

A Summary of the Research on the Discovery Method of Interdisciplinary Topics

Wei Xu

School of Economic & Management, Xidian University, Xi'an 710071, China.

evril2v@foxmail.com

Abstract

General Secretary Xi encouraged the professional development of interdisciplinary and related disciplines in the "Speech at the Symposium of Scientists". The current class is facing a series of major problems, and a single subject area has been unable to solve them. The knowledge of multiple disciplines is showing a trend of mutual integration, and it is found that the subject of interdisciplinary research is the starting point to promote the development of interdisciplinary. First, clarify the related concepts and classifications of interdisciplinary. Secondly, the discovery methods of domestic and foreign documents are summarized into document content perspective, author collaboration perspective, reference citation perspective, and multiple comprehensive perspectives. Finally, analyze the principles, advantages and disadvantages of the above perspectives. The method of integrating the three perspectives can make full use of the differences of each perspective, and complement each other with complementary advantages.

Keywords

Interdisciplinary; Theme Discovery; Document Content; Author Collaboration; Reference Citation.

1. Introduction

In August 2019, six departments including the Ministry of Science and Technology jointly issued a document [1], proposing guidance on following the laws of scientific research activities, talent growth, and transformation of results, transforming the government's science and technology management functions, focusing on planning, and focusing on services. In September 2020, President Xi Jinping proposed at the symposium of scientists [2] to encourage qualified universities to actively set up basic research, interdisciplinary and related disciplines. In December 2020, the Academic Degrees Committee of the State Council [3] decided to set up the category of "interdisciplinary" (the category code is "14"). The next month, when the person in charge talked about the significance of this move [4], he mentioned that interdisciplinary integration is a major feature of the current development of science and technology, an important source of the emergence of new disciplines, an effective path for cultivating innovative talents, and an economic society. For the inherent needs of development, the Party Central Committee and the State Council attach great importance to interdisciplinary development.

In recent years, due to a series of major problems faced by mankind, a single subject area has been unable to solve major and emerging problems [5]. It requires the integration and interaction of multiple subject knowledge, and new subject branches and new growth points continue to emerge. Interdisciplinary research is used as a means to deal with complex issues in various cognitive and social phenomena [5]. For example, the Nobel Prize was established in the early 19th century, and the proportion of winners belonging to interdisciplinary subjects was only 36.23%; by the end of the 19th century, the ratio reached 47.37% [6]. However, the

current mainstream knowledge organization method is organized according to the classification of disciplines. If the scientific research personnel under the background of each discipline are limited to deep work in their own fields, it will deviate from the actual needs of current scientific and technological development. Economic and social development has more urgent needs for high-level innovative, compound, and application-oriented talents [4], interdisciplinary cooperation is becoming more and more extensive, and the tasks of scientists have changed from solving simple problems oriented to a single discipline to solving interdisciplinary oriented The complex problem [7].

With the new round of scientific and technological revolution and the accelerated evolution of industrial changes, major innovations have the characteristics of interdisciplinary and technological integration. Because there are only a few major innovations based on new knowledge, more of them are the reintegration and innovative application of existing knowledge. Interdisciplinary integration is an important source of promoting major innovations [8]. However, it is difficult for most experts in the field to see that technologies outside the field have been cross-fused with the field [9]. As a result, interdisciplinary research has become an important breakthrough in the prediction of major innovations, and the discovery of interdisciplinary research topics is the starting point for promoting the development of interdisciplinary.

At present, scholars are committed to the discovery of interdisciplinary topics, mainly using the results of scientific researchers to conduct quantitative and qualitative analysis. After careful study of topic discovery methods, this article first discusses related definitions and classifications, and summarizes the discovery methods into document content perspective, author collaboration perspective, reference citation perspective, and multiple comprehensive perspectives.

2. Definition and Classification

The concept of interdisciplinary is not unified yet. It was first publicly proposed in 1926 by R.S. Woodworth of Columbia University in the United States [10]. He believed that this was a practical activity involving two or more disciplines that went beyond the boundaries of a known discipline. Klein [11] believes that "disciplinarity is a comprehensive scientific activity that is accompanied by the development needs of society and disciplines." Rafols [5] believe that "disciplinarity is essentially the mutual knowledge of different disciplines. Assimilation and integration are interoperable." The concept similar to interdisciplinary is interdisciplinary, and both reveal the process of dynamic integration of multiple disciplines. Considering the above definitions, this article believes that interdisciplinary is a comprehensive scientific practice activity with researchers as the main body and accompanied by the flow of knowledge between different disciplines.

Interdisciplinary research can be divided into two levels, namely the macro-intersection situation and the micro-intersection theme [9]. The detection of the interdisciplinary situation helps to understand the characteristics of knowledge absorption and diffusion of disciplines, promote interdisciplinary cooperation and sharing, and promote the integration and development of disciplines, so as to better cope with the challenges faced in the development of interdisciplinary. Cross-topics are an important hub for the convergence and fusion of knowledge from different disciplines. Its effective identification helps to detect emerging areas of science and technology, assists in scientific decision-making, and early plans to form forward-looking research. This article mainly conducts a microscopic research review of interdisciplinary research, that is, the study of interdisciplinary topics. This article believes that the subject of interdisciplinary is the common research content formed by multiple disciplines in the process of knowledge flow, and it often contains the frontiers of disciplinary development and new growth points.

3. Document Content Perspective

Based on the principle of spreading or merging knowledge of different disciplines through the medium of text content, it fully considers the potential semantic features of the text. Document content information mainly includes the title of the paper, abstract, research topic, body content, citation content, etc. [12]. The target document itself contains text content information, subject information or the subject classification number information of the thesis determined according to the "Chinese Library Classification". One of the more representative methods is keyword co-word frequency analysis [13]: Since the keywords of journal articles are natural language vocabularies that express the topics of journal articles, statistical analysis and clustering of keywords are used to analyze their differences in different disciplines. In order to achieve the purpose of discovering interdisciplinary topics. This method can go deep into the micro text structure, and the recognition result is more accurate, but the feature item extraction result is single and the process is cumbersome.

In domestic and foreign research, the method of discovering interdisciplinary topics from the perspective of document content is the most commonly used. First, establish a content relationship network based on the text content of scientific research results and the disciplines to which they belong, and then conduct thematic analysis of the text content. For example, Zhang [14] uses the LDA theme to cluster the term-document matrix generated by the basic discipline that has a strong connection with the management information system discipline, and obtains the cross-topic of this discipline and other disciplines. Wu [15] used an improved topic correlation analysis on the known interdisciplinary literature to extract common topics and their respective independent topic phases. Ruan [16] combined cluster analysis, LDA model, social relationship network and other methods to explore interdisciplinary research topics. Zhang [17] explored the intersection of professional knowledge in different disciplines through a combination of terminology and citation content. Dong [18] comprehensively used methods such as keyword co-occurrence network, high TI words, and emergency monitoring to conduct a comprehensive analysis of interdisciplinary topics in the field of library and information. Wang [19] used text mining technology to analyze the co-words of five related fields of nanotechnology, and revealed the cross-themes of these related fields. However, the subject discovery from a single perspective does not consider the process of dynamic knowledge flow.

4. Author Collaboration Perspective

Based on the principle of spreading or merging knowledge of different disciplines with the author as an intermediary, consider the relevant cooperation information of scientific researchers. In-depth mining and analysis of the information of the collaborators' research institutions, disciplinary backgrounds, and geographical and country distribution [20] can reveal the pattern and law of cross-science activities from another perspective [12]. First, establish an author's cooperation network based on the author's cooperation relationship in different disciplines, and then analyze the author's research theme. In theory, scientific research cooperation and subject themes should be intertwined and co-evolved [21]: On the one hand, the emergence of hot issues in major innovations will attract some researchers to conduct interdisciplinary cooperation and exploration; on the other hand, scientific researchers have adopted Interdisciplinary collaboration may stimulate the emergence of some new research topics. The method based on author cooperation can obtain direct and active interdisciplinary topics, but in reality, authors are more focused on their own subdivisions, resulting in less interdisciplinary cooperation, author's same name and organization changes, and the lack of unified standards for the classification of the author's disciplines and other issues, the topic recognition effect is not ideal.

There are still few related researches on the methods of discovering interdisciplinary topics from the perspective of author's cooperation. For example, Zhang [20] used the research themes summarized in the field of "the evolution of cooperation" and the co-evolution model of cooperation networks to understand the interdisciplinary subjects in the general sense from the perspective of complex systems. Schumme [22], in order to understand the crossover phenomenon of knowledge in different fields under different cooperation modes, analyzed the crossover of knowledge in specific subject fields from the perspectives of institution and regional cooperation. However, there are few interdisciplinary collaborations, the authors' subject classification lacks a unified standard, and the collaborators' duplicate names and difficulties in obtaining information from academic institutions have led to unsatisfactory results in topic discovery.

5. Reference Citation Perspective

Based on the principle that the knowledge of different disciplines can flow through the mutual citation of the discipline literature, it makes full use of the behavior of scientific researchers to "learn from" the views of others in their academic output. These "learning" contents include concepts, ideas, methods of collecting and analyzing data, and technical methods developed in other fields [12]. The citations of reference documents express the knowledge and information of other fields in the field [23]. First, identify the documents with interdisciplinary citation relationships, and then analyze the citation network of the cited documents. This method is simple and convenient, and can use the existing mature literature analysis tools, but it has the limitations of lag in citation and differences in motivation of citation.

The method of discovering interdisciplinary topics from the perspective of reference and citation is also a relatively common method. The interdisciplinary measurement index is often used as a measurement method, and the two complement each other. For example, Li [24] analyzed the word frequency distribution and k-clique number distribution law of the cross-subject keywords through the cross-citing papers of the main exchange journals of different disciplines, and analyzed the cross-research topics of the two disciplines. Han [25] developed an interdisciplinary literature discovery method based on the Rao-Stirling index to measure the interdisciplinary literature to find interdisciplinary literature and obtain cross-research topics. Small [26] used co-citation data between journal articles and clustered analysis of the intersectionality and similarity between disciplines, and then obtained cross-research topics. Leydesdorff [27] used the Rao-Stirling index to measure high-level cross-disciplinary literature and topics, including citing and cited citation matrices, and made a visual display. Chi [28] used the literature in the field of cross-cultural relations research from 1980 to 2010 as a data set, and established a citation network to identify the main research topics in this field and the relationship between the topics, and found interdisciplinary topics. But citations are lagging, and the motivations for citations are also different.

6. Multiple Comprehensive Perspectives

A single perspective has its own shortcomings. Only by discovering interdisciplinary topics through multiple indicators can it meet the comprehensive nature of scientific research activities. For example, some scholars combine the literature content with the reference perspective: Xu [12] uses a standardized thesaurus to conduct multi-disciplinary cross-topic exploration based on the two perspectives of citation content and citation relationship, and obtains the interdisciplinary relationship Cross themes under the limited scope of the vocabulary; Raimbault [29] uses two methods of citation network analysis and semantic analysis to build a large corpus of nearly 200,000 articles to analyze the subject of cross-disciplinary research in the journal "Geography Generalist". There are also scholars who

integrate three perspectives. For example, Huang [20] effectively decompose and integrate the three different dimensions of interdisciplinary research: external knowledge integration, internal knowledge convergence, and scientific activity mode to form a multi-dimensional comprehensive measurement system for interdisciplinary evaluation. However, it is only devoted to interdisciplinary measurement, and the analysis of interdisciplinary topics is not in place.

7. Summary

The subject of interdisciplinarity can be regarded as the common and core research content formed in the process of interdisciplinary of two or more disciplines. In the field of bibliometrics, interdisciplinary research is mainly based on scientific research results, and qualitative analysis and quantitative analysis need to be considered. The principles, advantages and disadvantages of each method are shown in Table 1.

Table 1. The principles, advantages and disadvantages of each method

Method	Principle	Advantage	Disadvantage
Document content perspective	The knowledge of different disciplines spreads or merges through the medium of text content.	Deeper into the micro text structure, the recognition result is more accurate.	The feature extraction result is single and the process is cumbersome.
Author collaboration perspective	The knowledge of different disciplines is diffused or merged through the intermediary of the author.	Interweaving and evolving together with subject themes can get direct and active themes.	There are few interdisciplinary collaborations, the collaborators have the same name, and the author's subject classification lacks a unified standard.
Reference citation perspective	The knowledge of different disciplines flows through the mutual citation of the discipline literature.	It is simple and convenient to use the existing mature bibliometric software.	Citation lag, citation motivation difference.

To sum up, most of the past researches on interdisciplinary subjects are based on a single perspective, and a single perspective is difficult to conduct a comprehensive and in-depth discussion on the complex issue of interdisciplinary, and its inherent methodological defects will be relatively highlighted [20]; Combining the content of the literature with the reference perspective, neglecting the information mining of the author, is not conducive to the overall management of scientific research personnel; although some scholars combine the three perspectives, they only measure and evaluate the cross-over results, and do not analyze it at the micro level. The generation principle of interdisciplinary topics does not reveal the information of scientific researchers involved in interdisciplinary. In future research, the three perspectives can be integrated, making full use of the differences of each perspective, complementing each other's advantages, and complementing each other, and discovering interdisciplinary research topics more scientifically and efficiently.

References

- [1] Information on: http://www.gov.cn/xinwen/2019-08/22/content_5423254.htm
- [2] Jinping Xi. Speech at the Symposium of Scientists. People's Publishing House, 2020.
- [3] Information on: http://www.moe.gov.cn/srcsite/A22/yjss_xwgl/xwgl_xwsy/202101/t20210113_509633.html
- [4] Information on: http://www.moe.gov.cn/jyb_xwfb/s271/202101/t20210113_509682.html

- [5] Rafols I, Meyer M. How cross-disciplinary is bionanotechnology? explorations in the specialty of molecular motors. *Scientometrics*. Vol.70(2007) No.3, p.633-650.
- [6] Chunmei Zhang, Fengxia Hao, Hongxiu Yan. The charm of interdisciplinary research——An analysis of the Nobel Prize in Natural Science for a Century. *Science Technology and Dialectics*. (2001) No.6, p.63-67.
- [7] Haiyun Xu, Chunxiao Yin, Ting Guo, et al. Overview of Interdisciplinary Research. *Library and Information Service*. Vol.59(2015) No.5, p.119-127.
- [8] Guangzu Bai, Yurong Zheng, Xinnian Wu, et al. Research and Empirical Research on Disruptive Technology Foresight Method Based on Literature Knowledge Association. *Journal of Information*. Vol.36(2017) No.9, p.38-44.
- [9] Haiyun Xu, Kun Dong, Ling Kui. Research on subject identification and prediction methods for interdisciplinary subjects. Science and Technology Literature Publishing House, 2019.
- [10] Ying Ye, Jiarong Zhang, Hui Zhang. The relationship between knowledge flow and interdisciplinary research. *Library and Information*. (2020) No.3, p.29-33.
- [11] Klein J T. a conceptual vocabulary of interdisciplinary science. *Practicing interdisciplinarity*. (2000), p.3-24.
- [12] Shurui Xu: Research on multidisciplinary cross-topic detection based on citation relationship and citation content (Master degree, Nanjing University of Science and Technology, China 2018).
- [13] LeiX ia. Research on interdisciplinary subject identification based on journal articles. *New Century Library*. (2019) No.12, p.62-67.
- [14] Bin Zhang. Interdisciplinary Subject Exploration: From the Perspective of Subject Clustering. *Information Science*. Vol.38(2020) No.10, p.49-55.
- [15] Lei Wu, Ruya Tian, Xuefu Zhang. Interdisciplinary theme discovery method and empirical research based on theme-related analysis: Taking the field of animal resources and breeding as an example. *Library and Information Service*. (2017) No.1, p.72-79.
- [16] Guangce Ruan, Lei Xia. Interdisciplinary Interdisciplinary Research Topic Identification——Taking Library and Information Science and Pedagogy as an Example. *Information Science*. Vol.38(2020) No.12, p.152-157.
- [17] Chengzhi Zhang, Shurui Xu, Chao Lu. Methods and empirical evidence for monitoring multi-disciplinarity phenomena using citation content. *Library and Information Service*. Vol.60(2016) No.19, p.108-115.
- [18] Dong K, Xu H, Luo R, et al. An integrated method for interdisciplinary topic identification and prediction: a case study on information science and library science. *Scientometrics*. Vol.115(2018) No.2, p.849-868.
- [19] Wang L, Notten A, Surpatean A. Interdisciplinarity of nano research fields: A keyword mining approach. *Scientometrics*. Vol.94(2013) No.3, p.877-892.
- [20] Ying Huang, Lin Zhang, Beibei Sun, et al. Interdisciplinary three-dimensional measurement-external knowledge integration, internal knowledge convergence and scientific cooperation mode. *Studies in Science of Science*. Vol.37(2019) No.1, p.25-35.
- [21] Wenqi Zhang: Research on the Evolution of Interdisciplinary Cooperation Network Based on Comprehensive Subject Analysis (Master degree, Dalian University of Technology, China 2016).
- [22] Schummer J. Multidisciplinarity, interdisciplinarity, and patterns of research collaboration in nanoscience and nanotechnology. *Scientometrics*. Vol.59(2004) No.3, p.425-465.
- [23] Porter A L, Cohen A S, Roessner J D, et al. Measuring researcher interdisciplinarity. *Scientometrics*. Vol.72(2007) No.1, p.117-147.
- [24] Changling Li, Feifan Liu, Fengjiao Guo. Using the overlapping community visualization software CFinder to analyze the subject of interdisciplinary research——taking information science and computer science as examples. *Library and Information Service*. Vol.57(2013) No.7, p.75-80.

- [25] Zhengqi Han, Xiaoping Liu, Han Xu. Discovery of interdisciplinary literature based on Rao-Stirling index: Taking nanoscience and nanotechnology as examples. Library and Information Service. Vol.62(2018) No.1, p.125-131.
- [26] Small H. Maps of science as interdisciplinary discourse: co-citation contexts and the role analogy. Scientometrics. (2010) No.83, p.835-849.
- [27] Leydesdorff L, Rafols I, Chen C. Interactive over-lays of journals and the measurement of interdisciplinarity on the basis of aggregated journal-journal citations. Journal of the American Society for Information Science and Technology. Vol.64(2013) No.12, p.2573-2586.
- [28] Chi R, Young J. The interdisciplinary structure of research on intercultural relations: a co-citation network analysis study. Scientometrics. Vol.96(2013) No.1, p.147-171.
- [29] Raimbault J. Exploration of an interdisciplinary scientific landscape. Scientometrics. Vol.119(2019) No.2, p.617-641.