Expertise Identification and Analysis Based on Citation-IDF Weight: A Case Study in Library and Information Science - Postprint

Authors: Xiaobo Tang, Zhou Heshen, Li Shixuan, Mou Hao

Date: 2023-04-01T00:00:00+00:00

Abstract

[Purpose/Significance] Identifying expert expertise facilitates the discovery of researchers with identical or similar research directions, holding significant importance for fine-grained expert evaluation and analysis. [Method/Process] We construct an expertise seed dictionary based on academic paper keywords and employ semantic similarity computation to expand and align the dictionary; by integrating citation frequency of expertise terms, author contribution rate, and inverse document frequency of expertise terms, we propose a citation-inverse-document weight calculation method for expert expertise terms; combining expertise weight scores and rankings, we identify experts' representative research expertise and conduct expert evaluation and analysis. [Results/Conclusion] Experimental validation demonstrates that the expert expertise identification method proposed in this study can objectively reflect the influence of expert expertise, while holding practical reference value in related domains such as fine-grained expert assessment, expert recommendation, and disciplinary hotspot analysis.

Full Text

Preamble

Expertise Identification and Analysis Based on Cited-Inverse Document Weight: A Case Study in Library and Information Science

Tang Xiaobo[1,2], Zhou Heshen[1], Li Shixuan[3], Mou Hao[4]

[1]School of Information Management, Wuhan University, Wuhan 430072
[2]Center for Studies of Information System, Wuhan University, Wuhan 430072
[3]School of Safety Science and Emergency Management, Wuhan University of

Technology, Wuhan 430070
[4]State Grid Sichuan Electric Power Company, Chengdu 610000

Abstract: [Purpose/Significance] Identifying expert expertise helps discover researchers with similar or related research directions and is crucial for conducting fine-grained expert evaluation and analysis. [Method/Process] This study constructs an expertise seed dictionary based on academic paper keywords and employs semantic similarity calculation for dictionary expansion and alignment. By integrating citation frequency of expertise terms, author contribution rates, and inverse document frequency of expertise terms, we propose a cited-inverse document weight calculation method for expert expertise terms. Combined with expertise weight scores and rankings, we identify representative research expertise of experts and conduct expert evaluation and analysis. [Result/Conclusion] Experimental validation demonstrates that the proposed expert expertise identification method can objectively reflect the influence of expert expertise and offers practical reference value for fine-grained expert evaluation, expert recommendation, and disciplinary hotspot analysis.

Keywords: informetrics, semantic mining, expertise identification, expert evaluation
Classification Number: G250
DOI: 10.13266/j.issn.0252-3116.2021.15.013

In October 2020, the General Office of the State Council issued the "Overall Plan for Deepening the Reform of Education Evaluation in the New Era," emphasizing the importance of research evaluation for university faculty and proposing to adhere to classified evaluation based on different disciplines and positions, promote representative achievement evaluation, explore long-cycle evaluation, improve peer expert review mechanisms, and focus on combining individual evaluation with team evaluation. However, with the continuous emergence of emerging and interdisciplinary fields, the characteristics of diverse information resources and research achievements—large quantity, multiple types, and rapid updates—make traditional informetric methods unable to meet the needs of scientific talent evaluation in the new era. Therefore, addressing the integration of philosophy and social sciences and establishing a fine-grained scientific talent evaluation management system to enhance diversified scientific research team construction and optimize the utilization of disciplinary resources has become an urgent problem to be solved.

With the proposal of the "small peer" concept, researchers have begun to conduct fine-grained evaluation and analysis of experts with the same or similar research directions within disciplinary fields. Identifying expert expertise can discover "small peer" expert groups and support expert selection and multidimensional expert evaluation and analysis work. Statistical methods are the most common approaches for expert expertise identification. For instance, Li Gang et al. extracted expert expertise based on word frequency and conducted clustering and visual analysis of similar research experts in China's library, information, and archival management field. Considering document location, Tang Xiaobo et al. constructed achievement characteristics for doctor profiles by counting keywords in academic achievements. Liu Xiaoyu et al. used keywords as candidate expertise terms, extracted author-keyword matrices, and combined TF-IDF weighting to build expert expertise. Some researchers have conducted expertise identification based on network analysis methods. For example, Zhu Weizhu et al. constructed a conceptual knowledge network based on word frequency analysis and used K-core hierarchy theory to divide the hierarchical structure of disciplinary fields and their research subgroups. Liu Ping and Zhou Menghuan proposed an expert expertise identification method based on co-word networks. Chen Gang et al. combined TextRank and concept linking technology to identify candidate expertise terms representing expert expertise and used the Analytic Hierarchy Process to assign weights to expertise terms based on author order and citation counts. Additionally, some researchers have identified expert expertise based on topic analysis. For example, Zhang Xiaojuan et al. used PLSA to model topics for each expert's paper output and analyze research fields of experts in library and information science. Chen Hongling et al. combined Word2vec word vector models with LDA topic models to construct expert features and identify academic communities.

Currently, expert expertise identification methods are relatively limited. Most researchers use statistical term frequency methods to build expert expertise labels and introduce certain subjective factors in term weight calculation. Expertise identification methods based on domain knowledge bases require collecting expert knowledge for domain ontology construction, while expertise identification methods based on topic analysis have poor interpretability. Current research on expert expertise identification mostly extracts representative expertise from relevant texts or network relationships of expert research achievements, ignoring factors such as the impact of achievements on the disciplinary field and the contribution size of experts in the achievements.

Therefore, this study proposes an expert expertise identification method based on cited-inverse document weight. We combine paper keywords with word vector models to automatically construct an expert expertise term dictionary. By integrating author contribution rate, citation frequency, and inverse document frequency of expertise terms, we propose an expertise term weight calculation method. We calculate and rank expert expertise weight scores to extract representative expertise labels of experts. The proposed method can objectively extract representative expertise of experts by considering factors such as the scale of researchers in related fields and the influence of experts in related fields, which is of great practical significance for expert evaluation, expert recommendation, and disciplinary hotspot analysis.

2 Related Research

2.1 Expert Academic Evaluation Research

Scholars have conducted multi-faceted explorations on expert evaluation. Traditional studies mainly evaluate scientific talents through bibliographic analysis and citation analysis. Classic expert evaluation methods include the h-index and p-index, which primarily construct expert evaluation indicators through the number of papers published and citations within a certain period. Some researchers have also optimized evaluation indicators and constructed derived expert evaluation indices from aspects such as paper count, author order, and publication time. However, Liu Zhongxing and Yang Jianlin pointed out that the use of personal academic evaluation indicators in China's library and information science field is still in the development stage, with scholars mainly focusing on h-index related indicators, while research on the integration of multiple indicators for comprehensive personal academic evaluation is relatively scarce, and academic evaluation research including personal academic evaluation still needs further improvement.

In recent years, social network analysis, topic analysis, and expert knowledge mapping have gradually become common methods for expert evaluation and analysis in disciplinary fields. Additionally, some researchers have constructed expert knowledge graphs for expert evaluation and analysis, including expert knowledge graphs based on cooperative relationships, document content analysis, link analysis, and comprehensive content and link analysis. However, current research on fine-grained expert evaluation and analysis is still relatively limited. Due to factors such as differences in disciplines or research directions, evaluating expert influence solely based on paper counts has limitations. Meanwhile, in expert evaluation related research, scholars usually select some experts in specific fields for analysis, and their research methods cannot conduct fine-grained influence evaluation on massive numbers of experts and scholars.

2.2 Keyword Extraction and Expertise Dictionary Construction

Expertise representation methods based on domain knowledge bases can accurately describe expert expertise. To construct an expertise dictionary that reflects domain knowledge, it is necessary to extract terms from research achievements that can reflect and distinguish research topics. A common expertise dictionary construction method is to use keywords provided by authors in papers. For example, Fan Xiaoyu et al. used literature keywords published by researchers to construct research topic and interest tags for experts. Some researchers have built expertise dictionaries by mining keywords from paper abstracts through statistical methods. For instance, Mao Jin et al. selected high-frequency nouns from expert research achievements to represent expert research expertise. Chen Gang et al. combined word co-occurrence networks with TextRank to form academic expertise candidate words. With the development of natural language processing, some researchers have studied how to identify keywords from academic paper abstracts and main texts, introducing word vector models and deep learning models into scientific keyword extraction tasks. Additionally, domain knowledge bases have attracted scholars' attention. For example, Lu Wei et al. combined the Chinese Library Classification with the Management Science Thesaurus to construct an expertise dictionary for library and information science experts, mapping research achievements of different experts. Hu Yuehong and Liu Ping extracted domain terms from academic papers and mined relationships between terms based on association rules and formal concept analysis to construct an informatics domain ontology.

Methods for constructing term dictionaries based on expert knowledge and domain ontology not only require massive expert domain knowledge but also often have lag when dealing with emerging research hotspots. While automatically constructing term dictionaries through algorithms like TextRank or natural language processing methods can reduce manual annotation costs for expertise ontology, they also bring problems such as low interpretability and inability to effectively represent relationships between words.

2.3 Author Order and Contribution Research

Against the background of disciplinary integration and interdisciplinarity, more and more experts tend to conduct research through collaboration. Different author orders can directly reflect the contribution size of experts. As shown in Figure 1 [Figure 1: see original paper], after conducting statistical analysis on authors of more than 50,000 papers published in the library, information, and documentation field, this study found that the number of single-author papers shows a decreasing trend. Author order is often correlated with the contribution size of experts in research, which also brings about the problem of allocating expert contribution proportions in scientific research achievements. Ding Jingda et al. proposed evaluating a expert's academic influence in a field by calculating the total citation frequency based on research contribution rates. This study adopts the author contribution rate allocation formula proposed by N.T. Hagen to calculate expert contribution in papers, applying the expert author order and contribution rate calculation method to expert expertise word weight calculation, thereby allocating citation frequency representing paper influence according to contribution rate to highlight important contributors and reflect the research influence of important authors in the research field. As shown in Formula (1):

where j represents the author's signature order, and m represents the number of authors in the paper.

3 Expertise Identification Model Framework Based on Cited-Inverse Document Weight

Extracting research expertise recognized by the research field from research achievements is a prerequisite for conducting fine-grained expert evaluation and analysis. This paper combines author contribution rate, citation frequency, and inverse document frequency of expertise terms to construct an expert expertise identification model based on cited-inverse document weight, as shown in Figure 2 [Figure 2: see original paper]. This framework mainly includes three parts: data preprocessing, expertise term dictionary construction, and expert expertise representation.

3.1 Data Preprocessing

To ensure data completeness, this study integrates Chinese journal paper data collected from multiple platforms during the data preprocessing stage and extracts standardized academic paper data for further analysis. The data preprocessing workflow in this paper mainly includes: (1) Data acquisition. Based on CNKI and Wanfang databases, we export metadata of target journal papers and use Selenium to build a crawler to collect paper citation data. (2) Data cleaning and filtering. Data cleaning mainly standardizes paper data from different databases. After merging data, we filter out samples with overly short titles/abstracts, empty author fields, and duplicates, and define rules to remove relevant records such as notifications and submission information.

3.2 Expertise Term Dictionary Construction

Keywords are words that highly condense and summarize paper content and can well reflect experts' research directions and capabilities. This study uses papers published in the field in the past 10 years as the research object, constructs an expertise seed dictionary from text keywords, and introduces the seed dictionary as an external dictionary into the segmentation tool. After preprocessing such as segmentation and stop word removal of titles and abstracts, we construct a Word2vec word vector model. We extract high-frequency words from paper titles and abstracts as expansion candidate words and conduct semantic similarity comparison based on the word vector model. We select expansion candidate words with high similarity to the seed dictionary to establish a keyword-expansion candidate word synonym table. In subsequent natural language processing, we use the synonym table to transform synonymous expansion candidate words with different forms in the text into standardized keywords. Simultaneously, we select candidate words with low similarity to keywords in the seed dictionary to build an expertise expansion dictionary, identifying high-frequency words with different meanings from keywords in the seed dictionary, and manually filter out words that cannot effectively reflect expert research directions and capabilities. Finally, we integrate the seed dictionary and expertise expansion dictionary to obtain a semantically expanded expertise dictionary containing 8,027 words.

3.3 Expert Expertise Representation

Expert expertise representation includes expertise word extraction and expertise word weight calculation. In the expertise word extraction part, we use the synonym table to transform high-frequency words in the original text into standardized expertise terms, fuse the expertise seed dictionary with the expertise expansion dictionary, and then use this dictionary to annotate expertise terms in the paper dataset. Finally, we extract each expert's expertise words and related paper information.

In the expertise weight calculation part, this study uses the citation count of papers where expertise terms appear as one of the main factors to objectively measure expertise weight scores based on the influence size generated by experts in related fields. Since the training corpus scale of word vector models has limitations, some words introduced during the semantic expansion stage cannot effectively reflect expert expertise. Meanwhile, inverse document frequency can reflect whether words have good category discrimination ability. Therefore, this paper introduces inverse document frequency into expertise word weight, as shown in Formula (2). By calculating the inverse document frequency of expertise terms in the paper dataset, we can filter out common words that cannot characterize paper research content on one hand, and consider the scale of related research fields as a factor on the other hand, avoiding the convergence of research content among domain experts and promoting the common development of multiple research directions. Additionally, introducing author contribution rate factors based on expert author order in weight score calculation can effectively highlight important researchers in related fields. In summary, this study proposes an expert expertise word weight calculation method, as shown in Formula (3). We select expertise words with more than 10 researchers in the field and rank them according to expertise word weight scores to finally obtain experts' representative expertise and weight scores.

where M represents the total number of papers, w represents the expertise term, m_w represents the number of papers containing w, and IDF_w represents the inverse document frequency of expertise term w. i represents the ith paper among n papers of the expert, and j represents the jth expert. Score_{jw} represents the weight score of expert j on expertise term w, D_{ij} represents the contribution degree of expert j in the ith paper, and cite_{ji} represents the citation count of the ith paper of expert j.

4 Experiments and Results Analysis

4.1 Data Collection

This study takes 20 journals in the library, information, and documentation field from the Nanjing University CSSCI source Chinese journal catalog (2019-2020) as the research object, collects academic paper metadata through CNKI, and supplements data with Wanfang. We collected information on 54,698 papers published between January 1, 2010 and April 25, 2020. The collected fields include source database, title, author, affiliation, literature source, keywords, abstract, publication time, first responsible person, funding, year, volume, issue, page numbers, classification number, and citation count. Metadata is mainly exported through data services provided by CNKI and Wanfang, and citation counts are collected using a Selenium-based crawler. During data preprocessing, we integrate paper data from Wanfang and CNKI databases, remove samples with overly short titles/abstracts and empty author fields, remove relevant records such as notifications and submission information, and merge duplicate records, finally obtaining 49,399 articles.

4.2 Experimental Process

To mine terms that can describe expert expertise, this paper constructs an expertise seed dictionary using keywords with word frequency greater than 3 in the paper dataset, totaling 7,990 words. We import the expertise seed dictionary into the jieba segmentation tool's external dictionary, preprocess the titles and abstracts of the paper dataset through segmentation and stop word removal, and set parameters to train a Word2Vec word vector model with dimension 100, context window size 5, and minimum word frequency 3. We extract high-frequency words with word frequency greater than 100 from titles and abstracts as expansion candidate words and conduct semantic similarity comparison between high-frequency words and keywords in the expertise seed dictionary based on the word vector model. If a high-frequency word can find a keyword with similarity greater than 0.9 from the expertise seed dictionary, we select the most similar keyword to build a keyword-expansion candidate word synonym table (see Table 1), establishing 94 keyword-expansion candidate word mapping pairs. If the similarity between a high-frequency word and all keywords in the expertise seed dictionary is below 0.6, we include the high-frequency word in the expertise expansion dictionary, delete meaningless words such as "within" and "two types," and finally construct an expertise expansion dictionary containing 37 words such as "core," "background," and "novel." Finally, we transform high-frequency words in papers into standardized expressions through the keyword-expansion candidate word synonym table, fuse the expertise seed dictionary with the expertise expansion dictionary, and build a semantically expanded expertise dictionary containing 8,027 words.

We preprocess the paper dataset by segmenting titles and abstracts, removing stop words, transforming some high-frequency words into standardized keywords through the synonym table, and concatenating the processed titles, abstracts, and article keywords to build the paper's word list. We retain expertise terms in the paper text through the semantically expanded expertise dictionary and select words that can well reflect expert expertise. Finally, we calculate the inverse document frequency of expertise terms in the preprocessed paper dataset. Simultaneously, we extract each expert's relevant author order information, paper citation counts, and calculate expert contribution rates in papers based on author order. We use Formula 3 to calculate expert expertise term weight scores and rank expert expertise according to weight scores to obtain experts' representative research expertise.

4.3 Results Analysis

To verify the effectiveness of the proposed expert expertise identification method based on cited-inverse document weight, this paper conducts three parts of empirical analysis: First, we compare the identification effect of the proposed method with the TF-IDF method; second, we extract representative expertise of multiple experts and conduct authoritative researcher analysis for specific research expertise and evaluation of expertise influence for experts at different research stages; finally, we select high h-index experts in the research field to extract their representative expertise for research team hot topic analysis.

4.3.1 Comparative Analysis of Expert Expertise Identification The TF-IDF algorithm is one of the commonly used expert expertise identification methods, consisting of term frequency and inverse document frequency, which considers the importance of keywords to documents and category discrimination ability. This study integrates each expert's related paper information, uses the semantically expanded expertise dictionary to build a TF-IDF matrix of expert keywords, and compares the top 10 expert expertise extracted by both methods using Qiu Junping as an example, as shown in Table 2 .

The results show that the cited-inverse document weight method identifies "CiteSpaceII" as Qiu Junping's highest-scoring expertise, while the TF-IDF method identifies "Five Metrics" as his most representative research expertise. Analysis of related research achievements shows that Qiu Junping published 4 papers related to "Five Metrics," mainly concentrated in 2019, with only 18 experts in this concept's related research. The cited-inverse document weight method selects Qiu Junping's highly cited research achievements to build his representative expertise labels and comprehensively considers the scale of researchers in different expertise terms to select expertise terms, such as in research related to "CiteSpaceII," the highest citation is 249 times, and research related to "disciplinary knowledge diffusion" is cited 48 and 40 times respectively. The TF-IDF method only considers the quantity of related research content and researcher scale when extracting expertise, making it easy to select expertise words in research content with smaller researcher scale. The average citation count of papers used by the proposed method is much higher than that of the TF-IDF method. Therefore, this paper believes that expertise extracted based on the cited-inverse document weight method can reflect the representative research directions recognized by peers and can mine newer and highly recognized research themes in the research field, which is important for promoting the common development of multiple research directions in the discipline.

To verify the effectiveness of the cited-inverse document weight method, this study randomly selected 100 experts with more than 3 publications, extracted the highest-scoring expertise of each expert using both methods, and conducted visual analysis of related papers of this expertise, as shown in Figure 3 [Figure 3: see original paper]. The papers used by the cited-inverse document weight method to extract expert expertise have a total of 132 papers, with an average citation count of 17.72 times per paper, while the TF-IDF method uses 155 papers to extract expert expertise, with an average citation count of 8.66 times per paper.

4.3.2 Expert Expertise Evaluation The expert expertise identification method proposed in this paper can conduct expert evaluation and analysis from multiple dimensions. By calculating and ranking expertise weight scores of domain researchers, we can mine authoritative experts in research fields or evaluate experts' research influence in the field. Taking "big data" related research as an example, we select papers with "big data" in the title, keywords, or abstract as the research object, count expert paper information in the field, and calculate expertise weight scores, as shown in Table 3 .

The results show that Han Cuifeng only has two big data research papers but obtained the highest expertise weight score. Analysis shows that his two papers were cited 314 and 119 times respectively, with first-author attribution. Meanwhile, although Su Xinning has a lower average citation count per paper, he has three sole-author papers in big data research, with the highest citation being 221 times, while Li Guangjian's two first-author papers with citations of 178 and 165 times have co-authors, making Su Xinning's score in the big data field relatively higher. It can be seen that the expert expertise weight calculation method proposed in this study has a strong tendency toward highly cited articles and is sensitive to author order.

Furthermore, analyzing experts' representative expertise and expertise weights can effectively evaluate their academic influence. Based on rankings of high-productivity authors and highly cited authors, high-productivity young authors and highly cited young authors in the CSSCI informatics field during China's "12th Five-Year Plan" period, we extract experts' representative research expertise by weight score and build radar charts while showing their influence rankings in this expertise. The final results are shown in Figure 4 [Figure 4: see original paper] and Figure 5 [Figure 5: see original paper]. Through comparative analysis of experts at different research stages, we find that high-productivity and highly cited research experts often have deep academic accumulation in multiple research directions, while young research experts can also achieve excellent results in major research directions through their research accumulation. The expert expertise identification method proposed in this study comprehensively considers the contribution size of experts in expertise fields and selects representative expertise for experts based on the scale of expertise term research fields, which can intuitively reflect the influence of expert research in disciplinary fields and is conducive to promoting individual achievement construction of experts and supporting multi-dimensional expert evaluation work.

4.3.3 High h-Index Expert Research Theme Analysis Citation behavior to some extent reflects the recognition of article content and direction by the disciplinary field. High h-index experts have both high publication volume and article citations. Analyzing the research content of high h-index experts can help understand hot topics in the field. This study calculates the h-index of experts based on academic papers published in the Nanjing University CSSCI source journal catalog in library, information, and documentation fields between January 1, 2010 and April 25, 2020, and analyzes high h-index experts. Taking experts with h-index scores greater than or equal to 20 as the research object, we identify the representative expertise of the above experts based on cited-inverse document weight, retain expertise words with higher weight scores among synonyms representing expert expertise, and finally obtain the representative expertise and weights of high h-index experts in the field, as shown in Table 4 .

Analysis shows that the main research areas of high h-index experts in library, information, and documentation fields include informetrics, government data openness, emergencies and emergency response, user behavior research, social media research, data analysis and knowledge discovery, and library management and analysis. Among them, informetrics and library management and analysis have obtained higher weight scores. The h-index cannot reflect experts' contributions in specific research directions in expert evaluation work and still requires manual screening of evaluation objects and research data to conduct expert evaluation and analysis work for specific research directions. The expertise identification method proposed in this study is an effective supplement to expert evaluation research and can enrich expert analysis and evaluation work from the influence generated by experts in each research direction.

5 Conclusion

This paper constructs a dictionary describing expert expertise based on word vector models and integrates citation frequency of expertise terms, author contribution rate, and inverse document frequency calculation formulas to propose an expert expertise identification method based on cited-inverse document weight. This method can extract expert expertise based on representative research achievements of experts, consider factors such as researcher scale and paper influence, and enrich existing expert expertise identification methods from the dimension of disciplinary field influence. Additionally, this method can mine authoritative experts for specific expertise, conduct fine-grained expert evaluation, and analyze disciplinary field hotspots. Experimental results preliminarily verify the effectiveness of the proposed expert expertise identification method and provide a new perspective for expert evaluation and disciplinary analysis.

However, the expert expertise identification method constructed in this paper still has certain limitations. For example, the dataset only uses papers from the Nanjing University CSSCI source Chinese journal catalog, and does not distinguish between different types of papers such as reviews and empirical studies in the data selection process. Additionally, constructing an expertise dictionary through paper keywords and semantic expansion of word vector models cannot effectively reveal the relationship between disciplines and terms, cannot effectively distinguish between terms describing research topics and research methods, and some fine-grained expertise terms still require expert knowledge for parsing to better describe expert expertise. Therefore, how to integrate disciplinary domain knowledge ontology, further optimize the expert expertise identification method, and construct a more comprehensive and systematic fine-grained expert evaluation model remains to be further explored.

References

[1] General Office of the State Council. Overall Plan for Deepening the Reform of Education Evaluation in the New Era [EB/OL]. [2021-01-15]. http://www.gov.cn/zhengce/2020-10/13/content_{5551032}.htm.
[2] Li Gang, Yu Yifei, Du Wen. Research on "Small Peer" Groups Among LIS Faculty (2001-2010) [J]. Library and Information Knowledge, 2011(6): 78-85.
[3] Tang Xiaobo, Gao Hexuan. Construction of Doctor Profiles Based on Feature Analysis and Tag Extraction [J]. Information Science, 2020, 38(5): 3-10.
[4] Liu Xiaoyu, Zhu Donghua, Wang Xuefeng, et al. Research on Multi-Expertise Expert Identification Method: A Case Study in Big Data Field [J]. Library and Information Service, 2018, 62(3): 55-63.
[5] Zhu Weizhu, Li Chunfa. Empirical Research on "Small Peer" Review Expert Selection Method Based on Conceptual Knowledge Network [J]. Journal of Intelligence, 2017, 36(7): 78-83, 88.
[6] Liu Ping, Zhou Menghuan. Expertise Mining Based on Co-word Networks [J]. Information Science, 2012, 30(12): 1815-1819.
[7] Chen Gang, Li Nan, Liang Bing, et al. Method for Identifying Scholar Academic Expertise Based on Achievement Characteristics [J]. Library and Information Service, 2019, 63(20): 96-103.
[8] Zhang Xiaojuan, Lu Wei, Cheng Qikai. Application of PLSA in Expertise Identification in Library and Information Science [J]. New Technology of Library and Information Service, 2012(2): 76-81.
[9] Chen Hongling, Yang Jiaying, Xu Xin. Academic Community Identification Based on Bibliographic Abstract Semantic Modeling: A Case Study of Domestic LIS Scholars [J]. Information Studies: Theory & Application, 2020, 43(5): 170-176.
[10] Qiu Junping. Introduction to Informetrics [M]. Wuhan: Wuhan University Press, 2019.
[11] HIRSCH JE. An index to quantify an individual's scientific research output [J]. Proceedings of the national academy of sciences, 2005, 102(46): 16569-16572.
[12] PRATHAP G. The 100 most prolific economists using the p-index [J]. Scientometrics, 2010, 84(1): 167-172.
[13] Yan M, Yu Z, Zhang Y, et al. An expert recommendation approach combining project correlation and professional ability [C]// 2015 12th international conference on fuzzy systems and knowledge discovery. Zhangjiajie: IEEE, 2015.

[14] Tang Puni. Application Research of p_r(y) Index and h_r(y) Index in Dynamic Evaluation of Scholar Academic Influence: A Case Study in Library and Information Science [J]. Information Studies: Theory & Application, 2020, 43(12): 63-67, 41.
[15] Xie Ruixia, Li Xiuxia, Han Xia, et al. Construction of Author Influence Evaluation Indicators Based on Weighted Citation Frequency and Author Order [J]. Information Science, 2018, 36(8): 90-93, 148.
[16] Liu Zhongxing, Yang Jianlin. Research on the Application of Personal Academic Evaluation Indicators in China's Library and Information Science Field [J]. Modern Information, 2020, 40(12): 140-149.
[17] Lü Penghui, Zhang Ling. Research on Disciplinary Knowledge Networks (II): Structure, Characteristics and Evolution of Co-citation Networks [J]. Journal of the China Society for Scientific and Technical Information, 2014, 33(4): 349-357.
[18] Lü Penghui, Liu Shengbo. Empirical Research on Disciplinary Knowledge Networks (IV): Structure and Characteristic Analysis of Collaboration Networks [J]. Journal of the China Society for Scientific and Technical Information, 2014, 33(4): 367-374.
[19] Hu Yuanjiao, Wang Hao. Construction and Analysis of Scholar Knowledge Maps for CSSCI [J]. New Technology of Library and Information Service, 2011(3): 38-43.
[20] Liu Kan, Zhou Lihong. Research on Expert Knowledge Maps [J]. Information and Documentation Services, 2012, 33(2): 18-22.
[21] Pan Youneng, He Huanzhen. Research on Expert Knowledge Map Construction Based on Co-authorship and Citation Weighting [J]. Journal of Intelligence, 2018, 037(8): 128-132.
[22] Fan Xiaoyu, Dou Yongxiang, Zhao Pengwei, et al. Research on Construction Method of Researcher Profiles Integrating Multi-source Data [J]. Library and Information Service, 2018, 62(15): 31-40.
[23] Mao Jin, Li Gang. A Method for Constructing Research Field Expert Maps Based on OKM [J]. Library and Information Service, 2014(14): 34-40.
[24] Xu Wenhai, Wen Youkui. A Chinese Keyword Extraction Algorithm Based on TF-IDF Method [J]. Information Studies: Theory & Application, 2008(02): 298-302.
[25] ZHU X, LYU C, JI D, et al. Deep neural model with self-training for scientific keyphrase extraction [J]. Plos one, 2020, 15(5): e0232547.
[26] Lu Wei, Liu Jie, Qin Xiyan. Expert Retrieval and Evaluation in Library and Information Science Based on Expertise Vocabulary [J]. Journal of Library Science in China, 2010, 36(2): 70-76.
[27] Hu Yuehong, Liu Ping. Research on Expertise Representation Based on Ontology Concepts [J]. Library and Information Service, 2012, 56(4): 17-40.
[28] Cui Linwei, Lu Ying. Analysis of Author Contribution Elements Based on Author Order: Taking Author Contribution Statements in Library and Information Service 2015-2016 as Examples [J]. Library and Information Service, 2017, 61(9): 80-86.
[29] RAHMAN M T, REGENSTEIN JM, KASSIM NLA, et al. The need to

quantify authors relative intellectual contributions in a multi-author paper [J]. Journal of informetrics, 2017, 11(1): 275-281.
[30] Ding Jingda, Wang Xinming. Method for Measuring Co-author Contribution Rate Based on Author Contribution Statements [J]. Library and Information Service, 2019, 63(16): 95-102.
[31] HAGEN NT. Harmonic allocation of authorship credit: source-level correction of bibliometric bias assures accurate publication and citation analysis [J]. Plos one, 2008, 3(12): e4021.
[32] ROBERTSON S. Understanding inverse document frequency: on theoretical arguments for IDF [J]. Journal of documentation, 2004, 60(5): 503-520.
[33] Tan Chunhui, Zeng Juan, Qiu Junping. Analysis of Domestic Informatics Research Trends During China's "12th Five-Year Plan" Period Based on CSSCI [J]. Journal of the China Society for Scientific and Technical Information, 2017, 36(7): 734-747.

Author Contributions

Tang Xiaobo: Overall paper concept, framework structure design, and revision;
Zhou Heshen: Experiments, paper content organization, and writing;
Li Shixuan: Research design and evaluation method improvement;
Mou Hao: Paper revision and experimental evaluation.

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

Source: ChinaXiv — Machine translation. Verify with original.