Identifying Core Research Topics and Visualizing Evolutionary Paths in Journal Articles: A Case Study of China's Healthcare Information Domain (Postprint)

Authors: Yue Lixin, Zhou Xiaoying, Chen Yini

Date: 2023-04-01T16:15:50+00:00

Abstract

[Purpose/Significance] This study proposes a methodology for identifying domain-specific core research topics and visualizing their evolutionary pathways, aiming to provide insights for research on topic evolution analysis and to reveal the evolutionary characteristics and developmental patterns of core topics within a field. [Method/Process] The approach employs the LDA model for topic identification, integrates multidimensional scaling analysis and visualization techniques to project LDA topic identification results onto a two-dimensional space for identifying inter-topic relationships and distinguishing core topics from secondary ones; additionally, a topic similarity algorithm is utilized to detect associations between topics across adjacent time periods, and a novel visualization method is proposed to construct cross-evolutionary pathways for different types of research topics, thereby revealing the dynamic changes of both core and secondary topics throughout the evolutionary process. [Results/Conclusion] An empirical study is conducted using China's medical and health information domain as a case example; the findings indicate that core research topics in this domain primarily include electronic health records and internet-based healthcare, among which core topics such as health management and smart healthcare exhibit promising evolutionary development trends.

Full Text

Research on Identification and Evolution Path Visualization Methods for Core Research Topics in Journal Articles: A Case Study of China's Medical and Health Information Field

Yue Lixin, Zhou Xiaoying, Chen Yini

School of Information Resources Management, Renmin University of China, Beijing 100872

Abstract

[**Purpose/Significance**] This paper proposes methods for identifying domain core research topics and visualizing their evolution paths, aiming to provide a reference for research on domain topic evolution analysis and to reveal the evolutionary characteristics and development patterns of core topics in a field. [**Method/Process**] The study employs the LDA model for topic identification, combines multidimensional scaling analysis and visualization techniques to map LDA topic identification results into two-dimensional space, identifies relationships between topics, and determines core and secondary topics. A topic similarity algorithm is used to detect associations between topics in adjacent periods, and a new visualization method is proposed to construct cross-evolution paths for different types of research topics, revealing the dynamic changes of core and secondary topics during evolution. [**Result/Conclusion**] An empirical study was conducted using China's medical and health information field as an example. The results show that core research topics in this field mainly include electronic health records and internet-based medical treatment, among which core topics such as health management and smart medical care demonstrate favorable development and evolution trends.

Keywords: core research topics; topic identification method; topic evolution path; visualization method; medical and health information

Classification Number: G251.2
DOI: 10.13266/j.issn.0252-3116.2020.05.010

Introduction

In recent years, analyzing research topics and their development trends in various fields using methods such as keyword frequency analysis, co-word analysis, citation analysis, topic detection and tracking, and topic evolution has become a hot spot in information science research both domestically and internationally. However, most existing studies rely on software tools such as CiteSpace, Ucinet, and SPSS for research topic identification and trend analysis of academic papers. As data volume and users' fine-grained demands change, macroscopic and static results like co-word networks and citation networks gradually fail to meet

the needs of disciplinary intelligence analysis. Currently, there are few research results on the dynamic paths of topic evolution based on topic type classification and their temporal changes. To address this gap, this paper first investigates methods for core research topic identification and evolution path visualization analysis, proposes a new visualization method for constructing cross-evolution paths of two different types of research topics—core topics and secondary topics, conducts empirical analysis using paper data from the medical and health information field in the CNKI journal full-text database, and applies visualization analysis results to specifically analyze the evolution process of core and secondary topics in the medical and health information field.

1 Literature Review

1.1 Topic Identification Topic identification refers to the use of bibliometrics, natural language processing, and other methods to mine and analyze research topics in scientific literature. Current topic identification methods mainly include co-word network-based approaches, community detection, and the LDA (Latent Dirichlet Allocation) topic model. Related studies include: A.D. Ritzhaupf et al. used co-word network analysis for topic identification and applied this method to analyze the main research topics and development trends of distance education in North America; Cheng Qikai et al. proposed a topic identification method based on community detection models; Wang Xiaoyue et al. proposed a disciplinary topic identification method based on the LDA model and conducted an empirical study using data from projects funded by the U.S. National Science Foundation to verify the feasibility of the method.

1.2 Topic Evolution Topic evolution refers to the dynamic change process of research topics embedded in journal papers over time. It mainly describes the growth, division, fusion, decline, and other states of research topics in a field within a certain period, helping to reveal the current status, changes, and trends of research. How to accurately and effectively identify the evolution 脉络 of research topics from massive academic papers and visualize them has become an urgent problem. Many scholars have conducted topic evolution research. For example, Li Xiangdong et al. proposed a topic evolution analysis method for scientific journals based on the LDA model, introduced time factors, and realized topic evolution in both intensity and content dimensions based on LDA topic identification and JS divergence calculation results. Liu Ziqiang et al. proposed a topic evolution analysis method from a multi-dimensional perspective, constructing a topic evolution model with three dimensions—topic intensity, topic structure, and topic content—and verified the effectiveness of the method through an empirical study on big data research in the domestic library and information science field. Zhou Yuan et al. integrated author features into topic analysis, constructed an author-topic evolution model based on a weighted Jaccard similarity algorithm, and could analyze scholars with high influence on a certain research topic in different periods.

Through analysis of existing research findings, it is found that: in terms of topic identification, most current research conducts static topic identification with insufficient analysis of the relative importance between topics. In fact, there are primary and secondary relationships between topics within different time periods, and treating all research topics equally to some extent limits the accuracy of analysis of disciplinary status and development trends. In terms of topic evolution, current research focuses on analyzing fusion and division processes by examining topic intensity, content, and other dimensional features, but the identification of relationships between research topics and changes in topic relationships at different evolution stages need further in-depth study. In terms of topic evolution path visualization, current methods mainly focus on analyzing associations between adjacent period topics, with few results analyzing the relationships among topics within the same time window. To address these limitations, this paper proposes a topic identification and evolution path visualization method based on topic category classification to improve upon the above shortcomings.

1.3 Topic Evolution Path Visualization Researchers in data mining and visualization have conducted extensive studies on topic evolution and proposed numerous topic evolution visualization methods and tools. For example, S. Havre et al. proposed the ThemeRiver visualization model, where the horizontal axis represents time, different colored lines represent topics, and line thickness represents topic intensity at different time windows, showing the overall topic evolution 脉络 of a field. M. Rosvall et al. proposed a community topic evolution visualization analysis method based on alluvial diagrams, displaying communities at different time windows on a horizontal time dimension and using different colored lines to represent community evolution paths. Wang Xiaoguang et al. developed the NEViwer software for topic evolution visualization analysis based on co-word network analysis, providing functions for colored network diagrams and alluvial diagram drawing, which can effectively reveal the macro process and micro details of topic evolution. Mu Dongmei et al. integrated "three metrics" theory, social network analysis methods, disciplinary knowledge structure theory, and knowledge mapping technology through optimization and collaborative integration, and explored method processes to reveal knowledge structures at high, medium, and low levels according to different levels of knowledge structure, providing a theoretical foundation for disciplinary structure visualization. Meanwhile, Mu Dongmei et al. used time-keyword co-occurrence analysis to construct a time-keyword two-dimensional matrix, adopted four methods—cluster analysis, social network analysis, temporal word frequency statistics, and topic classification—to visualize the time-keyword two-dimensional matrix, conducted multi-dimensional analysis of the dynamic knowledge structure of the LIS field, and based on temporal analysis and topic-keyword co-occurrence analysis, constructed a 2-mode network, using NetDraw to visually present the evolution patterns of various topics.

Analysis of existing research shows that: in topic identification, most current

research conducts static topic identification with insufficient analysis of relative importance between topics. In fact, there are primary and secondary relationships between topics within different time periods, and treating all research topics equally to some extent limits the accuracy of disciplinary status and development trend analysis. In topic evolution, current research focuses on analyzing fusion and division processes by examining topic intensity, content, and other dimensional features, but the identification of relationships between research topics and changes in topic relationships at different evolution stages need further in-depth study. In topic evolution path visualization, current methods mainly focus on analyzing associations between adjacent period topics, with few results analyzing relationships among topics within the same time window. To address these limitations, this paper proposes a topic identification and evolution path visualization method based on topic category classification to improve upon the above shortcomings.

2 Methodology for Core Research Topic Identification and Evolution Path Visualization Based on Topic Category Classification

2.1 Theoretical Basis Keywords and subject terms in journal papers are the refinement of their core content, and research topics are the basic units that effectively represent disciplinary knowledge. Therefore, bibliometrics and natural language processing methods can be used to identify research topics embedded in journal papers and analyze hotspots, frontiers, and development trends in a field. R.N. Kostoff et al. from the Office of Naval Research (ONR) divided research topics into pervasive themes and sub-themes, and through experimental analysis of the relationship between the two types of topics found that: pervasive themes and sub-themes have close associations—changes in pervasive themes cause changes in sub-themes, but changes in sub-themes basically do not cause changes in pervasive themes; when pervasive themes remain stable, sub-themes may still change. Pervasive themes and sub-themes together constitute a complete domain topic network. In topic evolution analysis, distinguishing the primary and secondary relationships of research topics and comprehensively considering the synergistic effects of both can improve the accuracy and effectiveness of analysis.

This study draws on the basic ideas of R.N. Kostoff's topic analysis research. Based on the concepts of "pervasive themes" and "sub-themes" proposed by Kostoff, this study divides paper topics into two categories—"core topics" and "secondary topics"—according to their importance, and proposes a core topic identification and evolution path visualization method based on topic category classification theory.

2.2 Method Process and Framework The basic steps and framework of the core topic identification and evolution path visualization method based on topic categories are as follows: First, determine the data source (database), retrieval strategy, and time region according to the field, and collect and organize journal

paper data. Second, based on data preprocessing and time window division, use the LDA model for topic identification. Third, combine multidimensional scaling analysis and visualization techniques to map LDA topic identification results to two-dimensional space, identify relationships between topics, and determine core and secondary topics. Fourth, use a topic similarity algorithm to detect associations between topics in adjacent periods, propose a new visualization display method, construct cross-evolution paths for different types of research topics, and reveal the dynamic changes of core and secondary topics during evolution.

The main content of the above steps is introduced in detail below.

2.2.1 Research Topic Identification Based on the LDA Model In recent years, many topic models have been proposed in academia, such as Latent Semantic Analysis (LSA), probabilistic Latent Semantic Analysis (pLSA), and the LDA model. Compared with LSA and pLSA models, the LDA model can not only predict the topic distribution of training set documents but also effectively predict the topic distribution of documents and words not in the training set. Therefore, the LDA model has gradually become a major tool for analyzing large-scale unstructured document collections. LDA is a three-layer (word, topic, and document) Bayesian probability model that assumes documents are composed of several latent topics, and topics are composed of all words in the vocabulary. The joint distribution probability of the LDA topic model is shown in Formula (1) [19]:

$$P(\theta, z, w) - P(\theta|w) \sum_{n=1}^{N} P(z_n|\theta) P(w_n|z_n, \beta) \quad (1)$$

where z represents topics, w represents topic words, N represents the number of words in the m-th document, and θ is sampled from the Dirichlet distribution with parameter α. Since the LDA topic model can analyze text topics more accurately and efficiently compared with other topic identification methods (such as keyword clustering and community detection), this paper will use the scikit-learn toolkit in Python for topic identification in the medical and health information field.

2.2.2 Core Research Topic Identification Based on MDS LDA topic identification results generally cannot directly analyze the relationships between different topics. To obtain core topics among research topics, this study uses Multidimensional Scaling (MDS) to construct a low-dimensional space using the similarity between topics, making the distances between LDA topics in this space as consistent as possible with the similarity between LDA topics in high-dimensional space, thereby visualizing the relationships between LDA topics and intuitively identifying core topics.

This study uses the pyLDAvis toolkit in Python to draw dynamic interactive LDA topic visualization maps to analyze the relationships between research topics and identify core and secondary research topics. pyLDAvis can adjust the parameter λ ($0 \leq \lambda \leq 1$) to control the topic-word relevance $r(w, k|\lambda)$, which can control the display of different subordinate word items for a topic. When $\lambda = 0$, it displays subordinate word items that are unique and relatively independent under the topic, meaning these word items often appear only in that topic; when $\lambda = 1$, it displays subordinate word items with higher distribution probability, but these high-probability word items often do not belong exclusively to that topic and also belong to other topics. The parameter λ calculation method is shown in Formula (2) [20]:

$$r(w, k|\lambda) = \lambda \log(\phi_{kw}) + (1 - \lambda) \log(p_w) \quad (2)$$

where w represents topic words, $w \in \{1, 2, 3, ..., V\}$; k represents topics, $k \in \{1, 2, 3, ..., K\}$; ϕ_{kw} represents Gibbs sampling parameters; and p_w represents the distribution probability of topic word w.

2.2.3 Visualization of Core and Secondary Topic Evolution Paths Based on the analysis of related research on topic evolution path visualization, current methods such as ThemeRiver, TextFlow [21], and NEViwer mainly focus on analyzing associations between adjacent period topics and treat all topics equally, making it difficult to effectively analyze the relationships among topics within the same time window and the evolution relationships of different types of topics. Therefore, this paper proposes a new domain core research topic identification and evolution path visualization method: a visualization method based on R language stream graphs to analyze the evolution paths of core and secondary topics, which can effectively reveal the dynamic changes of relationships such as division and fusion between core and secondary topics during evolution.

Compared with existing topic evolution path visualization methods, the evolution path visualization map designed in this paper can analyze the flow patterns of a certain type of research topic over time, analyze the cross-evolution 脉络 of different types of research topics such as core and secondary topics, and display the dynamic change process of association relationships. The basic style of visualization is shown in Figure 1 [Figure 1: see original paper], where blocks represent topics, stream graphs between blocks represent the evolution paths (association changes) of these topics over time, thickness represents the strength of associations between topics; block height represents topic intensity (higher document probability distribution results in larger topic blocks); core research topics are labeled with "core," and secondary research topics are labeled with "secondary."

3 Application of Topic Identification and Evolution Visualization Methods in the Medical and Health Information Field

3.1 Data Source and Preprocessing This study selects the CNKI journal full-text database as the data source and collects key bibliographic information such as titles, keywords, and abstracts. The specific retrieval strategy is as follows: Database: CNKI; Retrieval strategy: Subject = "medical and health information"; Time span: unlimited; Retrieval results: 704 papers; Retrieval date: June 3, 2018. The annual distribution of obtained documents is shown in Figure 2 [Figure 2: see original paper].

Since topic identification only requires analysis of text fields, titles, keywords, and abstracts are extracted separately, and the downloaded journal papers from four sub-periods are preprocessed for topic identification. The processing mainly includes format conversion, deduplication, removal of stop words and punctuation, etc. Chinese word segmentation is performed using Python's jieba toolkit before LDA topic identification. To improve the accuracy of word segmentation results, a segmentation dictionary is constructed based on keywords from the collected journal papers in the medical and health information field (keywords are highly condensed summaries of journal paper content and maintain the main vocabulary authors intend to express more accurately than the default dictionary in the jieba toolkit). The basic format of this dictionary is word-frequency-part of speech. Since this study does not involve part-of-speech analysis, the part-of-speech is ignored, and the custom dictionary is saved in txt format and called through jieba.load_{userdict}("dict.txt").

Researchers conducting topic identification and evolution analysis need to divide time windows to clarify the temporal dimension of topic evolution (dividing journal paper data into several continuous sub-periods). The main methods for dividing time windows are determining based on data time labels and using fixed annual time windows [22-23]. This paper uses the fixed annual time window method to divide the retrieval results into four sub-periods. The years and number of documents for each sub-period are shown in Table 1 :

Table 1 Division of Sub-periods for Topic Identification and Evolution

Period (Year)	Number of Documents
1996-2009	68
2010-2012	112
2013-2015	241
2016-2018	283

Since there was relatively little research related to domestic medical and health information before 2010, 1996-2009 is designated as sub-period I (68 papers). After 2010, research results gradually increased, so each three-year period is designated as a sub-period: 2010-2012 as sub-period II (112 papers); 2013-2015 as sub-period III (241 papers); and 2016-2018 as sub-period IV (283 papers).

3.2 Medical and Health Information Research Topic Identification Based on the LDA Model After Chinese word segmentation using Python's jieba toolkit, LDA topic identification is performed using Python's scikit-learn toolkit (conducting LDA topic identification sequentially for the four divided periods). After processing, the LDA topic identification results for each sub-period show only some research topics related to the medical and health information field (this paper only lists the top 5, with subordinate word tables showing partial results), followed by their corresponding keywords and subordinate keywords sorted by frequency, as shown in Table 2 .

Table 2 List of Research Topics and Subordinate Keywords in Different Periods of China's Medical and Health Information Field (Partial)

Period (Year)	Topics and Subordinate Keywords
Period I (1996-2009)	Health records | Health communication | Information technology | Regional healthcare | Elderly | Poverty-stricken areas | Mass media | Developed countries | Medical services | Health Bureau | B2C | Health network | Medical health services | Difficulty in seeing a doctor | Unified standards | Health information resource platform | Anyang City | Expensive medical care | Health management | Resident health information system | Consumers | Medical information | Special medical services | Non-profit | Survey report | Federation | Health information work | Medical health information | Elderly chronic diseases | Health information | Remote care | Online health information | Growth path | Medical health information | Medical records | Control rights | Golden Health Network | Privacy rights | Information highway | Medical network | Personal information | National level | Comprehensive

Period (Year)	Topics and Subordinate Keywords
Period II (2010-2012)	Electronic health records | Medical institutions | Medical services | Information technology | Cloud computing | Health information | Electronic health | Health records | One-card system | Healthcare | Information platform | Medical health | Traditional Chinese medicine clinical information standards | Electronic health records | Body sensor networks | Internet | Two-way referral | Health communication | Health information | Internet | Health information resource platform | Health monitoring | Cloud platform | Body sensor networks | Healthy lifestyle | Ningbo City | Information resources | Internet of Things | Medical health | Healthcare | Electronic health | Resident health | Electronic health records | Medical institutions | Medical services | Information technology | Cloud computing | Health information | Electronic health | Health records
Period III (2013-2015)	Medical health informatization | Medical health | Health records | Health information | Informatization | Mobile health | Service model | Cloud computing | Big data | Resident health records | Elderly | Medical-nursing integration | Information platform | Medical information | Home-based elderly care | Pharmaceutical e-commerce | New media | Medical services | Internet medical treatment | Electronic health literacy

Period (Year)	Topics and Subordinate Keywords
Period IV (2016-2018)	Health information acquisition | HADOOP | Smart medical care | Cloud computing | Medical services | Health information | Personal health information management | Electronic health records | Health management | Resident health card | Cloud computing | Health information | UGC | Online health community | Internet | Social media | Health education | Mobile medical treatment | APP | Health communication | Health information | Health management | Medical health | Mobile Internet | Wearable | Information platform | Medical health | Internet medical treatment | Big data | Internet | Medical services | Big data application | Health monitoring | Medical big data | Usage intention | Patients | Electronic health literacy | Intelligent health management | Mobile medical services | Self-efficacy | O2O medical services | Health management | Health management service industry | Health cloud | Infectious disease patients | Medical services | Electronic health records | Service platform | Internet | Health information | Health information acquisition | Faculty and staff | Public services | Information technology | Influencing factors | Medical service websites | Health information | Influencing factors | Health literacy | Health insurance | Accuracy rate | School hospital | Diabetes | Patients | Wearable devices

3.3 Core Research Topic Identification Results in Medical and Health Information Related Fields Based on MDS To better analyze the relationships between LDA topics, based on the LDA topic identification results from the previous step, Multidimensional Scaling (MDS) is used to construct a low-dimensional spatial distribution of LDA topics to visualize the relationships between LDA topics and discover core research topics in each period of China's medical and health information field.

The specific data processing procedure is as follows: Based on LDA topic identification results, the pyLDAvis toolkit in Python is used to draw interactive LDA topic visualization maps for the four sub-periods, as shown in Figures 3 [Figure 3: see original paper], 4 [Figure 4: see original paper], 5 [Figure 5: see original paper], and 6 [Figure 6: see original paper] (5 topics are listed for each period, added manually; 30 subordinate words are listed).

In Figures 3, 4, 5, and 6, the large circles on the left represent core topics, while small circles represent secondary topics. Due to differences in the number of documents and research topics in each period, the number of circles varies. The right side shows subordinate word items for each topic. Based on this, a brief analysis of core research topics in different periods of China's medical and health information field can be conducted:

Period I's core topics include health records, health platforms, and health management. During this period, medical and health-related research gradually developed, but related research was relatively lacking. Period II's core research topics include electronic health records, internet information platforms, and internet health communication. Due to technological development, research topics in this stage had new research content based on new technologies. Period III's core topics include medical health informatization, electronic medical treatment, and smart medical care. With further development of information technology, electronic health further developed and smart medical care emerged. Period IV's core research topics include internet medical treatment, electronic health literacy, and health management. During this period, relying on technology, research on internet medical treatment further increased, and the public paid more attention to health. Due to the development of internet technology, the public could access health information more conveniently, so research on health literacy gradually increased.

3.4 Visualization of Research Topic Evolution in China's Medical and Health Information Field

3.4.1 Cross-Evolution Analysis of Core and Secondary Topics

Based on the core research topic identification results for the four periods in China's medical and health information field, an evolution 脉络 map is drawn using the cross-evolution path visualization method for core and secondary topics proposed in this study to analyze the core research topics and their development and evolution process in each period, as shown in Figure 7 [Figure 7: see original paper] (color figure URL: https://www.informationscience.top/topicevlution.html).

The core and secondary topics selected in Figure 7 are overlapping topics based on LDA model identification of medical and health information research topics and MDS-based core research topic identification (if a topic does not exist in a certain period, the topic with the highest similarity is used instead based on topic similarity calculation). Figure 7 shows that core and secondary topics in each period are constantly developing and changing. Several representative topics are analyzed in detail below:

(1) **Electronic Health Records**: Electronic health records were secondary research topics in Periods I and II. With technological development, they became core research topics after Period III. Current research in this area mainly includes two aspects: first, comparative studies on electronic health

record construction at home and abroad and related domestic experience, introducing the shared service system construction of resident electronic health records in major foreign countries and providing suggestions for China's shared service system construction; second, the establishment of electronic health records and electronic health record management systems. The electronic health system architecture describes the overall technical composition of electronic health records and the relationships between technical elements, which is one of the core technologies of electronic health records.

(2) **Internet Medical Treatment**: Internet medical treatment developed under the background of rapid IT technology development, mobile communication entering the 4G era, internet applications evolving to Internet+, and the rapid development and popularization of big data and cloud computing technologies. Therefore, similar to electronic health records, it became a core research topic after Period III. Currently, internet medical treatment uses advanced information technology and Internet+ application platforms to improve the efficiency of medical resource use and enhance treatment and service levels, becoming an important direction for the development of China's medical and health industry in the near future. In the new era, technical factors play a crucial role in the application and promotion process of internet medical treatment. How to construct a development model for the internet medical treatment industry that can achieve long-term healthy and sustainable development is worth studying.

(3) **Health Communication**: Health communication has been a core research topic in all periods except Period II when it became a secondary topic. As an emerging branch of communication studies, this topic has received widespread attention due to its close association with personal life and significant social influence, and it has been an important research hotspot in each period. However, current domestic health communication research [24] remains at the primary stage of describing phenomena, discussing cases, and summarizing macro characteristics of the field. Exploring how to conduct health communication research under theoretical guidance, investigating important influencing factors in the health information communication process and health behavior development process, examining the interaction between people and information, people and people, and people and society, and studying the changes in people's concepts, attitudes, and behaviors brought about by health information communication will be very important issues.

(4) **Mobile Medical Treatment**: Mobile medical treatment refers to the provision of medical services and information through mobile communication technology. Due to its strong practicality, which can be achieved through a few simple APPs with low technical costs and ease of use, it has always been a core research topic during the evolution stage. Current research mainly focuses on [25-26]: first, specific implementation process

research of mobile medical treatment, conducting detailed mobile medical health needs analysis to summarize key points for application software design and mobile medical health development; second, sorting out the current status of mobile medical application industries at home and abroad, exploring the development models of typical applications, analyzing key elements, and proposing development suggestions suitable for China's national conditions; third, analyzing mobile medical treatment users and exploring personalized mobile medical health services for different users.

Based on the above analysis, changes in research topic types are closely related to technological development. Electronic health records and internet medical treatment have gradually become core research topics in the medical and health field based on new technologies. Health communication and mobile medical treatment, as core research topics in each stage, have new research content under technological development. Research on medical and health information has always been a secondary research topic because its correlation with technology is lower than other topics, but based on current research content, this topic is also gradually beginning to use new technologies and is very likely to produce new research content in the near future.

3.4.2 Development Trends In recent years, with the continuous development of internet technology, research topics in the medical and health information field have also been constantly changing. New technology topics continue to emerge and show a growing evolution trend, while some topics have gradually declined due to the impact of new technology topics. In addition, some key research topics have maintained good development momentum. Based on the previous analysis of core and secondary topics, the development trends of core research topics are visualized, as shown in Figure 8 [Figure 8: see original paper] (color figure URL: https://www.informationscience.top/yh.html).

According to Figure 8, five typical research topics are selected to analyze their development trends. These five topics can be divided into three categories: The first category is key research topics in the medical and health field that have not disappeared with social development but have new research content under the influence of new technologies. The second category is new research topics generated under the background of technological development, which are inevitable products of the times. The third category is disappearing research topics, whose development trends are also affected by technological development and are replaced by more advanced technologies in the new era, which is also an inevitable process of times.

(1) **Health Management**: Health management has always been a research focus in the medical and health information field, showing a gradually increasing trend over time. Past journal papers mostly introduced foreign health management practices and achievements. With the transformation of people's lifestyles and health concepts in recent years, journal paper research results reflect that health management has also shown new char-

acteristics under new health models and has gradually become an emerging health service concept and service mode. Health management has always been a field of public concern, so it has maintained an upward momentum during evolution. The health management service industry [27] is oriented by consumer health needs, replaces the single economic goal with multiple goals, and represents an industrial innovation development model and people-oriented development model for coordinated development between human progress and economy and society. Currently, China's health management service is in the initial stage of rapid industrial technology development and diversification formation. In this context, the development of health management services has driven scholars' research on health management service systems, currently focusing on: first, medical health management service systems centered on various diseases; second, health management service technology; third, the composition and comparative research of health management systems or systems.

(2) **Smart Medical Care**: Smart medical care is the deep integration of information technology with medical and health services and management. It has achieved rapid development with strong national support and has had a profound impact on medical service models, health management methods, and resident health management. However, there are still some problems and challenges in medical data, system security, construction guarantee, resource sharing, and evaluation systems. Therefore, on the future development path of smart medical care, the government still needs to strengthen macro guidance and expand the scope of information sharing to better meet patients' needs. Smart medical care is a research topic that has gradually increased in recent years. Its emergence and development are inseparable from Internet+, big data, and other technologies, and it is a product of the new era.

(3) **Medical Big Data**: Medical big data is an inevitable product under the big data environment and a new research hotspot and focus in the medical and health information field in recent years. As an extremely important part of big data, its application includes not only data information in medical aspects but also massive data digitally stored in health services and life health. Research on medical big data mainly includes: first, review research on the current status and development trends of medical big data. By studying related research on medical big data at home and abroad, clarify the current development status and research hotspots of domestic medical big data and predict future development trends; clarify the opportunities and challenges that medical big data may encounter in future development and formulate relevant measures to maximize development; research on the contribution of medical big data to other related fields. Second, medical service models under medical big data. Taking the construction of clinical data centers as the entry point and the clinical data center as the core of medical big data platforms, research on the construction of medical big data platforms has become a new hotspot [28].

Third, privacy protection issues in medical big data. Therefore, research on medical big data will be a research hotspot within the next few years.

(4) **Medical Information Technology**: Medical information technology introduces information technology into the medical field, constructs new medical service models, and provides better medical information service technology. Research on medical information technology first appeared in the 1990s when information technology had gradually begun to penetrate various fields. The medical field conformed to the development of the times by introducing information technology, which promoted the development of the medical and health industry. With further development of information technology, hotspots in medical information technology have gradually shifted to smart medical care, medical big data, and other fields for in-depth research. Research in journal papers on pure medical information technology has gradually decreased.

(5) **Regional Health Informatization**: Regional health informatization, similar to the development of medical information technology, was spawned by technological development and public demand, and similarly shifted to other fields due to continuous technological development and increasingly urgent public demand. The research scope of regional health informatization has gradually expanded from small regions and small scopes to the national level and even the international level. The meaning of "region" has gradually changed and been replaced by more vocabulary, so papers on this research topic have shown a substantial reduction in recent years.

Based on the above analysis, research topics in the medical and health information field have undergone significant changes over the past two decades. Changes in research topics are inseparably related to technological development. New technologies continuously inject new vitality into the medical and health information field, which is of great significance for improving medical service models, enhancing health service levels, and improving public health literacy. However, technological development also brings huge challenges to a certain extent. Issues such as public privacy and medical data leakage greatly test the professional quality of practitioners. Proper use of new technologies to provide better services for the public should be a key issue for consideration.

4 Discussion and Conclusion

The method proposed in this paper draws on the basic ideas of Kostoff's topic analysis research, dividing research topics into core and secondary topics. Based on MDS, it constructs association relationship detection for LDA topic identification results to identify core research topics. Compared with current core research topic identification and visualization analysis methods based on tools such as CiteSpace, UCINET, and SPSS, this method conducts further in-depth research on the association relationships between research topics and their changes at dif-

ferent evolution stages. In addition, this paper proposes a visualization method based on R language for the cross-evolution of core and secondary topics, which can visually display the development and evolution 脉络 of domain research topics and the dynamic change process of core and secondary topics in different time periods. The core technology topic identification and evolution visualization method based on large-scale scientific literature data helps identify core research content in a field and analyze the development direction of core research content, which is fundamental intelligence work for scientific innovation and has significant application value.

This study has two main limitations: First, the interpretation of LDA topic identification results (combinations of several topic words, difficult to interpret) relies on the professional knowledge of analysts. Therefore, more effective topic identification methods need to be explored to improve the semantic information content of results for easier interpretation. Second, the division between core and secondary topics in this study needs further refinement; for example, secondary research topics can be divided into emerging topics, declining topics, etc. Future research can further explore using semantically enhanced LDA models for topic identification to improve result interpretability and attempt to combine topic evolution life cycle division methods to conduct multi-level classification of topic types to increase the dimensions of topic evolution analysis.

References

[1] Wang Liya. Research progress on topic evolution [J]. Information Research, 2014(4): 29-32.

[2] Bai Rujiang, Leng Fuhai. Research on k-clique community knowledge innovation evolution method [J]. Library and Information Service, 2013, 57(17): 94-99.

[3] Ritzhaupf AD. An investigation of distance education in North American research literature using co-word analysis [J]. International review of research in open & distance learning, 2010, 11(1): 37-60.

[4] Cheng Qikai, Wang Xiaoguang. A research topic evolution analysis framework based on co-word network communities [J]. Library and Information Service, 2013, 57(8): 91-96.

[5] Wang Xiaoyue, Liu Ziqiang, Bai Rujiang, et al. Research front topic detection method based on fund project data [J]. Library and Information Service, 2017, 61(13): 87-98.

[6] Li Xiangdong, Zhang Jiao, Yuan Man. Research on scientific journal topic evolution based on LDA model [J]. Journal of Intelligence, 2014(7): 115-121.

[7] Liu Ziqiang, Wang Xiaoyue, Bai Rujiang. Research on multi-dimensional perspective-based disciplinary topic evolution visualization analysis method [J]. Journal of Library Science in China, 2016, 42(6): 67-84.

[8] Zhou Yuan, Zhang Chao, Tang Jie, et al. Intelligent identification of domain development paths based on topic change [J]. Library and Information Service, 2018, 62(14): 62-71.

[9] Havre S, Hetzler E, Whitney P, et al. ThemeRiver: visualizing thematic changes in large document collections [J]. Visualization & computer graphics IEEE transactions on, 2002, 8(1): 9-20.

[10] Rosvall M, Bergstrom CT. Mapping change in large networks [J]. Plos One, 2010, 5(1): e8694.

[11] Wang Xiaoguang, Cheng Qikai. Disciplinary topic evolution visualization analysis based on NEViwer [J]. Journal of the China Society for Scientific and Technical Information, 2013, 32(9): 900-911.

[12] Mu Dongmei, Zheng Xiaoyue, Ju Yuanhong, et al. Construction of disciplinary knowledge structure revelation model [J]. Library and Information Service, 2017, 61(12): 6-13.

[13] Zheng Xiaoyue, Mu Dongmei, Ju Yuanhong, et al. Exploration of disciplinary knowledge structure revelation process and method [J]. Library and Information Service, 2017, 61(12): 14-20.

[14] Mu Dongmei, Ju Yuanhong, Zheng Xiaoyue, et al. Research on disciplinary dynamic knowledge structure based on time-keyword co-occurrence analysis [J]. Library and Information Service, 2017, 61(12): 21-31.

[15] Zheng Xiaoyue, Mu Dongmei, Ju Yuanhong, et al. Research on disciplinary knowledge structure topic evolution pattern [J]. Library and Information Service, 2017, 61(12): 32-41.

[16] Kostoff RN, Eberhart HJ, Toothman DR, et al. Database tomography for technical intelligence: comparative roadmaps of research impact assessment literature and the Journal of the American Chemical Society [J]. Scientometrics, 1997, 40(1): 103-138.

[17] Landauer TK, Dumais ST. A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge [J]. Psychological review, 1997, 104(2): 211-240.

[18] Shen C, Li T, Ding CHQ. Integrating clustering and multi-document summarization by bi-mixture probabilistic latent semantic analysis (pLSA) with sentence bases [C]//AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2011: 914-920.

[19] Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation [J]. The journal of machine learning research, 2003(3): 993-1022.

[20] Sievert C, Shirley K. LDAvis: a method for visualizing and interpreting topics [C]//Proceedings of the workshop on interactive language learning, visualization, and interfaces. Baltimore: Association for Computational Linguistics, 2014: 63-70.

[21] Cui WW, Liu SX, Li T, et al. Textflow: towards better understanding of evolving topics in text [J]. Transactions on visualization and computer graphics, 2011, 17(12): 2412-2421.

[22] Wu Bin, Wang Bai, Yang Shengji. Event-based social network evolution analysis framework [J]. Journal of Software, 2011, 22(7): 1488-1502.

[23] Qian Tieyun, Li Qing, Xu Chengyu. Community core circle technology for scientific topic development segmentation [J]. Computer Science and Exploration, 2010, 4(2): 170-179.

[24] Zhao Jiakui, Lin Jun, Lu Ying, et al. Application status and countermeasures of new media in health communication [J]. Chinese Journal of Health Education, 2016, 32(10): 919-921.

[25] Hu Rong, Chen Huifang, Xu Weiguo. Impact of doctor-patient interaction in mobile medical services on patients' perceived value [J]. Management Science, 2018, 31(3): 75-85.

[26] Wang Xue, Shi Yuanbo, Huang Yueyang. Research on data mining based on Apriori algorithm in mobile medical terminal systems [J]. Digital Technology and Application, 2017(9): 60-61.

[27] Han Zhengchen, Zhang Jidong. Design and implementation of Android-based clinic appointment and health management system [J]. Information Technology and Informatization, 2016(12): 30-35.

[28] Lin Jing, Wu Xiangyang. Medical informatization construction based on "Internet+" [J]. Computer Knowledge and Technology, 2016, 12(23): 216-218.

Author Contributions

Yue Lixin: Responsible for literature review and data collection on topic identification and visualization, paper writing and revision.
Zhou Xiaoying: Responsible for paper framework and idea design, guiding paper writing and revision, corresponding author.
Chen Yini: Responsible for paper revision.

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