Text Mining-Based Research on Medication Consultation in Online Health Communities (Postprint)

Authors: Yang Ping, Xiao Yigui

Date: 2024-03-28T00:00:00+00:00

Abstract

[Purpose/Significance] Online health communities constitute crucial channels for individuals to acquire health information. Investigating users' medication consultation needs within these communities facilitates the optimization of pharmaceutical services and the sustainable development of online health platforms.

[Method/Process] Taking 39 Health Network as a case study, this research first crawled 59,048 medication consultation comments pertaining to gastrointestinal drugs using Python scripts and performed data preprocessing. Subsequently, text mining methods including TF-IDF, TextRank, and LDA topic models were employed to extract thematic keywords from the experimental data, complemented by keyword co-occurrence network analysis. Finally, a comprehensive analysis of the thematic characteristics of users' medication consultation needs in online health communities was conducted, and corresponding optimization recommendations were proposed.

[Results/Conclusion] The findings indicate that users of online health communities primarily focus on drug therapeutic efficacy, administration methods, adverse reactions, distinctions between medications, and precautions for special populations (e.g., pregnant women) regarding medication use. This study provides, on the one hand, a theoretical basis for pharmaceutical manufacturers to adjust and optimize drug labeling content, and on the other hand, offers guidance for online health communities to optimize the layout of drug labeling content and to establish or improve pharmaceutical science communication services.

Full Text

A Study of Online Health Community Medication Counseling Based on Text Mining

Yang Ping, Xiao Yigui
(College of Business, Hunan University of Technology, Zhuzhou 412007, China)

Abstract

[**Purpose/Significance**] Online health communities have become an important channel for people to obtain health information. Studying user medication consultation needs in these communities contributes to optimizing drug services and promoting sustainable development of online health platforms. [**Method/Process**] Taking 39Health.com as a case study, we first used Python to crawl 59,048 medication consultation comments related to gastrointestinal drugs and performed preprocessing. Second, we employed text mining methods including TF-IDF, TextRank, and LDA topic modeling to extract thematic keywords and conducted keyword co-occurrence network analysis. Finally, we comprehensively analyzed the thematic characteristics of user medication consultation needs and proposed optimization recommendations. [**Results/Conclusion**] The findings reveal that users primarily focus on drug therapeutic effects, administration methods, adverse reactions, drug differences, and precautions for special populations such as pregnant women. This study provides a theoretical basis for pharmaceutical manufacturers to optimize drug instruction content and offers direction for online health communities to improve instruction layout and establish or enhance drug education services.

Keywords: Online health communities; Medication counseling; Text mining

---

1. Introduction

1.1 Health Information and Services Research

Online health communities are social platforms centered on medical and health topics, built upon internet information technology. These platforms host diverse users including the general public, patients and their caregivers, physicians, and healthcare institutions, who engage in online interactions instead of traditional face-to-face communication. By providing health services, sharing medical information, and disseminating disease treatment experiences, these communities help overcome spatial and temporal limitations of medical resources.

Online health communities offer various health information and services where information and service delivery are inseparable and mutually reinforcing. From the perspective of information publishers, these platforms provide two main categories of information. The first includes platform-published information about hospitals, physicians, drugs, and diseases, which supports services such as hospital and physician recommendations, electronic appointment scheduling, drug purchasing, and disease queries. Scholars have conducted extensive research on hospital and physician recommendations and disease diagnosis based on such information. The second category comprises information published by users such as physicians and patients. Some researchers have studied personalized health education recommendation services based on health popularization articles published by physicians, as well as various online consultation services. Others have analyzed user health information needs, health information behaviors, and emotional characteristics based on patient-generated content such as disease and drug consultations, experience sharing, and emotional exchanges among patients.

1.2 Health Information Mining Technology Research

Medical and health information in online communities primarily exists as interactive text generated through social behaviors, characterized by large volumes and complex structures. How to mine thematic content from such information and explore users' health information needs and behaviors has long been a research focus.

Traditional approaches to mining thematic content primarily employed content analysis and manual statistical annotation. For instance, Jin Biyi and Xu Xin collected diabetes-related questions from Yahoo Answers to understand consumer information needs regarding diabetes, using manual coding, text processing, multidimensional scaling analysis, and central word clustering to identify hot topics such as daily disease management, disease diagnosis, and treatment. Shi Yilong and Xu Xin employed content analysis to compare autism-related Q&A data from Baidu Zhidao and Yahoo Answers, finding that American users had better mastery of basic disease knowledge than Chinese users, asked more detailed and diverse questions, and showed more proactive exploratory questioning about diseases. However, traditional content analysis and manual annotation consume substantial human resources and time. With the maturation of text mining technologies, increasing numbers of scholars have applied automatic identification techniques such as text clustering algorithms and LDA topic models to health community research. For example, Tang Xiaobo and Li Jin conducted cluster analysis on hypertension Q&A texts in online health communities, discovering that users were most concerned about disease treatment, complications, and lifestyle. Zhang Li and Zhang Zhen performed text analysis on pharmaceutical e-commerce reviews using LDA topic modeling and sentiment analysis to identify consumer needs and pain points regarding online drug purchases during the pandemic.

1.3 Medication Counseling Research

Medication counseling refers to pharmaceutical services where pharmacists apply pharmaceutical expertise and clinical skills to provide drug therapy and rational medication guidance to patients and caregivers. As special commodities, drug safety directly affects users' health, and rational and standardized use represents a key principle for ensuring safety. However, most patients currently lack adequate medical literacy and frequently have questions about drug usage, making medication counseling essential.

Current medication counseling methods are divided into offline and online channels. Traditional offline counseling typically involves face-to-face guidance from pharmacists at pharmacies or hospitals, which can build sufficient trust but often requires significant time investment from consultees. With the rapid development of "Internet + Healthcare," medication counseling services have gradually expanded to online platforms, allowing consultees to directly ask pharmacists medication-related questions via the internet. This provides tremendous convenience, particularly during the COVID-19 pandemic when people prefer contactless healthcare services, leading to massive growth in online medication counseling.

As a product of the "Internet + Healthcare" era, online health communities also feature online medication counseling functions, enabling consultees to ask pharmacists any drug-related questions. Current domestic and international research on medication counseling services primarily focuses on offline outpatient settings, though some scholars have explored online services, particularly regarding quality management and evaluation of online counseling and service models. For example, Mei Xin et al. analyzed the practical effects of pediatric pharmacists providing children' s medication counseling through the "Wen Yaoshi" platform, offering references for subsequent development of pharmacist participation in clinical rational medication standards and industry regulations.

In summary, research on health information in online health communities is relatively mature with abundant achievements, but studies focusing on drug information and services remain limited. As an important drug service in online health communities, medication counseling significantly impacts community development. Users, as the main subjects of medication counseling, generate content that directly reflects medication needs and holds important value for service improvement. Given the mature research foundation of text mining technology in the healthcare field, this study examines user medication consultation comments in online health communities, employing multiple text mining methods including TF-IDF, TextRank, and LDA topic modeling, along with keyword co-occurrence network analysis, to discover users' highly concerned medication needs.

2. Research Design

This study's research design primarily includes data collection and preprocessing, text mining, experimental results analysis, and results discussion. The specific research framework is shown in Figure 1 [Figure 1: see original paper].

First, we collect user comment data from medication consultation modules in online health communities and preprocess the collected data. Second, we use TF-IDF, TextRank, and LDA topic modeling methods to mine thematic keywords from medication consultation comments. Third, we conduct thematic feature analysis of medication consultation comments, including word frequency statistical analysis, text thematic analysis, and keyword co-occurrence network analysis. Finally, based on the above analysis, we perform comprehensive analysis and discussion. The following sections elaborate on data collection and preprocessing and the text mining components.

2.1 Data Collection and Preprocessing

First, we select representative online health communities with drug databases. Second, we use Python to write web crawler programs to collect user medication consultation-related information. Third, we perform cleaning, deduplication, segmentation, and stop-word removal on the obtained text data. Finally, we obtain the experimental dataset.

2.2 Text Mining

This study employs three mainstream keyword mining methods—TF-IDF, TextRank, and LDA topic modeling—to conduct text mining on the experimental dataset. The following sections briefly introduce these three methods.

2.2.1 TF-IDF Algorithm The TF-IDF algorithm is a weighting technology based on word frequency statistics. Its advantages include simplicity, speed, and high computational efficiency. It can represent the importance of feature terms across the entire corpus, based on the principle that a feature term's importance is proportional to its frequency in a document and inversely proportional to its frequency across the corpus. The TF-IDF algorithm consists of two components: TF and IDF. The TF algorithm operates on the principle that the more times a feature word appears in a document, the better it represents that document. The IDF algorithm operates on the principle that the fewer documents containing a feature word, the stronger its ability to distinguish between documents.

The TF-IDF algorithm calculation formula [26] is shown in (1):

$$\text{TF-IDF}(t,d) = \text{TF}(t,d) \times \text{IDF}(t)$$

where $\text{TF}(t,d)$ is the term frequency of word t in document d, and $\text{IDF}(t) = \log\left(\frac{N}{df(t)}\right)$, with N being the total number of documents and $df(t)$ being the document frequency of term t.

2.2.2 TextRank Algorithm The TextRank algorithm is a graph model-based ranking algorithm that considers both word frequency and relationships between words. Its basic idea originates from Google's PageRank algorithm and can operate independently of the corpus by converting text into graph structures.

It uses iterative calculations to compute each node' s weight value, where larger weight values indicate more important words or phrases [27]. Its calculation formula is shown in (2):

$$S(V_i) = (1-d) + d \times \sum_{V_j \in \mathrm{In}(V_i)} \frac{w_{ji}}{\sum_{V_k \in \mathrm{Out}(V_j)} w_{jk}} S(V_j)$$

Where V_i and V_j are word nodes, d is the damping coefficient (generally valued at 0.85), $\mathrm{In}(V_i)$ is the set of word nodes pointing to node V_i, $\mathrm{Out}(V_j)$ is the set of word nodes that node V_j points to, and w_{ji} and w_{jk} are the edge weights from node V_j to V_i and from V_j to V_k, respectively [28].

2.2.3 LDA Topic Model The LDA model is a semantic-based method that considers contextual semantic relationships. Before using LDA for topic mining, it is necessary to determine the optimal number of topics. Current research often uses perplexity or coherence scores to determine the optimal topic number [29]. This study calculates the perplexity value of medication consultation texts and plots the perplexity curve to determine the optimal number of topics. Perplexity, proposed by Blei D. M. et al. [30] in 2003, primarily evaluates language model quality. While smaller perplexity scores indicate better text prediction, the curve' s inflection point must also be considered for comprehensive evaluation of appropriate topic numbers. The specific calculation formula (3) is as follows:

$$\mathrm{Perplexity}(D) = \exp\left(-\frac{\sum_{d=1}^{|D|} \log p(w_d)}{\sum_{d=1}^{|D|} N_d}\right)$$

Where D represents the text documents in the target data, w_d represents the word sequence in document d, N_d represents the total number of terms in the document, and $p(w_d)$ represents the generation probability of each document in the document set.

LDA is an unsupervised machine learning technique proposed by Blei D. M. et al. to address limitations in early LSA and PLSA topic mining models [30]. The model is based on a three-layer Bayesian probability model of "document-topic-term" that uses probabilistic statistics to model documents. It can automatically identify hidden thematic information in large-scale document collections or corpora, is highly effective for processing massive text information, and can improve text classification accuracy [31]. The core expression of the LDA topic model is shown in formula (4):

$$p(w|d) = \sum_{t=1}^{K} p(w|t) \cdot p(t|d)$$

where w represents a term in the document, d represents a document in the corpus, and t represents a topic.

When using the LDA topic model for user medication consultation text mining, this study first determines three parameters: α, β, and K, where $\alpha = 50/K$, $\beta = 0.01$, and K is the optimal topic extraction number determined by the perplexity curve.

3. Data Collection and Analysis

39Health.com is a leading and representative online health community in China, possessing the country' s largest database of hospitals, physicians, drugs, and personal medical information at various levels. Therefore, this study selects 39Health.com as the data source. The platform's drug database module includes numerous drug categories such as antipyretic and analgesic drugs, gastrointestinal medicines, and respiratory system drugs. According to the "2022 Online Medication Trends White Paper" released by Ali Health Research Institute and Zhongkang Technology, with the rapid growth of user groups and health demands, the OTC market has emerged with multiple new trend tracks including gastrointestinal health. Additionally, according to the "2022 Annual Pharmaceutical E-commerce White Paper" released by Yaorongyun, digestive system drugs account for the largest proportion in online sales channels. Therefore, selecting gastrointestinal drugs as the research object provides better representativeness.

Using Python web crawler programs, we collected user medication consultation comment data for gastrointestinal drugs related to digestion aid, inflammation, and gastric pain from 39Health.com' s drug database module. Data collection occurred on October 12, 2023, including data items such as drug names, user condition descriptions, and question times. We ultimately obtained 59,048 user comments across 2,996 drugs.

3.2 Data Preprocessing

Data preprocessing primarily includes data cleaning, deduplication, segmentation, and stop-word removal. Since crawled data is often messy and contains substantial irrelevant content such as website links that affect final results, data cleaning is essential. Additionally, many functionally identical drugs on 39Health.com have completely identical consultation content. For example, Qingkailing Granules (Baiyunshan) and Qingkailing Granules (Yuanda) have no significant difference in efficacy and their drug instructions differ, yet their consultation content is identical, requiring deduplication. After preprocessing, we obtained 28,250 user comment text documents as our experimental dataset.

We then performed segmentation and stop-word removal on the experimental dataset. Using Python's jieba tool for segmentation and considering the particularity of drug terminology where jieba' s segmentation of professional drug terms is insufficient, we extracted drug names from the collected dataset as an additional jieba dictionary, ultimately obtaining 2,298 drug names including Danqi

Tablets, making segmentation more scientific. For stop-word removal, we comprehensively used Chinese stop-word lists from Harbin Institute of Technology, Sichuan University, Baidu, and custom stop-words.

3.3 Text Thematic Feature Analysis

Using Python programs on the preprocessed data, we generated word cloud diagrams of medication consultation comments and extracted thematic keywords through TF-IDF, TextRank, and LDA topic modeling. Finally, we conducted keyword co-occurrence network analysis. Specific results are shown in Figure 2 [Figure 2: see original paper], Figure 3 [Figure 3: see original paper], Figure 4 [Figure 4: see original paper], and Table 1 .

3.3.1 Word Cloud Analysis The word cloud visualizes word frequency statistical analysis. High-frequency words include "effect," "efficacy," "side effects," "treatment," "administration," "duration," "results," "difference," etc., indicating that user medication consultation comments primarily involve drug effects, administration methods, adverse reactions, and drug differences. Notably, questions about drug effects, results, and efficacy—related to drug effectiveness—are most frequently described.

3.3.2 Text Thematic Keyword Analysis This study calculated the top 10 thematic keywords for medication consultation comments and their respective TF-IDF and TextRank values. Before LDA topic extraction, we calculated perplexity to determine the optimal topic number. After calculation and plotting the perplexity curve (Figure 3 [Figure 3: see original paper]), we found that perplexity was lowest at 5 topics, which also represented an inflection point, making 5 the appropriate optimal topic number. Using the gensim toolkit, we calculated the LDA distribution of comment texts with parameters $\alpha = 50/K$, $\beta = 0.01$, optimal topic number $K = 5$, and iteration passes = 100, extracting the top 10 keywords for each topic. Results are shown in Table 1 .

Table 1 Medication Consultation Thematic Keyword Extraction Results

- **TF-IDF**: "efficacy" : 0.29, "side effects" : 0.20, "administration" : 0.17, "effect" : 0.17, "treatment" : 0.11, "duration" : 0.10, "results" : 0.07, "difference" : 0.07, "diarrhea" : 0.05, "Jianweixiaoshi Tablets" : 0.04
- **TextRank**: "efficacy" : 1.0, "effect" : 0.99, "administration" : 0.59, "side effects" : 0.54, "treatment" : 0.53, "results" : 0.33, "difference" : 0.25, "diarrhea" : 0.14, "pregnant women" : 0.10, "gastritis" : 0.10
- **LDA Topic1**: 0.047"*efficacy" + 0.027*"administration" + 0.020"*treatment" + 0.016*" effect" + 0.013 "*Xiaoyaowan" + 0.011*" results" + 0.011 "*difference" + 0.008*" useful?" + 0.008 "*amoxicillin" + 0.008*" duration"
- **LDA Topic2**: 0.046"*effect" + 0.040*"efficacy" + 0.031"*side effects" + 0.026*" duration" + 0.019 "*treatment" + 0.018*" Jianweixiaoshi Tablets" + 0.015 "*administration" + 0.014*" results" + 0.013 "*difference" + 0.009*" injection"

- **LDA Topic3**: 0.035 "*effect*" *+ 0.026*" administration" + 0.025 "*treatment*" *+ 0.025*" efficacy" + 0.019 "*side effects*" *+ 0.016*" difference" + 0.012 "*Huoxiangzhengqi Liquid*" *+ 0.009*"enteric-coated tablets"+ 0.009"*lactation period*" *+ 0.009*" ofloxacin"
- **LDA Topic4**: 0.035 "*effect*" *+ 0.022*" side effects" + 0.017 "*efficacy*" *+ 0.016*" administration" + 0.015 "*difference*" *+ 0.014*" duration" + 0.014 "*Xiangshayangweiwan*" *+ 0.012*"stomach pain"+ 0.012"*treatment*" *+ 0.011*" vitamins"
- **LDA Topic5**: 0.057 "*efficacy*" *+ 0.054*" effect" + 0.038 "*side effects*" *+ 0.029*" administration" + 0.023 "*results*" *+ 0.015*" treatment" + 0.014 "*price*" *+ 0.013*" gastritis" + 0.011 "*duration*" *+ 0.011*" injection"

Table 1 reveals several key findings. First, keywords extracted by TF-IDF and TextRank, as well as those from each LDA topic, show high similarity, all containing terms such as "efficacy," "treatment," "effect," "results," "side effects," and "difference," indicating that drug effects, side effects, and drug differences are highly concerned content. Second, the five topics extracted by LDA show high similarity, making it difficult to distinguish each topic' s representative content through simple observation. However, examining differences between topics reveals that drug variations are the key factor causing topic differences, reflecting that regardless of drug type, user consultation content typically involves drug efficacy, administration methods, drug differences, and drug prices. Finally, synthesizing Table 1 results allows us to categorize user medication consultation content into several aspects: drug therapeutic effects (indicated by high-frequency terms like "efficacy," "effect," "treatment," "duration," "results"), drug categories (with drugs like Jianweixiaoshi Tablets, Xiaoyaowan, amoxicillin, Huoxiangzhengqi Liquid, ofloxacin, and Xiangshayangweiwan being commonly consulted for gastrointestinal diseases), disease types (with terms like "diarrhea," "gastritis," and "stomach pain" reflecting common symptoms), and administration methods, adverse reactions, and drug differences (indicated by frequent appearance of "administration," "side effects," and "difference").

3.3.3 Keyword Co-occurrence Network Analysis Word cloud and thematic keyword analysis can identify key consultation content but cannot reveal association strength between different keywords. This study uses keyword co-occurrence networks to further discover intrinsic connections between thematic keywords. Keyword co-occurrence network analysis calculates the frequency of multiple keywords appearing simultaneously in texts to determine their similarity relationships and conduct analysis [32]. This method uses high-frequency keywords as nodes and pairwise co-occurrence relationships as foundations to numerically process word relationships and graphically reveal structural relationships between words. Using Python, we segmented medication consultation texts, extracted keywords, and generated co-word matrices. We then selected the top 25 most frequent keywords for co-occurrence analysis and used Ucinet' s NetDraw visualization function to draw the keyword co-occurrence network, shown in Figure 4 [Figure 4: see original paper]. In the diagram, nodes represent high-frequency keywords with larger nodes indicating higher frequency, and connecting lines between nodes represent keyword co-occurrence relationships with thicker lines indicating more frequent co-occurrence and stronger connections.

In Figure 4, terms such as "efficacy," "effect," "side effects," "administration," "treatment," and "results" are important nodes representing core consultation content. The co-occurrence network reveals several consultation aspects: drug effectiveness (with strong connections between "administration" and "efficacy/effect/results," showing users' concern about post-administration effects), drug safety (with close associations between "adverse" and "reaction" and between "administration" and "side effects," reflecting concerns about adverse reactions and safety), administration methods (with tight connections between "before meals," "after meals," "efficacy," and "effect," indicating user attention to administration timing), and applicable populations (with children and pregnant women being special populations requiring numerous precautions, and the appearance of "pregnant women" reflecting their frequent consultations with pharmacists for safe medication use).

3.4 Discussion and Recommendations

Comprehensive experimental analysis reveals that user medication consultation content primarily involves drug therapeutic effects, administration methods, adverse reactions, drug differences, and precautions for special populations such as pregnant women.

Based on these findings and actual online health community operations, we propose several recommendations for optimizing drug services. First, users typically consult drug instructions before purchasing or using medications, with medication counseling serving as a supplementary service. However, our findings show that despite instructions covering core content like therapeutic effects, administration methods, and adverse reactions, users still consult pharmacists, indicating room for instruction optimization. Online health communities and pharmaceutical companies should supplement instructions with highly concerned content to address practical user questions and reduce pharmacist consultation pressure. Second, online health communities can optimize instruction layout by highlighting content related to drug effects, administration methods, and precautions to make key information immediately visible. Finally, communities can employ our research methods to mine user concerns and establish or improve drug education and promotion services.

4. Conclusion

This study collected user medication consultation comment data for gastrointestinal drugs related to digestion aid, inflammation, and gastric pain from 39Health.com' s drug database module. Through word frequency statistical analysis, text thematic feature analysis, and keyword co-occurrence network analysis, we identified users' most concerning issues during medication use. Results show that online health community users pay significant attention to drug effects, administration methods, adverse reactions, drug differences, and precautions for special populations such as pregnant women. These findings provide theoretical basis for pharmaceutical manufacturers to optimize drug instructions and offer direction for online health communities to improve instruction layout and establish or enhance drug education content.

References

[1] Cao B, Huang W, Chao N, et al. Patient activeness during online medical consultation in China: Multilevel analysis[J]. Journal of Medical Internet Research, 2022, 24(5): e35557.
[2] Fu Y, Tang T, Long J, et al. Factors associated with using the Internet for medical information based on the doctor-patient trust model: a cross-sectional study[J]. BMC Health Services Research, 2021, 21(1): 1268.
[3] 52nd Statistical Report on China' s Internet Development [EB/OL]. (2023-08-24) [2023-10-07]. https://www.cnnic.cn/n4/2023/0828/c88-10829.html.
[4] Yang P, Xiao Y, Zhong J. Research on users' drug purchase intention in online health communities based on the SOR model[J]. Journal of Mianyang Normal University, 2024, 43(1): 47-57.
[5] Wang J, Yao T, Wang Y. Patient engagement as contributors in online health communities: the mediation of peer involvement and moderation of community status[J]. Behavioral Sciences, 2023, 13(2): 152.
[6] Qiao W, Yan Z, Wang X. Join or not: the impact of physicians' group joining behavior on their online demand and reputation in online health communities[J]. Information Processing & Management, 2021, 58(5): 102634.
[7] Wang R, Wang J. Research on doctor recommendation in online consultation platforms from a user cognitive perspective[J]. Library and Information Service, 2023, 67(10):
[8] Huang J, Zhang Y, Zhang C, et al. Research on offline hospital recommendation based on online consultation text information[J]. Information Exploration, 2023(9): 88-93.
[9] Yuan H, Deng W. Doctor recommendation on healthcare consultation platforms: an integrated framework of knowledge graph and deep learning[J]. Internet Research, 2022, 32(2): 454-476.
[10] Kumar P M, Lokesh S, Varatharajan R, et al. Cloud and IoT based disease prediction and diagnosis system for healthcare using Fuzzy neural classifier[J]. Future Generation Computer Systems, 2018, 86(1): 527-534.
[11] Zhou H, Zhang P. Research on hybrid recommendation method for health popularization articles integrating LDA and TF-IWF[J]. Library Research, 2022, 52(3): 26-35.
[12] Yu B, Lu C. Research on thematic characteristics and potential value of information in online health communities: A case analysis of "Baidu Gout Bar" based on LDA model[J]. Price: Theory & Practice, 2022(3): 195-198, 206.
[13] Wu J, Wang Y, Li M, et al. User knowledge discovery in online health communities based on weighted knowledge networks[J]. Data Analysis and

Knowledge Discovery, 2019, 3(2): 108-117.
[14] Lin P, Lü J. Research on identification of question-answer information adoption in online health communities based on Stacking ensemble learning[J]. Information Science, 2023, 41(2): 135-142.
[15] Dong W, Li J, Tao J. Identification of active users and analysis of their interaction types in online health communities[J]. Journal of Library and Data, 2020, 2(1): 89-101.
[16] Lü J, Lin P. Analysis of the impact of user health information literacy and Q&A text emotional features on Q&A adoption in online health communities[J]. Intelligent Computer and Applications, 2023, 13(1): 5-11.
[17] Jin B, Xu X. Analysis of diabetes health information needs in social Q&A communities[J]. Chinese Journal of Medical Library and Information Science, 2014, 23(12): 37-42.
[18] Shi Y, Xu X. Analysis of autism information in Chinese and American online Q&A communities[J]. Chinese Journal of Medical Library and Information Science, 2015, 24(4): 5-8, 31.
[19] Tang X, Li J. Thematic analysis of information needs in online health communities[J]. Digital Library Forum, 2019(2): 12-17.
[20] Zhang L, Zhang Z. Research on the needs of online pharmaceutical consumers under the COVID-19 pandemic based on text mining[J/OL]. Operations Research and Management Science, 1-8[2024-02-20]. http://kns.cnki.net/kcms/detail/34.1133.G3.20230901.0849.002.html.
[21] Wang H, Yan S, Zhen J, et al. Formulation and analysis of medication consultation standards[J]. Herald of Medicine, 2022, 41(10): 1439-
[22] Mulder M B, Doga B, Borgsteede S D, et al. Evaluation of medication-related problems in liver transplant recipients with and without an outpatient medication consultation by a clinical pharmacist: a cohort study[J]. International Journal of Clinical Pharmacy, 2022, 44(5): 1114-1122.
[23] Deng Z, Xu C. An example of outpatient medication consultation using pharmacist service patient flow[J]. Chinese Journal of Hospital Pharmacy, 2021, 41(20): 2142-2145.
[24] Mei X, Ji L, Peng X, et al. Practice of pediatric pharmacists providing internet medication consultation services based on the "Wen Yaoshi" platform[J]. China Pharmacy, 2023, 34(12): 1520-1523.
[25] Jing L, He T. Chinese text classification model based on improved TF-IDF and ABLCNN[J]. Computer Science, 2021, 48(S2): 170-175, 190.
[26] Zhou H, Liu J, Zhang P, et al. Research on the usefulness of comments in online health communities from a complex network perspective[J]. Information Science, 2022, 40(9): 88-97.
[27] Yang D, Hu C. Keyword extraction method for scientific texts based on improved TextRank[J/OL]. Journal of Computer Applications, 1-9[2024-02-20]. http://kns.cnki.net/kcms/detail/51.1307.tp.20230825.1605.006.html.
[28] Yang Y, Zhao G, Yuan Z, et al. TextRank keyword extraction method integrating semantic features[J]. Computer Engineering, 2021, 47(10): 82-88.
[29] Ran C, Li W. Construction of enterprise competitor identification model based on LDA: A case study of NIO Automobile Co., Ltd.[J]. Information

Studies: Theory & Application, 2023, 46(8): 88-95.
[30] Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation[J]. Journal of Machine Learning Research, 2003, 3(4-5): 993-1022.
[31] Yang L, Wang Z, Hou G. Research on topic mining in online health communities based on Q-LDA topic model[J]. Data Analysis and Knowledge Discovery, 2019, 3(11): 52-59.
[32] Kim C, Na Y. Consumer reviews analysis on cycling pants in online shopping malls using text mining[J]. Fashion and Textiles, 2021, 8(1): 38.