



# Study on Document Measurement and Visual Analysis of Big Data Education Evaluation

Liang Yang<sup>1</sup>, Xu Luo<sup>1</sup>, and Ao Gao<sup>2</sup>(✉)

<sup>1</sup> Computing Center, Shenyang Normal University, Shenyang 110034, Liaoning, China

<sup>2</sup> National Engineering Research Centre of Fire and Emergency Rescue, Shenyang Fire Science and Technology Research Institute of MEM, Shenyang 110034, Liaoning, China  
gaoao@syfri.cn

**Abstract.** Machine learning, recommendation algorithm and other new big data technology, has become an important tool for learning analysis education evaluation. Educational evaluation is an important means to evaluate the value of educational activities, and big data technology is becoming an important part of educational evaluation. It has become an urgent problem to sort out the status quo of big data education evaluation research and clarify the context of its research development. In this paper, we use bibliometrics to analyze the data of 391 articles indexed by Web of Science in recent ten years, and use Cite Space software to generate visual results. Through the analysis of the results, it is concluded that big data technology is of great significance for educational evaluation.

**Keywords:** big data · education evaluation · educational activities · visual analysis

## 1 Introduction

Educational evaluation is an important research direction of pedagogy and an important means to embody educational value and evaluate educational level. Educational evaluation is defined as the process of judging the extent to which educational behaviors meet the needs of individuals and society, judging the value of educational activities, and expecting to realize the increase of educational value. Educational evaluation is not only a value evaluation activity, but also a scientific cognitive activity. The field of education has gradually attracted the attention of researchers. The value of big data does not lie in the data itself, but in the knowledge extracted from the data through data mining analysis. The popularization of big data technology in education makes it easier to obtain the values and rules of educational data through data mining [1]. The multi-dimensional and comprehensive nature of big data pushes educational evaluation from small sample data to evidentiary decision with comprehensive and full coverage data. The core value of education big data is reflected in improving the education process and increasing the education income. The reconstruction of the education evaluation system by big data provides an effective help for the scientific nature and rationality of education management and decision-making. With the application of big data technology to the

process of education evaluation, the connotation and extension of education evaluation are gradually changing. In terms of connotation, big data builds a systematic, scientific and diversified education evaluation system, following the laws of education essence and talent training. From the perspective of extension, the education evaluation system that conforms to the law of big data application pays more attention to the micro level of education and better reflects the basic requirement of people-oriented education.

Data visualization is the terminal link of the big data industry chain, which visually displays the results of data analysis. Compared with the traditional two-dimensional table, data visualization has the advantages of large information content and more intuitive performance, and can even realize real-time interaction with users. The application of visualization technology injects new vitality into bibliometric, which makes it easier to use analysis tools to explore research hotspots and trends in a certain field. Cite Space is a visual analysis tool based on Java language, which can sort out and display the research trends and their relationships of scientific literature. Cite Space can process the data downloaded from the current mainstream literature indexes at home and abroad, such as web of science, CSSCI and CNKI, conduct data analysis based on keywords at different levels from macro to micro, and visually present the analysis results.

Studies have shown that in 2012, research in the field of big data began to enter the practical application stage [13], and the research on big data education evaluation as a whole showed a steady increasing trend, and the research is still in a period of rising popularity [15]. In this paper, papers related to big data education evaluation are retrieved from the core collection of Web of Science database during the period from 2012 to 2022 as the research object. Download the paper record information to select full record and cited references, the file format is plain text. Cite Space software was used to analyze the downloaded documents, such as national cooperation network analysis, document co-citation cluster analysis and keyword emergence analysis. Sort out the development context of big data education evaluation research and find the relevant trends and hot spots. To provide reference and help for researchers engaged in related research, better grasp the direction of research development.

## 2 Research Methods and Data Sources

The research data comes from the core collection of Web of Science database, which is the most important citation database retrieval platform in the world. Its subjects cover a wide range of subjects and the data is detailed and accurate, providing effective support for interdisciplinary research such as big data education evaluation. The data retrieval strategy is as follows: (TS = (education evaluation or educational evaluation)) and TS = (big data), the literature type is “Articles or Proceedings Papers”, The time span was selected as “2012–2022”, and a total of 545 records were included. Then the category was limited to “Education Educational Research” and “Computer Science,” and 391 valid records were obtained after the search was removed. The exported file is a text file, and the retrieval time is June 15, 2022. The data obtained basically cover the important literature of related research in the past 10 years.

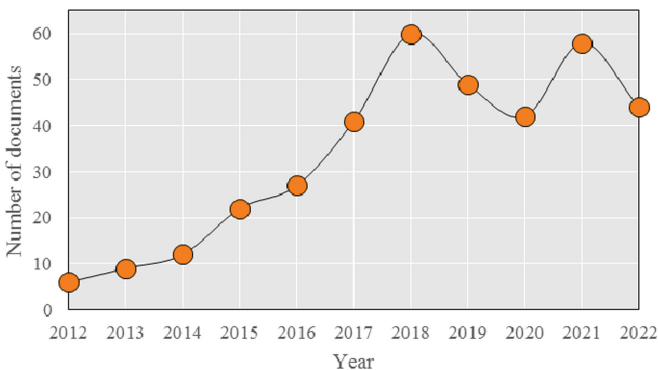
First of all, Excel2019 is used to carry out annual statistics on the number of published papers and draw data point broken line graph. Secondly, Citespace6.1. R2 software

was used to set parameters: Time Slicing (2012–2022), and each slice was 1 year. The selection of time slice, node shape and color are based on the principle of clear and accurate visual effect, and the default values of other parameters are used. (1) Based on the WOS data, Node Type is selected as country for the analysis of national and regional cooperation network, and Threshold parameter is set to 10. Here we can see the development of big data education evaluation related research in various countries, as well as the research cooperation between countries. For the visualization of national cooperation network, we can grasp the current research status of big data education evaluation from a macro perspective. (2) Cluster analysis of literature co-citation was conducted based on WOS data, and the results were presented in timeline view. The clustering results show the field of big data education evaluation, the knowledge basis of research and the mainstream research direction. (3) Perform keyword breakout analysis, set parameter Minimum Duration to 1, Parameter  $\gamma$  to 0.5, and other parameters to default values. Generate the statistical table of the time period and intensity of keyword outburst. Keyword emergent analysis is helpful to grasp the changes of relevant hot spots in the research field.

### 3 Results and Discussion

Excel files are directly exported from Web of Science, and frequency statistics are made for data according to “Publication Year”, and data point broken line graph is generated as shown in Fig. 1. The number of articles published increased from 6 in 2012 to 60 in 2019. The number of articles published in 2019–2020 decreased slightly, but the overall number remained above 40. Given the timing of the data download, the number of 2022 releases should be increased later. As can be seen from Fig. 1, the number of papers related to big data education evaluation has shown a trend of rapid growth in the past 10 years, especially in the last five years. It can be preliminarily judged that the research on big data education evaluation is still in a hot period.

Since 2012, 82 countries or regions have published papers related to big data education evaluation. Table 1 lists the 12 countries or regions with the largest number of documents. China ranks first with 160 articles, followed by the United States (45) and



**Fig. 1.** Statistics of periodicals published in 2012–2022

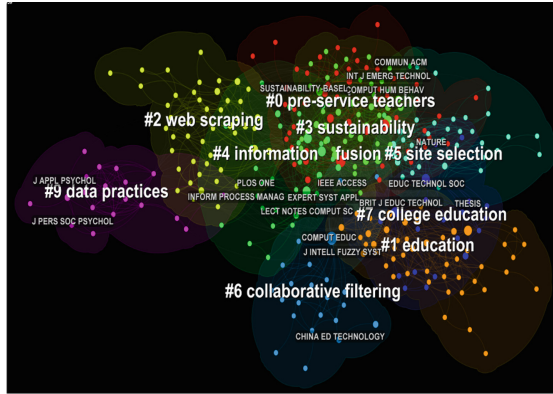
**Table 1.** Ranking of countries and regions in the number of papers published

Ranking	Countries	Count	Times Cited, All Databases	Year
1	CHINA	160	159	2013
2	USA	45	295	2012
3	SPAIN	22	30	2013
4	INDONESIA	11	10	2017
4	INDIA	11	20	2014
6	ENGLAND	10	87	2012
7	BRAZIL	10	12	2013
8	ITALY	10	12	2013
9	SAUDI ARABIA	9	87	2017
10	RUSSIA	7	7	2017
11	POLAND	7	30	2015
12	SWEDEN	7	35	2012

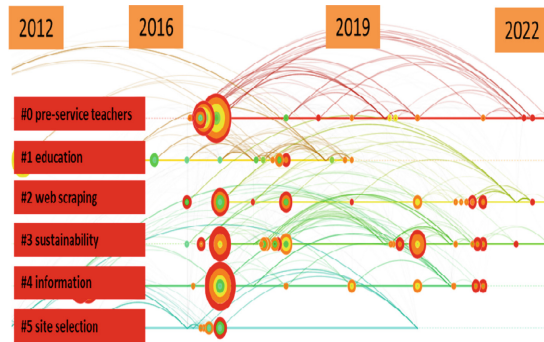
Spain (22), which shows that these three countries are in a leading position in the research of big data education evaluation. The number of citations of papers is an important indicator of the quality of papers. In the statistics of the number of citations of papers by the database of the whole network, the United States (295) is the highest, China (159) is the second, and Britain (87) and Saudi Arabia (87) are tied for the third place. It is noteworthy that although the number of papers published in the UK and Saudi Arabia as a whole is small, the papers have received a high degree of attention. According to the detailed data, the paper of Mehmood [11] in Saudi Arabia has been cited for the highest number of times among all documents, reaching as many as 74 times, including 73 times in the web of science database.

In order to solve the problem of big data application in education evaluation, the United States has earlier established a three-dimensional education data network covering the United States, including national, state, school district and school levels, and established education data standards to regulate and ensure data quality. In October 2020, China issued the overall plan for deepening the reform of educational evaluation in the new era, which proposed to use big data to generate evaluation tools, promote the classified evaluation of colleges and universities, guide the scientific positioning of colleges and universities, and create characteristics and levels. The data futures program developed by the British Bureau of higher education statistics is used to build a data architecture, speed up the collection of data on teaching quality, and provide a decision-making system for the reform of the higher education sector [16]. At the macro policy level, all countries are actively promoting the application of education big data and guiding colleges and universities to establish an education evaluation system in line with their own national conditions.





**Fig. 3.** Connection of Document co-citation Clustering



**Fig. 4.** Document co-citation Clustering Visualization

cluster #4 are two closely related links in the process of big data processing. Cluster #7 is a subset of cluster #1. Most of their studies focus on the related fields of pedagogy, and they are not closely related to other clusters.

The largest six clusters are selected and displayed in a timeline diagram, as shown in Fig. 4. Clustering network modularity  $q = 0.6111$ , weighted average contour value  $s = 0.8079$ , indicating that the clustering result is reasonable. The documents with the same clustering theme are arranged on the same horizontal line in chronological order, and the relatively large circle represents the highly cited documents in this category.

Clustering #0: Pre-service Teachers contains 56 nodes with a contour value of 0.783. The extracted subject words include deep learning, sentimental analysis, effective big data sharing prototype and natural language processing. The publication period is 2017–2022. Deep learning is a new research direction in the field of machine learning, which is very helpful for the interpretation of text, image, sound and other data. In the field of big data education evaluation, current research has used deep learning technology for emotional analysis to realize the correlation with curriculum evaluation [2]. There is

also research on the use of internet of things, supercomputing and deep learning technology to achieve the development and management of intelligent education environment [11]. Moreover, the deep learning algorithm based on natural language analysis realizes composition scoring [18] and other related research.

Clustering #1: Education contains 48 nodes, with an outline value of 0.818. The extracted subject words include learning analytics, artificial intelligence, clustering algorithm and software engineering competency, which are published from 2012 to 2019. Clustering algorithm is the most commonly used technology in big data analysis, which is used to gather similar data points. The main research content of this category is to apply different clustering algorithms to analyze learners' learning ability and professional skills, match industrial needs [6], conduct learning analysis [7], and realize the clustering and visualization of interactive education network resources [12]. Research related to AI technology includes monitoring and early warning of students' online learning performance [14], or evaluating the quality of education management of a discipline to help the discipline achieve intelligent teaching [5, 9, 20, 21].

Clustering #2: Web Scraping contains 47 nodes with a contour value of 0.749. The extracted subject words include covid-19 pandemics, process evaluation, data collection, recommended algorithm and collaborative filtering. The publication period is 2017–2022. The global pandemic of covid-19 has promoted the rapid development of online teaching and provided rich data sources and practical experience for big data education evaluation. Data collection is the premise of big data analysis. Only with enough magnitude and high-quality education data can we carry out accurate and standardized education evaluation. Recommendation algorithm can provide the most appropriate learning content for each learner through big data analysis [4]. Collaborative filtering is the most mainstream algorithm in recommendation algorithm at present. The research includes using deep learning and collaborative filtering technology to capture the data of online courses for process evaluation [17]. Use the recommendation algorithm to explore the relationship between the influencing factors of moral education evaluation [19]. And collect the data in the mixed learning environment, and use the results of learning analysis for teaching intervention [10].

Clustering #3: Sustainability consists of 45 nodes with a contour value of 0.712. The extracted subject words include machine learning, e-learning and virtual education, which are published from 2017 to 2022. S. M. Aslam et al. Summarized e-learning and machine learning algorithms from 1993 to 2020, providing ideas for online learning evaluation [3]. The research content also includes the teaching quality evaluation system based on face feature recognition [8] and the attention to students' privacy in the virtual education environment [22].

Clustering #4: Information contains 38 nodes with an outline value of 0.817. The extracted subject words include educational big data, educational data mining and subject knowledge graph, which are published from 2017 to 2022. Educational big data refers to the data set generated or actively collected in the process of complete educational activities. Bai et al. systematically summarized the prediction models and evaluation indicators in educational big data. The research content includes the method of automatically generating subject knowledge map based on educational big data. There are



also studies on data mining of educational big data to form a pre training model of College Students' development indicators, or to predict students' grades. Some studies have also used bootstrap based framework for mobile education big data mining to judge the learning effect of learners.

Clustering #5 Site Selection contains 37 nodes with a contour value of 0.82. The extracted subject words include academic big data, correlation analysis, community detection and disaster recovery. The publication period is 2014–2020. From a more macro perspective, this clustering studies the application of new technologies in the field of big data, which can also be applied to big data education evaluation. Some studies have focused on the rational evaluation and selection of academic big data information system disaster recovery (DR) sites. There is also research to explore the community structure composed of big data in different fields, and use potential correlation analysis and research to promote community detection, so as to achieve educational evaluation.

Through the analysis of the above clustering, we can see that the advanced technologies in the field of big data have been applied to the research of education evaluation. From the perspective of time dimension, 2017 is a key point, where the largest number of large nodes are arranged vertically, indicating that the research related to big data education evaluation was in an explosive period at that time. On the other hand, there is also a lot of overlap between these clusters, because many keywords in information technology are interrelated. For example, the field of artificial intelligence includes machine learning technology, and deep learning technology is the frontier of machine learning.

## 4 Conclusion

Using the bibliometric analysis method, this paper combs and counts the papers related to big data education evaluation on WOS, and finds that the research related to big data education evaluation has developed rapidly in the past decade, and entered the outbreak period after 2017. Combined with national and regional cooperation networks, it can be seen that China is currently in the lead in the number of papers published in this field, but the number of citations and the breadth of international cooperation are slightly low. Through literature co citation cluster analysis, we obtained the knowledge base related to big data education evaluation and found the frontier of current research. Finally, use keyword emergence analysis to grasp the research hot keywords. It provides a reference for future research on big data education evaluation.

**Acknowledgment.** This work was supported by Scientific Research Project of Shenyang Normal University (JS202014).

## References

1. Ang LM, Ge FL, Seng KP (2020) Big educational data & analytics: survey, architecture and challenges. *IEEE Access* 99:1
2. El-Demerdash AA, Hussein SE, Zaki JFW (2022) Course evaluation based on deep learning and SSA hyperparameters optimization. *Comput Mater Continua* 1. <https://doi.org/10.32604/CMC.2022.021839>



3. Aslam SM et al (2021) Feature evaluation of emerging E-learning systems using machine learning: an extensive survey. *IEEE Access* 9:69573–69587
4. Birjali M, Beni-Hssane A, Erritali M (2018). A novel adaptive e-learning model based on big data by using competence-based knowledge and social learner activities. *Appl Soft Comput*
5. Dai DD (2021) Artificial intelligence technology assisted music teaching design. *Sci Program* 2021: 9141339
6. Khousa EA, Atif Y (2018) Social network analysis to influence career development. *J Ambient Intell Humaniz Comput* 3
7. Liu F (2021) Design and implementation of intelligent educational administration system using fuzzy clustering algorithm. *Sci Program* 2021:9485654
8. Yuan F, Nie Y (2021) Online classroom teaching quality evaluation system based on facial feature recognition. *Sci Program* 2021:7374846
9. Qin J, Wang Y, Zhao Q, Tan L, Luo Y (2022). Evaluation of college students' ideological and political education management based on wireless network and artificial intelligence with big data technology. *Secur Commun Netw* 2022:5802525
10. Gong L, Liu Y (2019) Design and application of intervention model based on learning analytics under blended learning environment. *Information and Education Technology*
11. Mehmood R, Alam F, Albogami NN, Katib I, Albeshri A, Altowaijri SM (2017) Utilearn: a personalised ubiquitous teaching and learning system for smart societies. *IEEE Access* 5:2615–2635
12. Mohith M, Yi Z, Yeo Steve H, Song Jun S (2018). ClusterEnG: an interactive educational web resource for clustering and visualizing high-dimensional data. *PeerJ Comput Sci* <https://doi.org/10.7717/peerj-cs.155>
13. Ping Q, He J, Chen C (2017) How many ways to use CiteSpace? A study of user interactive events over 14 months. *J Assoc Inf Sci Technol* 68(5):1234–1256
14. Zhang T, Xiao W, Hu P (2022) Design of online learning early warning model based on artificial intelligence. *Wirel Commun Mobile Comput* 2022:3973665
15. Wang W, Lu C (2020) Visualization analysis of big data research based on Citespace. *Soft Comput* 24:2
16. Williamson B (2018) The hidden architecture of higher education: building a big data infrastructure for the 'smarter university.' *Int J Educ Technol High Educ* 15(1):12
17. Ziqiao W, Ningning Y (2021) Education data-driven online course optimization mechanism for college student. *Mob Inf Syst*. <https://doi.org/10.1155/2021/5545621>
18. Yuan Z (2021) 'RETRACTED: interactive intelligent teaching and automatic composition scoring system based on linear regression machine learning algorithm. *J Intell Fuzzy Syst* 40(2):2069–2081
19. Huafeng Y (2021) Apriori algorithm optimization based on Spark platform under big data. *Microprocess Microsyst*. <https://doi.org/10.1016/j.micpro.2020.103528>
20. Zhang J, Vijayalakshmi S (2021) Reform and innovation of artificial intelligence technology for information service in university physical education. *J Intell Fuzzy Syst* 2. <https://doi.org/10.3233/JIFS-189372>
21. Yu Z (2022) Intelligent transmission algorithm of college physical training course information based on big data. *Wireless Commun Mobile Comput*. 2022:4725406
22. Zeide E, Nissenbaum H (2018) Learner privacy in MOOCs and virtual education. *Theory Res Educ* 16(3):280–307

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

