Multi-Attribute Weighting-Based Keyword Extraction Method for Social Q&A Communities (Postprint)

Authors: Yu Bengong, Li Ting, Yang Ying

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

Abstract

[Purpose/Significance] Existing keyword extraction methods are ill-adapted to the characteristics of social Q&A community texts—namely short length, colloquial expression, and sparse datasets—and rarely consider the impact of user attention levels on term importance, rendering them ineffective for keyword extraction from such texts. Therefore, this paper proposes a multi-attribute weighted keyword extraction method tailored for social Q&A communities. [Method/Process] The proposed method enhances traditional TF-IDF by incorporating an adjustment function and part-of-speech features, and comprehensively measures term weights through linear weighting that fuses four user attention attributes: answer count, follow count, view count, and comment count. [Result/Conclusion] Experimental results indicate that the proposed method can more effectively extract keywords from social Q&A community texts.

Full Text

Preamble

Vol. 62 No. 5 March 2018 ChinaXiv Cooperative Journal

A Multi-attribute Weighted Keyword Extraction Method for Social Q&A Communities

Yu Bengong[1,2], Li Ting[1], Yang Ying[1]

[1] School of Management, Hefei University of Technology, Hefei 230009

[2] Key Laboratory of Process Optimization and Intelligent Decision-making, Ministry of Education, Hefei University of Technology, Hefei 230009

Abstract

[**Purpose/Significance**] Existing keyword extraction methods are ill-suited for the characteristics of social Q&A community texts, which feature short length, colloquial expression, and sparse datasets. Moreover, these methods rarely consider the influence of user attention on term importance, making them ineffective for extracting keywords from such texts. To address this gap, we propose a multi-attribute weighted keyword extraction method tailored for social Q&A communities. [**Method/Process**] The proposed method improves traditional TF-IDF by introducing a tuning function and part-of-speech weighting, and comprehensively measures term weight through linear fusion of four user attention attributes: answer count, follow count, view count, and comment count. [**Result/Conclusion**] Experiments demonstrate that this method can more effectively extract keywords from social Q&A community texts.

Keywords: social Q&A community; keyword extraction; TF-IDF; multi-attribute weighted
Classification Number: TP391
DOI: 10.13266/j.issn.0252-3116.2018.05.015

Introduction

With the development of information technology and widespread internet adoption, social Q&A communities such as Zhihu and Quora have become important channels for information exchange and knowledge sharing [1]. Social Q&A communities represent a hybrid of traditional Q&A websites and virtual communities, supporting users in self-generating content around shared interests and goals. Users serve as both beneficiaries and contributors of information resources [2]. Texts in social Q&A communities fully reflect users' knowledge domains and interests, holding significant value for natural language processing research areas including online public opinion analysis, user interest mining, and community knowledge discovery. Keyword extraction is a fundamental and core task in natural language processing that substantially impacts the effectiveness of related applications.

Currently, mainstream keyword extraction methods fall into three categories: machine learning-based methods [3-4], semantic-based methods [5], and statistics-based methods [6]. Machine learning methods automatically extract keywords by training models, but require large-scale corpora and extensive parameter tuning to ensure accuracy. Semantic methods analyze and extract keywords by constructing semantic relationship networks among terms. However, lacking standardized semantic definitions, these methods are susceptible to subjectivity and heavily depend on background knowledge bases, dictionaries, and lexicons with strict format requirements. Due to the free-form nature of user-generated content, social Q&A community texts are characterized by short length, colloquial expression, and sparse datasets [8], with rapid text updates [9], making it difficult to establish standard corpora and background knowledge bases. Consequently, these two approaches are unsuitable for keyword extraction from social Q&A community texts.

Statistics-based methods extract keywords by analyzing text features [10], with TF-IDF (Term Frequency-Inverse Document Frequency) being the most widely applied. This method is simple, universal, and imposes minimal restrictions on text length and language norms, but suffers from low accuracy [11-12]. To address this limitation, scholars have conducted extensive research. Studies show that incorporating attributes such as part-of-speech [13], term association [11,14], term position [11,15], and term span [15] into frequency analysis can effectively avoid errors in traditional keyword extraction. Furthermore, research on virtual communities has revealed that users serve as producers, disseminators, and consumers of information, with behavioral data such as browsing and replying reflecting their attention to content. These attributes should be considered when measuring term importance [12,16]. While existing studies have achieved progress in improving keyword extraction efficiency through attribute incorporation, they rarely consider applicability to social Q&A community texts and cannot effectively process such datasets. On one hand, current methods designed for Chinese texts or web pages may not suit the short, colloquial, and sparse nature of social Q&A community texts. On the other hand, compared to other data sources, social Q&A community texts have different structures and forms of user attention, requiring distinct attributes for measuring term importance. Therefore, developing a comprehensive method to measure social Q&A community text attributes and proposing a suitable keyword extraction approach remains an important research question.

2 Related Keyword Extraction Methods

2.1 TF-IDF-Based Keyword Extraction Methods

TF-IDF [17-18] is the most frequently used term weighting method. TF represents term frequency, while IDF represents the proportion of documents containing a given term. Given a document set P = {(p_j) | j = 1, 2, …, N} and the set of all terms T = {(t_i) | i = 1, 2, 3, …}, the weight calculation formula for term t_i is [16]:

W i j = t f i f × i d f i = t f i j × l o g

where n represents the number of documents in set P containing term t_i, and tf_{ij} represents the frequency of term t_i in document p_j. TF-IDF operates on two principles: terms appearing more frequently are more important; and terms appearing frequently in one document but rarely in others can effectively distinguish that document and thus hold higher importance. Although simple and effective, this method does not always achieve satisfactory results [14,19], prompting numerous improvements.

Luo Yan et al. proposed an improved TF-IDF method incorporating statistical patterns of co-frequency terms [11]. Zhang Jian'e combined TF-IDF with term association for Chinese text keyword extraction [14]. Zhang Jin improved TF-IDF by introducing term position and span for intelligence keyword extraction [15]. Luo Fanming constructed a keyword importance evaluation function integrating term deviation, position weighting, and TF-IDF [20]. Qian Aibing weighted TF-IDF with term length and position for news webpage keyword extraction [21]. Zhang Baofu adjusted TF-IDF weighting using inter-class and intra-class information distribution entropy [22].

2.2 Part-of-Speech-Based Keyword Extraction Methods

Generally, verbs, nouns, adjectives, and adverbs convey primary text information, while function words such as auxiliary words, conjunctions, and pronouns mainly serve grammatical functions and hold little value for summarizing content. Numerous scholars have investigated the impact of part-of-speech attributes on keyword extraction. Zhang Jian'e et al. statistically analyzed part-of-speech distributions in manually annotated keywords, finding that nouns, verbs, adjectives, and adverbs accounted for 95.5% of all keywords. They proposed a multi-feature fusion method combining term frequency, association, part-of-speech, and position features [14]. Yuan Jinsheng et al. developed a comprehensive weighting scheme for Chinese news webpages that assigns different weights to nouns, verbs, temporal words, locative words, adjectives, and adverbs [23]. Jiang Changjin et al. constructed a weighted formula incorporating term frequency, position, and length, assigning higher weights to nouns and noun phrases [24]. Li Xiangdong et al. applied part-of-speech and position attributes to modify term weights in an LDA generative model for extracting coarse-grained text features [25]. Lu Yonghe et al. proposed a feature weighting method influenced by part-of-speech [26]. Zhou Peng introduced centrality, part-of-speech, and position attributes for keyword extraction in microblog public opinion research [27].

2.3 User Attention Attribute-Based Keyword Extraction Methods

In virtual communities, users freely browse and comment, with these behaviors recorded as view counts, comment counts, and other metrics. User interests and concerns vary significantly, resulting in substantial differences in behavioral data across texts. Unlike literature-sourced texts, virtual community texts possess not only intrinsic term attributes like frequency and part-of-speech but also user attention attributes such as view and reply counts. Current research suggests that term importance in virtual communities depends on both term frequency and user attention levels. Huang Lucheng et al. incorporated user follow and answer counts to measure term importance in social Q&A community topic detection, quantifying user attention through statistical analysis of these metrics [8]. Liao Xiao et al. argued that term importance in enterprise virtual communities is influenced by term frequency, user view count, and reply count, combining media attention calculation methods with term frequency for comprehensive analysis [16].

Social Q&A communities are open platforms with integrated Q&A and social functions, allowing universal user participation in questioning, answering, and best answer selection. Users share knowledge through answers, track topic changes through follows, and express opinions through comments. Consequently, user attention manifests in four aspects: answer count, follow count, view count, and comment count. Due to the free-form nature of user expression, social Q&A community data exhibits short text length, colloquial expression, and sparse datasets. While statistics-based keyword extraction research has achieved results, existing methods are inadequate for social Q&A community texts due to insufficient attribute consideration or incompatibility with text characteristics.

3 Multi-Attribute Weighted Keyword Extraction for Social Q&A Communities

3.1 Keyword Extraction Method and Process

To address limitations of existing methods when applied to social Q&A communities, this paper proposes a Multi-attribute Weighted Keyword Extraction Method (MW-KEM). The workflow is shown in Figure 1 [Figure 1: see original paper]. Term frequency partially reflects term importance; assigning different weights to different part-of-speech terms helps highlight effective words and improves extraction efficiency; higher user attention indicates stronger content representation. In social Q&A communities, user attention is reflected through answer count, follow count, view count, and comment count. Therefore, this method employs six attributes as indicators—Frequency (FR), Part-of-Speech (POS), Answer count (RE), Attention count (AT), Browse count (BR), and Comment count (CO)—to comprehensively measure term importance through linear weighting. The weight W_{ij} of term t_i in document p_j is calculated as:

$$W\_\{ij\} = {}_1 \times FP\_\{ij\} + {}_2 \times RE(p\_j) + {}_3 \times AT(p\_j) + {}_4 \times BR(p\_j) + {}_5 \times CO(p\_j)$$

where FP_{ij} represents the combined weight of term frequency and part-of-speech; RE(p_j), AT(p_j), BR(p_j), and CO(p_j) represent user attention attribute weights quantified using the TF-PDF (Term Frequency-Proportional Document Frequency) [28] topic attention calculation method; and _m, m = 1, 2, …, 5 are attribute weight coefficients determined using the Analytic Hierarchy Process (AHP) [29-30].

3.2 Frequency and Part-of-Speech Weighting Based on Improved TF-IDF

Due to short text lengths in social Q&A communities, TF values are small, making traditional TF-IDF results heavily influenced by IDF and prone to two deficiencies: (1) When the number of documents n containing term t_i approaches the total document count N, the IDF value becomes very low, causing the overall weight to be too small and potentially overlooking terms that appear in multiple documents but effectively distinguish content; (2) When n approaches 0, the IDF value becomes excessively high, causing the overall weight to be too large and potentially misselecting low-frequency terms as keywords [12]. To resolve this, we introduce a power function $y = x^3$ to adjust n values, increasing TF-IDF results when n is large and decreasing them when n is small. Let $n' = a(n - N/2)^3 + b$, where a and b are constants. To ensure the function endpoints are (0, 0) and (N, N), we obtain $a = 4/N^2$ and $b = N/2$. The tuning function becomes:

$$n' = (2/N)^2(n - N/2)^3 + N/2$$

Figures 2 [Figure 2: see original paper] and 3 [Figure 3: see original paper] illustrate the tuning function and adjusted IDF function with N = 1000. Figure 3 shows that when $n < N/2$, the IDF value is smaller than traditional TF-IDF, with more pronounced changes at smaller n values; when $n > N/2$, the IDF value is larger than traditional TF-IDF, with more pronounced changes at larger n values.

To further improve extraction capability, we assign weights based on part-of-speech attributes. Typically, nouns, verbs, adjectives, and adverbs constitute the vast majority of keywords, while function words, conjunctions, and auxiliary words primarily serve grammatical functions without summarizing content. Therefore, this method assigns higher weights to nouns, verbs, adjectives, and adverbs, and zero weight to other part-of-speech terms. The part-of-speech attribute is integrated into the TF-IDF calculation as:

$$fp_{ij} = pos \times tf_{ij} \times idf_i = pos \times tf_{ij} \times \log N/n'$$

where n and tf_{ij} maintain their traditional TF-IDF meanings, and fp_{ij} represents the combined weight of term frequency and part-of-speech for term t_i in document p_j. The pos weight follows conventional assignments: 1.5 for verbs and nouns, and 1 for adjectives and adverbs. For cross-attribute comparison, we normalize the combined frequency and part-of-speech weight using the maximum and minimum fp_{ij} values in document p_j:

$$FP_{ij} = (fp_{ij} - \min(fp_{ij})) / (\max(fp_{ij}) - \min(fp_{ij}))$$

3.3 User Attention Attribute Weighting Based on TF-PDF

In social Q&A communities, user attention primarily manifests through answer count, follow count, view count, and comment count. We adopt the TF-PDF topic attention calculation method [28] to quantify these four attributes:

$$\omega_k(p_j) = a_k(p_j) / (\ a_k(p_j)) \cdot \exp(a_k(p_j) / \sqrt{(\ (a_k(p_j))^2)}),\ k = 1, 2, 3, 4$$

where $a_k(p_j)$, k = 1, 2, 3, 4 correspond to answer count, follow count, view count, and comment count for document p_j, respectively; and $\omega_k(p_j)$, k

= 1, 2, 3, 4 correspond to answer attribute weight RE(p__j), attention attribute weight AT(p__j), browse attribute weight BR(p__j), and comment attribute weight CO(p__j), respectively.

3.4 Attribute Weight Coefficient Assignment Using AHP

We employ the Analytic Hierarchy Process to determine attribute weight coefficients. AHP, proposed by Professor T.L. Saaty, is a commonly used and effective method for determining index weights, typically consisting of four steps: establishing the hierarchical structure, constructing judgment matrices, hierarchical single sorting, and hierarchical total sorting [29-30].

3.4.1 Establishing the Hierarchical Structure Model

The AHP structure model is shown in Figure 4 [Figure 4: see original paper]. The goal layer assigns appropriate weights to terms; the criterion layer comprises two major categories: term attributes and user attention attributes; the scheme layer includes specific attributes requiring weight determination.

3.4.2 Constructing Judgment Matrices

We adopt the consistent matrix method using a 1-9 scale. Five experts were invited to score the importance of each layer's indicators. The integrated expert opinions yielded judgment matrices shown in Table 1 and Table 2 .

3.4.3 Hierarchical Single Sorting and Consistency Testing

We introduce random consistency index RI values for 1-9 order judgment matrices [31-32], shown in Table 3 . Using the root method, we calculated factor weights from the judgment matrices, with results presented in Tables 1 and 2. Since 2-order judgment matrices are always perfectly consistent, consistency testing for matrix A is omitted. For matrix B2, the maximum eigenvalue λ*{max} = 4.0336, consistency index CI = (*λ{max} - n)/(n - 1) = 0.0112. From Table 3, RI = 0.96, thus CR = CI/RI = 0.0012 < 0.1. Generally, CR < 0.1 indicates satisfactory consistency, so matrix B2 passes the test.

3.4.4 Hierarchical Total Sorting and Comprehensive Consistency Testing

Based on single-level sorting results, we calculate total weights for each layer relative to the goal layer. Summarizing and normalizing weights from top to bottom yields the coefficient values shown in Table 4 .

The hierarchical total sorting consistency ratio CR = 0.0012 < 0.1, indicating satisfactory consistency. Therefore, in formula (2) for term weight calculation in individual documents, the coefficients $_1$, $_2$, $_3$, $_4$, $_5$ are 0.5, 0.102, 0.263, 0.044, and 0.092, respectively.

4 Experiments and Analysis

4.1 Experimental Method

Zhihu represents a typical domestic social Q&A community [1]. In such communities, user answers revolve around specific questions, where question content clearly summarizes the corresponding answers. Therefore, we selected the top 1000 posts from Zhihu's "automotive design" topic by comprehensive ranking. Using the Octoparse collector, we extracted text data including question content, tags, and supplements, along with numerical data including answer count, follow count, comment count, and view count. Each post's data was stored as a single document, yielding 848 documents after deduplication, totaling 94,306 characters. We performed word segmentation, stop word removal, and part-of-speech tagging using the HanLP toolkit. To further improve processing effectiveness, we expanded the HanLP dictionary with 41,645 automotive-related terms collected from platforms such as "Autohome" and "PCauto."

Following the methodology described in literature [8] for social Q&A community topic identification and analysis, we applied Huang Lucheng et al.'s keyword extraction approach to our experimental data, denoted as COM. We compared MW-KEM's performance against both COM and traditional TF-IDF. Two experimental categories were designed by controlling document quantity and extracted keyword count to evaluate the three methods under different conditions. The first category randomly selected N documents as a corpus, with each method extracting the top N/3 weighted terms as keywords. Varying N values examined performance across different corpus sizes. The second category used a fixed corpus while varying the number of extracted keywords to evaluate performance across different keyword quantities.

4.2 Results Analysis

Since no objective evaluation metrics exist for text keyword extraction, we manually annotated a reference keyword set using Zhihu question tags. Comparing machine-extracted keywords against manual annotations, we evaluated results using Precision (P), Recall (R), and F-measure (F):

$P = |A \quad B| / |A|$
$R = |A \quad B| / |B|$
$F = 2PR / (P + R)$

where A represents machine-extracted keywords and B represents manually annotated keywords.

To compare the three methods across different corpus sizes, we conducted experiments with varying N values, with results shown in Table 5 .

The results demonstrate that MW-KEM achieves higher precision, recall, and F-values than both comparison methods, indicating superior keyword extraction capability. Figure 5 [Figure 5: see original paper] plots F-value trends across document quantities. It shows that traditional TF-IDF performance degrades as corpus size increases, COM performs comparably to TF-IDF, while MW-KEM not only yields significantly higher F-values but also exhibits an upward trend with increasing corpus size, reflecting its robust performance.

To compare keyword extraction capabilities across different keyword quantities, we randomly selected corpora of 200 and 500 documents, each with 90 manually annotated reference keywords. We extracted 10, 20, …, 90 keywords using each method, with partial results in Table 6 . Figures 6 [Figure 6: see original paper] and 7 [Figure 7: see original paper] plot F-value trends. Experiments show that with small keyword quantities, all three methods perform similarly: high precision but low recall, resulting in poor comprehensive performance. As keyword quantity increases, precision decreases while recall increases, with F-values showing an upward trend. When extracted keywords exceed 20, MW-KEM's F-value consistently surpasses both TF-IDF and COM, demonstrating stronger extraction capability. Notably, MW-KEM achieves an F-value of 54.2% when extracting 90 keywords from 200 documents, and 71.2% from 500 documents, confirming that dataset size impacts effectiveness and validating MW-KEM's improving performance with increasing corpus size.

Analysis across different corpus sizes and keyword quantities confirms that introducing the tuning function, part-of-speech attributes, and user attention attributes into traditional TF-IDF effectively improves keyword extraction efficiency for social Q&A community texts. The method retains the advantages of statistics-based approaches: it requires no large corpora or parameter training, remains simple and convenient, and avoids dependence on semantic background knowledge bases, ensuring objectivity in keyword extraction.

Conclusion

Building upon statistics-based keyword extraction methods, this paper comprehensively considers attributes affecting term weights in social Q&A communities and proposes a multi-attribute weighted keyword extraction method suitable for such texts. The method linearly fuses term frequency, part-of-speech, and user attention attributes (view count, comment count, etc.), improves traditional TF-IDF through a power-function-based tuning function for quantifying frequency and part-of-speech attributes, and introduces TF-PDF for quantifying user attention attributes. Validation confirms the method's effectiveness in extracting keywords from social Q&A community texts.

Limitations include: experimental data solely from Zhihu's "automotive design" section represents a single source, and performance could be further improved on smaller datasets. Future research will focus on enhancing extraction efficiency and practical application, particularly integrating user-generated content keywords with user innovation analysis to identify hot topics, core knowledge, and knowledge domains in user innovation.

References

[1] Chen Juan, Deng Shengli. Empirical analysis of user experience factors in social Q&A platforms: A case study of Zhihu [J]. Library and Information Service, 2015, 59(24): 102-109.
[2] Yuan Hong, Zhao Juanjuan. Research on user-resource interaction in Q&A communities [J]. Library and Information Service, 2014, 58(18): 102-109.
[3] Witten IH, Paynter GW, Franke E, et al. KEA: Practical automatic keyphrase extraction [C]//Proceedings of the fourth ACM conference on Digital libraries. New York: ACM, 1999: 254-255.
[4] Horita T, Kimura F, Maeda A. Automatic keyword extraction for wikification of East Asian language documents [J]. International journal of computer theory and engineering, 2016, 8(1): 32-37.
[5] Fang Jun, Guo Lei, Wang Xiaodong. Semantic-based keyword extraction method [J]. Computer Science, 2008, 35(6): 148-151.
[6] Fei Hongxiao, Kang Songlin, Zhu Xiaojuan, et al. Research on Chinese word segmentation based on word frequency statistics [J]. Computer Engineering and Applications, 2005, 41(7): 67-68.
[7] Wang Lixia, Huai Xiaoyong. Semantic-based Chinese text keyword extraction method [J]. Computer Engineering, 2012, 38(1): 1-4.
[8] Huang Lucheng, Jiang Linshan, Miao Hong, et al. Topic identification and analysis based on network Q&A communities: A case study of "elderly" topics on Zhihu [J]. Library and Information Service, 2016, 60(5): 93-100.
[9] Chen Juan, Gao Shan, Deng Shengli. User characteristic identification and behavioral motivation analysis in social Q&A: A case study of Zhihu [J]. Information Science, 2017(5): 69-74.
[10] Fu Zhu, Wang Yuefen, Chen Bikun. Research hotspots of knowledge flow at home and abroad: Statistical analysis based on word frequency [J]. Library Science Research, 2016(14): 2-12.
[11] Luo Yan, Zhao Shuliang, Li Xiaochao, et al. Text keyword extraction method based on word frequency statistics [J]. Computer Applications, 2016, 36(3): 718-725.
[12] Chen Weihe, Liu Yun. Short Chinese text keyword extraction method based on word/phrase length and frequency [J]. Computer Science, 2016, 43(12): 50-57.
[13] Zhang Jian'e. Chinese text keyword extraction method based on multi-feature fusion [J]. Information Theory and Practice, 2013, 36(10): 105-108.
[14] Zhang Jian'e. Chinese keyword extraction method based on TFIDF and term association [J]. Information Science, 2012(10): 110-112, 123.
[15] Zhang Jin. Intelligence keyword extraction method based on improved TF-IDF [J]. Intelligence Magazine, 2014(4): 153-155.
[16] Liao Xiao, Li Zhihong, Xi Yunjiang. Modeling and analysis method for user innovation knowledge in enterprise communities based on weighted knowledge networks [J]. System Engineering Theory and Practice, 2016, 36(1): 94-105.
[17] Salton G, Buckley C. Term-weighting approaches in automatic text retrieval [J]. Information processing & management, 1988, 24(5): 513-523.

[18] Paik JH. A novel TF-IDF weighting scheme for effective ranking [C]//Proceedings of the 36th international ACM SIGIR conference on research and development in information retrieval. New York: ACM, 2013: 343-352.
[19] Shi Congying, Xu Chaojun, Yang Xiaojiang. Review of TFIDF method research [J]. Computer Applications, 2009, 29(s1): 167-170.
[20] Luo Fanming, Yang Haishen. Intelligence keyword extraction method based on statistical features in the big data era [J]. Information and Documentation Services, 2013, 34(3): 19-20.
[21] Qian Aibing, Jiang Lan. Chinese webpage keyword extraction based on improved TF-IDF: A case study of news webpages [J]. Information Theory and Practice, 2008, 31(6): 147-152.
[22] Zhang Baofu, Shi Huaji, Ma Suqin. Improved research on TFIDF text feature weighting method [J]. Computer Applications and Software, 2011, 28(2): 17-20.
[23] Yuan Jinsheng, Mao Xinwu. Chinese news webpage keyword extraction method based on combined features [J]. Computer Engineering and Applications, 2014, 50(19): 222-226.
[24] Jiang Changjin, Peng Hong, Chen Jianchao, et al. Automatic summarization based on keyword weighting and sentence features [J]. Journal of South China University of Technology (Natural Science Edition), 2010, 38(7): 50-56.
[25] Li Xiangdong, Ba Zhichao, Huang Li. A text feature selection method based on weighted LDA model and multi-granularity [J]. New Technology of Library and Information Service, 2015, 31(5): 42-49.
[26] Lu Yonghe, Wang Hongbin. Feature weighting method influenced by part-of-speech in text classification [J]. New Technology of Library and Information Service, 2015, 31(4): 18-25.
[27] Zhou Peng, Cai Shuqin, Shuang Yuanshi, et al. Microblog public opinion event content aggregation based on keyword extraction [J]. Intelligence Magazine, 2014(1): 91-96.
[28] Yeh M, Cheng W, Dai GZ. Design and implementation of online hot topic discovery model [J]. Wuhan University journal of natural sciences, 2006, 11(1): 21-26.
[29] Saaty TL. Modeling unstructured decision problems: The theory of analytical hierarchies [J]. Mathematics and computers in simulation, 1978, 20(3): 147-158.
[30] Liu Kaidi, Pang Yanjun, Zhou Shaoling, et al. Path problems in multi-criteria sorting and extension of analytic hierarchy process [J]. System Engineering Theory and Practice, 2015, 35(4): 973-979.
[31] Deng Aidong. Application of multi-level fuzzy comprehensive evaluation in library crisis management [J]. Modern Intelligence, 2008, 28(6): 117-119.
[32] Li Yaping, Jiao Jianling. Online transaction process efficiency evaluation [J]. Journal of Hefei University of Technology: Natural Science Edition, 2009, 32(8): 1204-1207.

Author Contributions

Yu Bengong: Supervised research design, provided guidance on methodology and research process, and offered revisions.
Li Ting: Designed research framework, conducted data collection and experimental analysis, and drafted and revised the manuscript.
Yang Ying: Supervised experimental research and provided ideas and suggestions for paper revision.

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