

Bibliometric analysis on data mining applied in education and teaching field

Xin QI, Linsi Huang

Jiangxi Agricultural Engineering College, Jiangxi, Zhangshu 331200

Abstract: Enter the era of big data, with the rapid development of computer science, data mining technology has been widely used in various fields. Among them, the application research of data mining in education and teaching has always been a hot field of concern for teaching and research workers. To use the document visualization software VOSviewer to make statistics on the academic journals included by Wanfang from 2010 to 2022, we analyze the number of published documents in this field, hot keywords and other information in recent years, and construct a keyword map. The results show that: 1) From the perspective of the number of documents published in the field of data mining applied in education and teaching in the past decade, it has roughly experienced a period of slow growth, growth and rapid growth; 2) The current research focus of data mining in the field of education and teaching is mainly on data mining technology, talent training, teaching reform, learning situation analysis and other fields; 3) In the future, industrial integration and innovation, information integration and data governance will be the research directions worthy of attention.

Keywords: Data mining; Education and teaching; Bibliometrics; Keyword heat analysis; VOSviewer;

This is a school-level project of Jiangxi Agricultural Engineering Vocational College: "Research on the Application of Big Data Technology in Education and Teaching" (Project No.: XJXM-21-05-05-05)

Introduction

Data mining technology appeared in the 1980s, and was limited by the level of information technology and computer facilities at that time. Until the beginning of this century, data mining technology has developed rapidly. The application research of data mining in education and teaching is also increasing.

Bibliometrics takes the external characteristics of documents as the research object, and uses statistical methods for data analysis and data mining, it can comprehensively display the development history, research status of this field. In recent years, domestic experts and scholars have carried out a lot of research on medical care, logistics and transportation, education and teaching and other fields by using bibliometrics.

This paper will adopt the bibliometric analysis method, take "data mining" and "education and teaching" as the research objects, carry out research from the research hotspots, development trends and other aspects, explore the development process of data mining in the field of education and teaching, and predict the future development direction of this field, with a view to providing reference for further research in this field.

1 Methods and data collection

1.1 Research methods

The study of knowledge carriers by quantitative methods has always been the basis of bibliometrics. We use VOSviewer software to study the distribution structure, quantity structure and change rule of the literature, focusing on the visualization of the analysis results. Compared with other similar software, VOSviewer is easier to learn in operation.

1.2 Data acquisition method

The data in this paper is sourced from Wanfang database, and is retrieved in a thematic way, with "data mining" and "education and teaching" as the joint search words. A total of 31149 search results are obtained. In order to help visual analysis, it is necessary to restrict the retrieval scope: 1) restrict the retrieval type to include only journal papers and conference papers; 2) The publication time is limited from 2010 to 2022. According to the above restrictions, a total of 1150 results were used as the data source for this study.

2 Bibliometric analysis

2.1 Analysis of the number of papers in different periods

According to the theory of bibliometrics, the growth law of the number of documents reflects the development level of a certain field to a certain extent. This paper makes statistics on the relevant documents of data mining applied in the field of education and teaching from 2010 to 2022. From the statistical results in Figure 1, it can be seen that the law of document growth conforms to the logic curve model, and the number of documents shows a rapid upward trend, which fully shows that with the continuous development of data mining technology, its application research in education and teaching has increased significantly. In combination with the number of papers issued, using the annual number of papers issued 50 and 100 as the critical value, the "application research of data mining in education and teaching" is further divided into slow growth stage, growth stage and rapid growth stage. See Figure 1 for details.

2.2 Keyword heat analysis

As the most prominent content tag of the document, keyword heat analysis is usually used to predict the core research topic and development trend in a certain field. In order to reduce the impact of non-hot tags and improve the accuracy of visual analysis, this paper

limits the frequency of keywords to 10 times, and filters out 54 keywords that meet the conditions. The “Network Visualization”, “Overlay Visualization” and “Density Visualization” are drawn respectively. This paper further counts the high-frequency keywords in this field. See Table 1 for details.

Due to the limited length of the article, only some high-frequency keywords are listed. Next, high-frequency keywords are classified according to the research topic. See Table 2 for details.

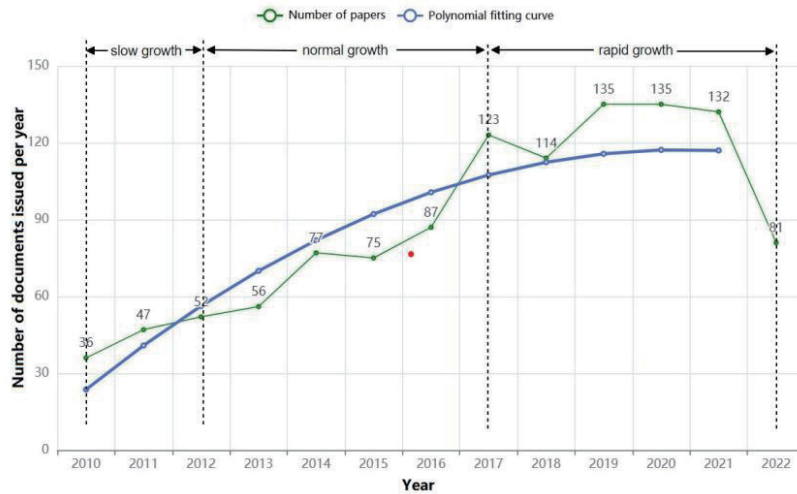
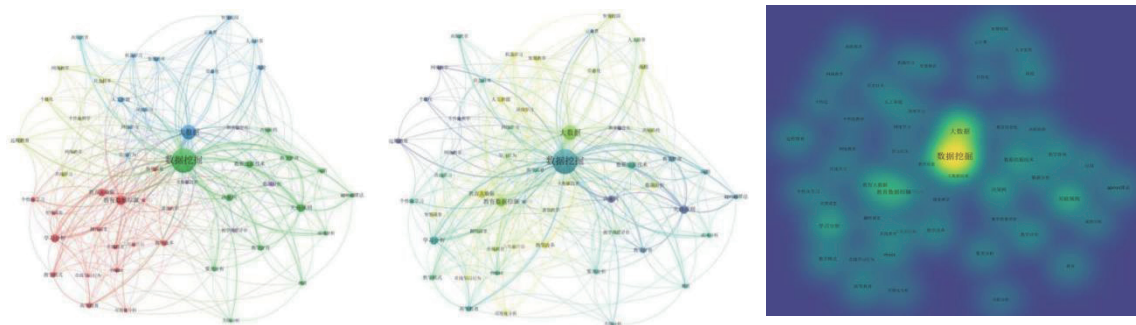


Figure 1 Statistics on the application of data mining in education and teaching from 2010 to 2022



(a) Network Visualization (b) Overlay Visualization (c) Density Visualization

Figure 2 Keyword atlas of literature from 2010 to 2022

Table 1 Application of data mining in education and teaching

Top 5 high-frequency keywords

no.	keyword	frequency	correlation strength
1	big data	176	217
2	Education data mining	75	81
3	Association rules	60	106
4	Learning analysis	51	88
5	Education big data	46	56

Table 2 Research topics and keyword clustering of data mining in the field of education and teaching

research topic	keyword clustering
data mining technology	association rules, decision trees, apriori algorithm, clustering analysis, machine learning, artificial intelligence
personnel training	teaching management, education big data, teaching evaluation, mooc, teaching mode
teaching reform	distance education, smart education, online education, personalized teaching, smart classroom, flipped classroom
learning situation analysis	score analysis, personalized learning, learning behavior

3 Frontier analysis

3.1 Research direction prediction

Clustering results based on keywords, combined with the research direction with high attention in recent years, the following prediction is made for the domestic

3.2 research direction.

Information integration:Information integration is the standardization of complex data forms to meet the needs of information integration and data sharing. At present, information integration is the development goal pursued by the government and the industry, and it is also a problem that data mining has been facing. In the future, promoting the pace of information resources integration and accelerating data fusion and sharing will become an area of continuous concern.

Cross-border data flow:With the successive promulgation of the Data Security Law and the Information Protection Law, together with the Network Security Law, the basic framework of China's cross-border data flow legislation system has been established. In the future, the global competition for data resources will become increasingly fierce, and large-scale cross-border data flow will become an inevitable trend.

Data governance:The goal of data governance is to improve data quality and maximize data value on the basis of reducing data management costs. At present, the research on data governance is still in its infancy. With the growth of national policy support and industrial demand, improving the management ability of massive data through data governance will become a key development area in the future.

4 Conclusion and prospect

This paper collects the research literature of data mining in the field of education and teaching in the past ten years, uses the knowledge map software VOSviewer to classify and count information such as keywords, analyzes the research hotspots and forecasts the research direction, and draws the following conclusions:

1) Policy guidance and promote the long-term development of the industry. In the past 12 years, the scientific and technological literature in the field of application research of data mining in education and teaching has increased year by year. With time as the boundary, it can be divided into slow growth stage, growth stage and rapid growth stage.

2) According to the results of visual clustering analysis, the current research hotspots can be summarized into the following four categories: ① data mining technology; ② Talent training; ③ Teaching reform; ④ Analysis of learning situation.

3)In the future, industrial integration and innovation, information integration and data governance will be the key development directions.

References

- [1]CHEN Quan. Discuss the development trend of AI and data mining[J]. Public Investment Guide,2021(35):146-148.
- [2]YIN Shuoxin. Bibliometric and visualized analysis of researches on gastric cancer stem cells[J].Chinese Journal of Tissue Engineering Research,2023,27(6):941-947.
- [3]XIAO Pengfei. Bibliometric Analysis of Global Aflatoxin Research[J]. Food Science,2022,43(15): 378-388.
- [4]YANG Jianjian, MA Xiaoqin, GUO Ran. Bibliometric analysis on data mining applied in domestic nursing field[J]. Journal of Nursing and Rehabilitation,2021,20(5):16-19.
- [5]LIU Li, FU Xianjun, LI Xuebo. Bibliometric Analysis of CNKI Traditional Chinese Medicine Data Mining[J]. Modernization of Traditional Chinese Medicine and Materia Medica-World Science and Technology,2020,22(11):3887-3893.
- [6]LAN Xu. Research on the Documentary Metrology and Knowledge Map Visualization of China's Traffic Culture Research in Recent 70 Years[J]. Public Communication of Science & Technology,2021,13(6):31-33.
- [7]ZHANG Kuangda, FAN Mingyue. Visualized Research on the Status Quo, Hotspots and Trends of Cross-border Logistics Research--Analysis of Knowledge Graph Based on CiteSpace[J]. Logistics Engineering and Management,2022,44(5):16-22.
- [8]ZOU X, YUE W L, LE VU H. Visualization and analysis of mapping knowledge domain of road safety studies[J].Accident Analysis and Prevention,2018,118: 131-145.
- [9]SHEN Lingyun, DONG Li. Knowledge map and basic trend of "the Belt and Road" transportation research - based on quantitative statistics of CNKI (2014-2017)[J].Journal of Qijing Normal University,2019,38(3):53-57.
- [10]WEN Xiaodong. Knowledge atlas analysis of research hotspots and trends of practical teaching in higher vocational education -- based on core journal data from 1993 to 2017[J]. Vocational & Technical Education Forum,2017(30):85-89.
- [11]ZHANG Yan,LI Yanzeng,SUN Dandan. Hotspots and Trends of Nursing Blended Learning in China[J]. China Continuing Medical Education,2021,13(28):104-108.
- [12]LIANG Yu, LI Nuo. Research topics, hotspots and frontiers of digital teaching resources in international Chinese education[J]. Overseas Chinese Education,2021(1):3-12.
- [13]CHEN Xinhang, LI Haichao, WU Hongbin. Research characteristics and development trend of postgraduate medical education in China[J]. Chinese Journal of Graduate Medical Education,2021,5(2):97-103.
- [14]WU Aizhi. Information Technology Progress and Bibliometrics Development[J]. Journal of Modern Information,2016,36(2):32-37.
- [15]JIN Peidong,KANG Zhongyuan.On the Mathematical Models of the Growth of Scientific Literature.Journal of the China Society for Scientific and Technical Information,2000,19(1):90-96.