Comparative Study of Multi-Platform Consumer Review Topics Based on Co-Occurrence Networks and Sentiment Analysis: Postprint

Authors: Zhou Tingwei

Date: 2023-10-08T00:00:00+00:00

Abstract

[Purpose/Significance] This study aims to empirically analyze the similarities and differences in thematic tendencies of user evaluation content for the same product across virtual lifestyle communities, social platforms, and shopping platforms.

[Method/Process] The "你今天真好看" App, Weibo, and JD.com were selected as experimental subjects, and a total of 54,071 user comment texts on the same skincare product were collected. LDA topic modeling, co-occurrence network analysis, and machine learning-based sentiment analysis methods were employed to conduct a multi-platform comparative analysis of user comment texts.

[Results/Conclusions] The study found that the eight major comment themes across the three platforms exhibited similarities and differences in thematic characteristic words, co-occurrence networks, and thematic sentiment, with content tendencies aligning with the characteristics of each platform.

Full Text

A Comparative Study of Multi-Platform Consumer Review Topics Based on Co-Occurrence Network and Sentiment Analysis

School of Management, Shanghai University, Shanghai 200444

Abstract

[Purpose/Significance] This study aims to empirically analyze the similarities and differences in thematic tendencies of user evaluation content across virtual life communities, social platforms, and shopping platforms for the same product.

[Method/Process] Selecting the "You Really Beautiful"App, Weibo, and JD.com as experimental subjects, we collected 54,071 user comment texts about the same skincare product. We employed the LDA topic generation model, co-occurrence network analysis, and machine learning-based sentiment analysis methods to conduct a multi-platform comparative analysis of user comment texts. [Result/Conclusion] The study reveals that across the three platforms, eight comment topics exhibit both similarities and differences in thematic feature words, co-occurrence networks, and topic sentiment, with content tendencies aligning with the characteristics of each platform.

Keywords: multi-platform comparison, text topic clustering, co-occurrence network analysis, sentiment analysis

1. Introduction

In the era of big data and mobile internet, users can exchange information online [1], and instant information presented in the form of comments is permanently retained on platforms. In this context, users gather together through shared interests, professions, goals, and needs, forming network-based virtual communities [2]. Compared with traditional physical communities, virtual communities are free from temporal and spatial constraints, thus reaching a broad audience [1]. In the new media era, virtual community marketing has gradually become the primary mode of information dissemination and value transfer in e-commerce. Virtual communities provide vast amounts of information for consumers to browse, fulfilling their functional needs for decision-making assistance.

Among these, the textual content of online reviews plays a crucial role in influencing consumer behavior, though not all reviews carry equal importance [3]. Some literature has categorized user review topics in virtual communities based on user needs into platform promotion, reward systems, infrastructure maintenance, professional features, posting norms, professional resource construction, management team building, social integration, and user interaction [4]. However, such research primarily focuses on virtual community construction and examines user participation characteristics and impacts. Studies on user reviews within virtual communities concentrate on single communities, lacking research on the thematic tendencies of characteristic user review content in virtual communities and cross-platform comparative studies.

This study collected 54,071 comment data entries for the same skincare product category from three channels: the "You Really Beautiful" App, Weibo, and JD.com. Using skincare as a case study, we compared skincare virtual communities, social platforms, and shopping platforms, employing LDA (Latent Dirichlet Allocation) topic generation models, co-occurrence networks, and sentiment analysis to conduct multi-platform comparative research on user comment texts. This paper aims to analyze the similarities and differences in comment topic

tendencies across platforms, compare content differences under the same topic, and provide brand enterprises with precise decision-making references for new product formula design, existing product optimization, target user positioning, product demand acquisition, actual usage feedback, product spokesperson selection, and product pricing, while also offering references for related commercial applications.

2. Literature Review

2.1 Virtual Communities and Users

Current research on virtual communities primarily focuses on virtual academic community construction and user participation in virtual brand communities. Zhou Yang et al. [5] used qualitative analysis software NVIVO to construct a model of participation behavior in virtual academic communities, finding that users at different levels exhibit different participation characteristics. C. Huaruo et al. [6] discovered that continuous knowledge-sharing intention in virtual academic communities is significantly positively correlated with knowledge-sharing satisfaction, which is influenced by heuristic factors (knowledge-sharing quantity, source credibility) and systematic factors (knowledge-sharing quality, knowledge-sharing usefulness). M. Tabish et al. [7] used Partial Least Squares (PLS) analysis to argue that virtual community participation positively affects brand trust, which in turn positively influences brand choice.

To uncover hidden user needs in virtual community reviews, relevant literature has analyzed user comment texts. Although online review content critically impacts consumer behavior, Chen Xiuxiu [3] found that not all reviews carry equal weight. Further research has identified connections between user needs and brands, uncovering the value of review texts containing user demand content. For instance, P. Cara [8] studied how consumers' perceptions of online review source credibility and usefulness affect intentions to engage with retail store online communities and purchase intentions, finding that consumers desire engagement with retail store network communities in shopping contexts. U. Chakraborty [9] used structural equation model bootstrapping methods to test the mediating effects of brand equity dimensions between purchase intention and online reviews, ultimately finding that marketers should focus more on brand awareness and perceived value, as these ultimately affect consumers' purchase intentions. Qu Huijun [10] studied the pairwise relationships among user participation motivation, interaction vitality, trust, and purchase intention in Weibo brand communities, finding that all four factors are mutually positively correlated.

In summary, current literature on virtual communities mainly includes two aspects: characteristics of different user participation behaviors in professional virtual communities, participation willingness, and impact on virtual communities; and text mining analysis of user review content to study hidden user needs for virtual communities and brands, as well as factors influencing user

trust and purchase decisions. However, the practical application value of user participation in commerce is vaguely defined, and user feedback on brand products constitutes an important manifestation of their co-created value, yet few studies have deeply investigated this aspect of user participation. Most research on virtual community reviews focuses on virtual brand communities, with few studies addressing virtual life communities.

2.2 Topic Mining, Co-Occurrence Networks, and Sentiment Analysis

Existing literature typically employs two methods for text topic clustering. The first is traditional models based on text similarity calculations [11-12], which operate on original texts and measure similarity through character matching degree or distance between two texts. While simple in principle and easy to implement, these models rely on word vector space distance [13], suffering from high text vector dimensionality and poor semantic sensitivity [14]. The second method uses the LDA topic generation model.

The LDA model, first proposed by D. M. Blei et al. [15], is an unsupervised machine learning technique based on Bayesian probability and one of the two major thrust models in natural language processing. It can identify latent topic information in large-scale corpora. In the LDA topic generation model, each document is represented as a multinomial distribution of topics, and each topic is represented as a multinomial distribution of words [16]. Due to its strong ability to process large-scale corpora and reduce dimensionality, the LDA topic generation model has become a popular direction in topic mining research [17].

The model can be applied to various contexts. In historical research, He Lin et al. [18] used LDA to automatically identify and classify event trigger words in historical texts such as *Zuo Zhuan*. In judicial practice, Wang Hui et al. [19] used LDA to extract case triple information, solving the problem of insufficient internal information association and inefficiency in existing judicial data analysis methods. In social platform public opinion management, Zhang Lei et al. [20] constructed a Weibo user topic generation model for university teacher ethics public opinion based on the LDA model, guiding public opinion by identifying optimal transmission paths and achieving prediction and control optimization for sudden public opinion events. In e-commerce, Cui Ning et al. [21] built an online product review analysis model based on LDA topic models and partial order set theory to help merchants grasp customer consumption needs and potential tendencies. In summary, the LDA topic generation model is widely applicable and can efficiently and accurately mine text topic features.

Machine learning-based sentiment analysis, as supervised learning, requires manual annotation of positive and negative samples in the text corpus before selecting appropriate algorithms to train classifiers [22]. Natural language processing classification algorithms mainly include Naive Bayes, KNN, Support Vector Machines, and neural networks. Research shows that when classifying text information of fewer than 100,000 entries using machine learning methods, the Naive

Bayes method achieves good results [23]. This study will use the Naive Bayes algorithm to train classifiers for processing the positive and negative polarity of text sentiment across the three platforms.

3. Methodology

3.1 Data Collection

This study selected user comments from skincare virtual communities, social platforms, and shopping platforms between June 2020 and March 2022, choosing "Freeplus Mild Soap" as the target product for comment data collection. According to iMedia Research data [24], 90.4% of Chinese netizens use facial cleanser to clean their faces, indicating that facial cleanser use is extremely common among consumer groups. During the 2021 JD.com "618" promotion event, Freeplus Mild Soap ranked first in sales in the facial cleanser category and has consistently remained at the top of the comprehensive ranking list in the skincare facial cleanser category on the "You Really Beautiful" App. Therefore, selecting Freeplus Mild Soap as the target object yields representative and generalizable research conclusions. Specific data sources are as follows:

(1) User comment texts from the skincare virtual community were crawled from the "You Really Beautiful" App. Launched at the end of 2016, "You Really Beautiful" is the world' s first mobile application that can detect skin quality through makeup-free photos. As a professional skincare virtual community, it provides skin quality detection while recommending suitable skincare products based on detection results. The platform features a product database of over 2 million skincare products, more than 15 million product reviews, and 12.9 million users. As of January 2023, it has provided over 237 million photo-based skin testing services (data from the App' s official website). This study crawled user comment data under the "Freeplus Mild Soap" product, ultimately obtaining 31,359 valid comment texts.

(2) User comment texts from the social platform were crawled from Weibo. Weibo is a broadcast-style social media and network platform based on user relationships for information sharing, dissemination, and acquisition through a follow mechanism. As of September 2021, Weibo' s monthly active users reached 573 million, with 94% from mobile devices, and daily active users reached 248 million [25]. This study crawled user blog post data under the "Freeplus Cleanser" topic, ultimately obtaining 12,404 valid comment texts.

(3) User comment texts from the shopping platform were crawled from JD.com. As a Chinese self-operated e-commerce enterprise, JD.com online sales include 13 major categories with 31.5 million SKUs (Stock Keeping Units) of products, including cosmetics and other personal care items. This study crawled user comment data under the "Freeplus Mild Soap" product, ultimately obtaining 10,308 valid comment texts.

This study obtained a total of 54,071 text entries from the three platforms.

3.2 Review Text Topic Generation Model

The main process of this study is divided into the following steps: (1) Use Python crawlers to obtain comment texts for "Freeplus Mild Soap" on the "You Really Beautiful" App, Weibo, and JD.com, forming a multi-platform comment text database to prepare for subsequent text analysis. (2) Preprocess comment texts by adding domain-specific terms, performing word segmentation, removing stop words, and annotating part-of-speech. (3) Use the LDA topic generation model to cluster comment texts from the three sources separately, determine the optimal number of topics based on perplexity, and manually define each topic. (4) Use the top 150 feature words clustered from each platform' s user comment texts as keywords to study the co-occurrence networks within and between topics, and conduct sentiment analysis on comment texts and topics. (5) Compare the co-occurrence networks and sentiment analysis results across the three platforms to explore the focus areas of each platform. The specific process is shown in Figure 1 [Figure 1: see original paper].

This section elaborates on the more important parts of the research process:

(1) **Data Preprocessing**. After obtaining comment texts, we removed stop words, annotated part-of-speech, and performed word segmentation with reference to the Harbin Institute of Technology stop word list (767 entries). To ensure the completeness and accuracy of word segmentation, we added a manually defined popular word list (97 entries) after reading all comments to assist Python' s third-party package Jieba in word segmentation.

(2) **LDA Topic Generation Model**. This study used Python' s Latent Dirichlet Allocation function to perform LDA clustering on each platform' s user comment texts separately and used the third-party package pyLDAvis for visual analysis of topic generation results to avoid overfitting. Based on LDA clustering results and referring to the elbow method [26] to find perplexity inflection points for the three platforms, the optimal topic numbers are shown in Figure 2 [Figure 2: see original paper]. According to results from all platforms, after deduplication of feature words under individual topics, we constructed a total feature word database for the three platforms. We ranked feature words by word frequency for each platform to obtain each platform' s respective feature word lists.

(3) **Topic Co-Occurrence Network Analysis**. After multiple experiments in KH Coder, network diagrams with 150 members yielded the best visualization effects. Therefore, based on LDA clustering results, we selected the top 150 words from feature word frequency tables to draw network diagrams. Verification showed that the proportion of coded words in total word frequency was 85.36% for "You Really Beautiful" App, 95.12% for Weibo, and 92.7% for JD.com, indicating representative results. Finally,

based on topic feature words and their co-occurrence relationships within the same comment text, we defined topic names. Using KH Coder software, we assigned topic codes to feature words by replacing multiple topic feature words with the same topic name as synonyms, enabling the study of co-occurrence relationships between topics—i.e., the co-occurrence phenomenon of feature words from different topics within the same comment text. We used cosine distance for co-occurrence analysis and semantic network construction between words and between topics. The topic-topic co-occurrence matrix example is shown in Table 1 , where diagonal elements represent the total frequency of a topic appearing in user comment texts obtained from that platform, and off-diagonal elements represent the total co-occurrence count of two topics within the same comment text. This paper used KH Coder and Gephi software to visualize co-occurrence networks for each platform, employing the minimum spanning tree method to extract the backbone of co-occurrence networks for simplification and more intuitive results.

(4) **Topic Sentiment Analysis**. Based on topic clustering and definition results, we divided word sentiment polarity into positive, neutral, and negative according to literature [27], labeling them as 1, 0, and -1 respectively for training and statistics. We used the Tf-idf method to extract features to highlight important words and suppress secondary ones [22]. For training data, we adopted the Naive Bayes method, calculating posterior probability categories based on known prior probabilities and conditional probabilities through word-class combination probabilities [23]. We trained classifiers using transformed word vectors with a training-to-testing set ratio of 8:2, and used Python' s Sklearn package to evaluate accuracy, obtaining results of 0.8201 for "You Really Beautiful," 0.7812 for Weibo, and 0.8059 for JD.com.

4. Results

4.1 LDA Topic Clustering Analysis

This study used Python for LDA clustering. Based on clustering results, we summarized eight topics across the three platforms. Among them, six topics are common to all three platforms: product formula (ingredients/skin feel), suitable skin types (user skin problems), product price, user experience (usage experience, usage combinations), user purchase reasons (why purchased, competing products, repurchase intention), and merchant activities. Additionally, product spokesperson is a unique topic to Weibo, and purchase platform/logistics is unique to JD.com. LDA topic clustering results are shown in Table 2 .

Some feature words appeared under two topics, but considering topic definitions, the actual meanings of feature words under each topic do not overlap, so they were retained. We calculated the total proportion of each platform' s topic feature words in that platform' s comment text database (i.e., after topic coding

processing), as shown in Figure 3 [Figure 3: see original paper]. Due to topic co-occurrence phenomena, the sum of proportions for each topic across platforms exceeds 100%. We selected four topics with relatively high proportions across all platforms for analysis through word cloud comparison of actual user comment text content.

(1) **Product Formula (Ingredients/Skin Feel)** topic feature word clouds are shown in Figure 4 [Figure 4: see original paper]. User comments across the three platforms show high attention to the basic formula attributes of facial cleanser, such as amino acids versus soap-based formulas, and the basic functional attribute of facial cleanser (i.e., cleansing power) and the characteristic of this amino acid facial cleanser (i.e., non-drying effect after washing). The "You Really Beautiful" App shows more attention to other functional ingredients beyond basic facial cleanser attributes, while also focusing on whether ingredients are irritating to skin or potentially cause acne—concerns not present in the other two platforms. Weibo users focus more on whether pregnant women can use this product, including whether ingredients are plant-derived and whether they contain fragrances, pigments, or preservatives. The first two platforms show inclusion relationships with JD.com' s feature words under this topic.

(2) **Suitable Skin Types (User Skin Problems)** topic feature word clouds are shown in Figure 5 [Figure 5: see original paper]. The "You Really Beautiful" App has significantly more feature words that clearly include those from the other two platforms under this topic. This App provides more detailed descriptions of facial areas (e.g., cheeks, forehead, nose, T-zone) and more specific descriptions of existing skin problems beyond general terms like pores, oil, and sensitive skin, adding terms like blackheads, acne, combination skin (partially neutral and partially oily), and oily acne-prone skin. Compared with the other two platforms, it also includes descriptions of applicable climate conditions, such as seasonal changes, autumn/winter, and periods.

(3) **User Experience (Usage Experience, Usage Combinations)** topic feature word clouds are shown in Figure 6 [Figure 6: see original paper]. User comments across the three platforms show considerable attention to basic facial cleanser efficacy (i.e., foaming ability and cleansing power). The "You Really Beautiful" App has significantly more feature words than the other two platforms under this topic. This App tends to focus on usage experience, covering various product effects comprehensively, including oil control, water-oil balance, reducing skin burden, particle addition, and post-use skin problems (e.g., breakouts, closed comedones, eye irritation). Weibo tends to focus on usage combinations, including using with foaming nets and subsequent pairing with other skincare products like Estée Lauder and CPB masks and creams. JD.com users place more emphasis on product packaging boxes, product capacity, and quality.

(4) **User Purchase Reasons (Why Purchased, Competing Products,**

Repurchase Intention) topic feature word clouds are shown in Figure 7 [Figure 7: see original paper]. Users across the three platforms primarily purchase this product based on its main characteristic attributes: brand (Freeplus), skincare category (facial cleanser), main ingredient (amino acids), and main function (cleansing), and all mention competing product brands like POND' S, Curel, and Elta. The "You Really Beautiful" App shows more positive expressions of repurchase intention, with first-time purchases often influenced by online media dissemination. Weibo and JD.com users' first-time purchase or usage reasons come from recommendations or gifts from friends and relatives, reviews under the same product on the same platform, and trust in the brand.

4.2 Keyword and Topic Co-Occurrence Network Analysis

Each platform' s keyword co-occurrence network is shown in Figure 8 [Figure 8: see original paper], where bubble size represents word frequency and line darkness represents edge weight. Color partitioning basically aligns with LDA topic clustering results.

In user comments from the "You Really Beautiful" App, high-frequency key nodes with connections to multiple words are relatively more numerous than in the other two platforms, with the connection between product formula and usage experience receiving the most attention. Weibo features the unique topic of product spokesperson. The abnormal phenomenon where user purchase reasons account for 99.9% of topics in Figure 3 for Weibo occurs because, as China' s mainstream public social media platform, brands place large amounts of product advertisements on it, causing abnormally high word frequency for brand and product name-related terms. The user purchase reasons topic occupies an extremely high proportion, demonstrating that Weibo' s marketing characteristics far exceed other attributes. JD.com features the unique topic of purchase platform/logistics. Based on relationships between keywords, users in this topic focus most on seller and courier service attitudes and logistics delivery speed. As a shopping platform, JD.com' s keyword co-occurrence network also shows the only strong associations among the three platforms: price with "Double Eleven" activities, "gifting" with "outer packaging," and "online" with "reviews."

Each platform' s topic co-occurrence network is shown in Figure 9 [Figure 9: see original paper], where line thickness represents co-occurrence frequency between topics. (1) User experience is the core topic in the "You Really Beautiful" App, with the highest co-occurrence frequency with product formula. Product formula, user experience, user purchase reasons, and suitable skin types all show strong pairwise associations, confirming that users focus most on the connection between product formula and user experience, which constitutes user purchase reasons, while user experience is also influenced by whether users belong to the suitable skin type corresponding to the product formula. (2) In the Weibo platform, although the product spokesperson topic has a high proportion in word frequency, it is relatively isolated overall, constituting only a small part

of user purchase intention. Merchant activities replace suitable skin types as the fourth element in the core topic group, and its high-frequency co-occurrence relationship with user purchase reasons, combined with the statistical result that product price is the highest proportion topic among the three platforms in Figure 7, indicates that Weibo users have higher information dissemination enthusiasm for merchant activities and prices. (3) In JD.com, the unique topic of purchase platform/logistics forms strong co-occurrence relationships with other topics, caused by JD.com' s inherent shopping platform attributes and its characteristic of self-operated logistics, not because other topics are influenced by purchase platform/logistics.

4.3 Topic Sentiment Analysis

This study obtained sentiment distributions for each topic across the three platforms, shown in Table 3 . Using overall proportion as weights, we calculated weighted average sentiment intensity (positive, neutral, negative) for each topic across the three platforms, shown in Figure 10 [Figure 10: see original paper].

Comprehensive sentiment intensity across eight topics shows total three-pole sentiment intensity of 0.75, 0.1, and 0.15 for the "You Really Beautiful" App; 0.07, 0.79, and 0.05 for Weibo; and 0.83, 0.04, and 0.02 for JD.com. Weibo' s extremely high proportion of neutral comments results from setting marketing promotion text sentiment polarity as neutral during machine learning training, confirming that Weibo user comment texts contain large amounts of marketing placement content. JD.com' s highest positive sentiment proportion results from the platform' s unique "habitual positive review" attribute among its users, which also appears extensively in comment texts. JD.com' s relatively higher neutral comments result from many comments like "sensitive skin should be usable," "okay," and "not bad," reflecting that platform users have relatively lower enthusiasm for detailed descriptions of their skin problems. The "You Really Beautiful" App' s user comment texts show higher negative sentiment proportions across all topics, as negative comments mainly focus on skin problems caused by product use. Notably, product price has a negative comment proportion as high as 0.51 on this platform, meaning over half of users consider the product price high. According to Baidu Index statistics, as of April 7, 2022, unlike Weibo (83.21%) and JD.com (85.94%) whose users are aged 20-39, this platform's users show a younger characteristic, with 73.87% under age 29. Users in this age group have lower economic levels, and the price of 125 RMB/100g for Freeplus Mild Soap represents a relatively high consumption level for this demographic. Overall, sentiment fluctuations across topics on the three platforms remain basically stable.

5. Discussion and Conclusion

This study compared comment texts from skincare virtual communities, social platforms, and shopping platforms. Based on LDA topic generation model clustering results, we categorized user comment texts into six common topics:

product formula (ingredients/skin feel), suitable skin types (user skin problems), product price, user experience (usage experience, usage combinations), user purchase reasons (why purchased, competing products, repurchase intention), and merchant activities, plus two unique topics: product spokesperson (social platform) and purchase platform/logistics (shopping platform). We summarize two main findings:

(1) The skincare virtual community provides the most detailed and comprehensive descriptions of product formula, suitable skin types, user experience, and user purchase reasons, making it the best channel for brands to acquire user needs and purchase decision factors. Social platforms aggregate large amounts of advertising information, making mining of authentic user feedback difficult. Shopping platform users have higher willingness to express positive emotions in comments but relatively lower enthusiasm for detailed skin problem descriptions, showing stronger privacy concerns and lower sharing desire.

(2) Three topics—product formula, user experience, and user purchase reasons—form strong associations across all three platforms, constituting the common components of each platform' s core topic group. The product spokesperson topic, though high in word frequency proportion, is relatively isolated overall, constituting only a small part of user purchase reasons. Each platform' s unique core topic group reflects its characteristics: suitable skin types for skincare virtual communities, merchant activities for social platforms, and purchase platform/logistics for shopping platforms.

The skincare industry is currently undergoing major reforms, with the *Cosmetics Supervision and Administration Regulations* implemented on January 1, 2021, emphasizing product efficacy verification and formula safety. Therefore, besides existing safety experiments like Bovine Corneal Opacity and Permeability (BCOP) and Hen' s Egg Test-Chorioallantoic Membrane (HET-CAM/CAMVA), brand enterprises should 正视 and actively acquire negative sentiment comment texts from skincare virtual communities regarding users' existing skin problems and product usage feedback. Considering users' focus on the connection between product formula and user experience, which constitutes purchase reasons, and that user experience is influenced by whether users belong to suitable skin types corresponding to product formulas, brand enterprises must pay sufficient attention and make corresponding formula improvements or specify applicable populations to fundamentally reduce negative comment frequency, improve product user satisfaction and brand trust, and thereby increase product sales.

The shopping platform' s inherent closeness to consumption determines that users can conveniently obtain merchant activity and price change information. Although skincare virtual community users are sensitive to product prices, they rarely mention merchant activities. Brands could address this pain point by coordinating with skincare virtual communities to add links to shopping platform official flagship stores and latest activity introductions on product detail pages, driving conversion from skincare virtual community users to shopping platform

users.

This study' s contributions include: (1) Comparing user comment texts from virtual communities, social platforms, and shopping platforms for the same product through topic clustering, keyword co-occurrence networks, topic co-occurrence networks, and sentiment analysis to identify content tendency similarities and differences; (2) Filling the research gap regarding virtual life communities; (3) Providing eight topics for the practical application value of user participation in commerce, which can also serve as measurement indicators for constructing user comment text value assessment models for deeper research on user value co-creation.

References

[1] WANG Lin, LIANG Menghua. Research on user behavior patterns in network communities based on knowledge publishing—Take Keep health community as an example[J]. Information Research, 2022(1): 1-9.

[2] YANG Ruixian, ZHANG Guangyi. Research on knowledge exchange process and mechanism in academic virtual communities[J]. Modern Information, 2020, 40(10): 52-61.

[3] CHEN Xiuxiu. Research on the impact and mechanism of virtual community interactivity on consumer brand attitudes based on Weibo comments[D]. Jiangmen: Wuyi University, 2019.

[4] TAN Chunhui, LI Yuepeng. Research on optimization strategies for virtual academic communities based on user comments and Kano model[J]. Information Studies: Theory & Application, 2021, 44(11): 108-115.

[5] ZHOU Yang, TAN Chunhui, ZHU Chenliang, et al. Research on user participation behavior in virtual academic communities based on grounded theory—Take Xiaomuchong as an example[J]. Information Science, 2022, 40(1): 176-183.

[6] CHEN H, LIU F, WEN Y, et al. Compilation and application of the scale of sustainable knowledge sharing willingness in virtual academic community during the times of the coronavirus pandemic (COVID-19)[J]. Frontiers in Psychology, 2021(12): 627833.

[7] TABISH M, BASHIR M A, ALAM M M, et al. The role of virtual community participation and engagement in building brand trust: evidence from Pakistan business schools[J]. The Journal of Asian Finance, Economics and Business, 2022, 9(3): 399-409.

[8] CARA P, BODKIN D. CHARLES. Online retail store community engagement and its impact on purchase intentions in the context of online reviews[J]. International Journal of Online Marketing, 2022, 12(1): 1-15.

[9] CHAKRABORTY U. The impact of source credible online reviews on purchase intention[J]. Journal of Research in Interactive Marketing, 2019, 13(2):

401-414.

[10] QU Huijun. The chain mediating effect of interactive vitality and trust—The influence of user participation motivation on purchase intention in Weibo virtual brand community[J]. Journal of Zhengzhou University (Philosophy and Social Sciences Edition), 2019, 52(2): 113-117.

[11] GUO Hengrui, WANG Zhongqing, ZHU Qiaoming, et al. Chinese social text event clustering method based on semi-supervised learning[J]. Journal of Chinese Information Processing, 2022, 36(2): 80-91.

[12] XIE Hong. Improved Jaccard coefficient text similarity calculation based on word frequency ratio[J]. Neijiang Science and Technology, 2021, 42(8): 27-28.

[13] WANG Chunliu, YANG Yonghui, DENG Fei, et al. A survey of text similarity calculation methods[J]. Information Science, 2019, 37(3): 158-168.

[14] QIU Xianbiao, CHEN Xiaorong. A text similarity calculation algorithm based on feature weighting[J]. Journal of Guizhou University (Natural Sciences), 2018, 35(1): 45-49.

[15] BLEI D M, NG A Y, JORDAN M I. Latent Dirichlet Allocation[J]. Journal of Machine Learning Research, 2003, 3(4/5): 993-1022.

[16] HAJJEM M, LATIRI C. Combining IR and LDA topic modeling for filtering microblogs[J]. Procedia Computer Science, 2017(112): 761-770.

[17] LI Luping, ZHAO Xiaobing. A survey of topic discovery methods based on topic models[J]. Journal of Minzu University of China (Natural Sciences Edition), 2021, 30(2): 61-70.

[18] HE Lin, MA Xiaowen, YU Xuehan, et al. Research on automatic identification and classification of event trigger verbs in classical texts: Text experiments based on Zuo Zhuan[J]. Library and Information Service, 2022, 66(5): 133-141.

[19] WANG Hui, SUN Guodao, DING Weijie, et al. Mining and visual analysis of evolutionary patterns in judicial judgment document data[J]. Journal of Computer-Aided Design & Computer Graphics, 2022, 34(3): 470-481.

[20] ZHANG Lei, TAN Huiwen, ZHANG Xuan, et al. Research on evolution and path conduction of university teacher ethics public opinion based on LDA model[J]. Information Science, 2022, 40(3): 78-86.

[21] CUI Ning, ZHAO Zongliang, WU Ruixue. Research on online product reviews based on LDA topic model and partial order set[J]. Information Research, 2021(12): 71-76.

[22] HUANG Fangyuan. Sentiment analysis of hospital evaluations using machine learning algorithms[J]. Information Technology and Informatization, 2021(2): 37-39.

[23] SHANG Yongmin, ZHAO Yuqin. Online review sentiment analysis and implementation based on machine learning[J]. Journal of Dali University, 2021,

6(12): 80-86.

[24] iMedia Data Center. 2022 China facial cleanser consumer behavior insights[EB/OL]. [2022-04-05]. https://data.iimedia.cn/31026926/31026945/44290232.

[25] 21st Century Economic Network. Weibo' s secondary listing in Hong Kong: 75% of users are post-90s, can the social "veteran"start a second spring?[EB/OL]. [2022-04-05]. http://www.21jingji.com/article/20211119/herald/55ae24eb120e9a397954803e832d4ada.html.

[26] CSDN. Elbow method + perplexity for determining the number of topics in LDA topic model[EB/OL]. [2022-04-05]. https://blog.csdn.net/weixin_{43343486}/article/details/109255165.

[27] CHI Maomao, PAN Meiyu, WANG Weijun. Cross-platform comparative study of shared accommodation and hotel user review texts: Based on LDA topic social network and sentiment analysis[J]. Library and Information Service, 2021, 65(2): 107-116.

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