20): 138-144.

[8] CUNHA MPE. Serendipity: why some organizations are luckier than others[EB/OL]. [2018-02-18]. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.540.9982&rep=rep1&typ

[9] MCCAY-PEET L, TOMS EG. The process of serendipity in knowledge work[M]. New Jersey: ACM, 2010.

[10] RUBIN VL, BURKELL J, QUAN-HAASE A. Facets of serendipity in everyday chance encounters: a grounded theory approach to blog analysis[J]. Information research, 2011, 16(3): 488.

[11] ERDELEZ S. Towards Understanding Information Encountering on the Web[C]//Proceedings of the 63rd annual meeting of the American Society for Information Science. Medford, NJ: Information Today, Inc., 2000, 47: 363-371.

[12] KURAMURA 伦久. Re-examination of user behavior models for information encountering: based on surveys of information encountering on the Web[J]. Library and information science, 2006(55): 47-69.

[13] LAWLEY J, TOMPKINS P. Maximising serendipity: the art of recognising and fostering unexpected potential? A systemic approach to change[EB/OL]. [2018-02-18]. http://www.cleanlanguage.co.uk/articles/articles/224/1/Maximising-Serendipity/Page1.html.

[14] MAKRI S, BLANDFORD A. Coming across information serendipitously? Part 1[J]. Journal of documentation, 2012, 68(5): 684-705(22).

[15] ERDELEZ S. Information Encountering: It's more than just bumping into information[J]. Bulletin of the American Society for Information Science & Technology, 1999, 25(3): 26-29.

[16] HEINSTROM J. Psychological factors behind incidental information acquisition[J]. Library & information science research, 2006, 28(4): 579-594.

[17] CAI Yixin, HUANG Yuanhe. Research on online consulting encountering experiences and personal characteristics[J]. Library and Information Science, 2010, 36(2): 16-34.

[18] TIAN Lizhong, YU Biyong. Research on influencing factors of researchers' information encountering[J]. Information Science, 2013, 31(4): 69-75.

[19] GUO Haixia. Investigation and research on information encountering in network browsing[J]. Journal of Intelligence, 2013, 32(4): 47-50.

[20] YUAN Hong, WANG Zhipeng. Research on information encountering phenomena in digital library utilization[J]. Library and Information Service, 2014, 58(17): 104-111.

[21] ERDELEZ S. Investigation of information encountering in the controlled research environment[J]. Information processing & management, 2004, 40(6): 1013-1025.

[22] PAN Shuguang. Research on information encountering in different contexts[J]. Information Research, 2012(8): 15-18.

[23] DU Xue, LIU Chunmao. Survey and experimental research on personality characteristics of influencing factors of network information encountering[J]. Library and Information Service, 2015, 59(11): 119-126.

[24] ERDELEZ S. Information encountering: a conceptual framework for accidental information discovery[C]//Proceedings of an international conference on research in information needs, seeking and use in different contexts. London: Taylor Graham, 1997: 412-421.

[25] RUBIN VL, BURKELL J, QUAN-HAASE A. Everyday serendipity as described in social media[J]. Proceedings of the American Society for Information Science & Technology, 2011, 47(1): 1-2.

[26] LIU Rujun, JIA Bin, XIN Yang. A multi-label feature selection algorithm based on information entropy[J]. Journal of Computer Research and Development, 2013, 50(6): 1177-1184.

[27] JIA Ping, DAI Jianhua, PAN Yunhe, et al. A new attribute reduction algorithm based on mutual information gain ratio[J]. Journal of Zhejiang University (Engineering Science), 2006, 40(6): 1041-1044, 1070.

[28] DING Shifei, QI Bingjuan, TAN Hongyan. An overview on theory and algorithm of Support Vector Machines[J]. Journal of University of Electronic Science and Technology of China, 2001, 40(1): 2-10.

[29] ZHANG Zhenhai, LI Shining, LI Zhigang, et al. A network anomaly detection model based on information gain feature selection[J]. Computer Applications, 2016, 36(S2): 49-53.

[30] YU Biyong. Empirical analysis of concepts and characteristics of information encountering: taking researchers as an example[J]. Journal of the China Society for Scientific and Technical Information, 2012, 31(7): 759-769.

[31] ZHOU Yanmei, LIU Dongsu, WANG Yanxi, et al. Investigation and analysis of university students' information behavior and information service strategies[J]. Library and Information Service, 2015, 59(6): 61-67.

[32] CNNIC. Characteristics of Chinese mobile internet users' online behavior[EB/OL]. [2017-03-25]. http://www.cnnic.cn/hlw fzyj/hlw xzbg/ydhlwbg/201211/P020121116518463145828

[33] 2014 China Mobile Internet User Behavior Research Report[EB/OL]. [2017-03-25]. http://www.useit.com.cn/thread-5941-1-1.html.

[34] CNNIC. 2015 Research Report on Chinese Youth Online Behavior[EB/OL]. [2017-11-25]. http://www.cnnic.net.cn/hlw fzyj/hlw xzbg/qsnbg/201608/P020160812393489128332.pdf.

[35] FOSTER AE, FORD NJ. Serendipity and information seeking: an empirical study[J]. Journal of documentation, 2002, 59(3): 321-340.

[36] ZHOU Pei, HUANG Chunyan. Meta-ethnographic analysis of information

encountering research[J]. Library and Information Service, 2014, 58(14): 115-120.

[37] TIAN Mei, ZHU Xuefang. Cultivation of university students' information encountering capability based on current information literacy models[J]. Library and Information Service, 2015, 59(17): 41-46.

[38] WANG Wentao, XIE Yangqun. Review of information encountering model research[J]. Library and Information Service, 2014, 58(21): 130-135.

Author Contributions

Tian Mei: Topic selection, conceptualization, and paper writing; Zhu Xuefang: Guidance on paper writing.

English Abstract

Study of Network Information Encountering Influence Factors for Undergraduate Group Based on Support Vector Machine

Tian Mei[1],[2], Zhu Xuefang[2]
[1] Research Center of Health Information Resources, Management Institute, Xinxiang Medical University, Xinxiang 453003
[2] School of information management, Nanjing University, Nanjing 210023

Abstract: [Purpose/significance] In the current Web 2.0 network environment, information encountering is one important method to get information for the undergraduate group. This study is of important significance of improving the ability of information encountering and information literacy for university students. [Method/process] Aiming at university students, this paper studies the sensitive influence factors of information encountering in the environment of network. Specifically speaking, this paper uses information gain to analyze the correlation between each influence factor and information encountering frequency, and then builds the model of sensitive influence factor. Furthermore, support vector machine (SVM) is introduced to establish the prediction model for information encountering frequency. [Result/conclusion] There exists 10 most sensitive influence factors for information encountering which are located in four dimensions including information user, encountering information, network environment and situation factors. The predicted classification accuracy can reach 82.96%, which demonstrates SVM works well to predict information encountering frequency.

Keywords: information encountering; information behavior; support vector machine; influence factors; information gain

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv — Machine translation. Verify with original.