

# Big data method and its application in innovation education research

Yankui Song<sup>1</sup>, Chuijiao Jie<sup>2</sup> and Zhijin Xu<sup>2\*</sup>

<sup>1</sup>Department of College English, Capital Normal University, Beijing, 100048, China

<sup>2</sup>Department of Management, Capital Normal University, Beijing, 100048, China

**Abstract.** Big data technology is a new stage of information development. In recent years, it has been widely used in many fields, especially in social science research. This paper analyzes the development status and significance of the combination of big data technology and social science research, on the basis of summarizing and combing the concept of big data and its important role. Taking the application of big data method in the research of innovation education as an example, this paper makes a series of visualization analysis with Citespace software on the related literature with the theme of “big data and innovation education” collected by CNKI, such as annual analysis, literature source analysis, co-occurrence analysis of authors, organization analysis, keyword clustering analysis and keyword timing analysis. This paper also draws the corresponding knowledge mapping, clarifies its research status, hot spots and development trend, and provides scientific basis for the research of innovation education. Thus the paper believes that the research on big data and innovation education needs to strengthen interdisciplinary communication and cooperation, refine and deepen the research theme and content.

## 1 Introduction

With the rapid development of information technology, human beings have entered the “big data era” of data explosion. Big data technology is a new stage in the development of informatization and a “sharp tool” to boost economic development, improves social governance and people’s lives. In recent years, big data technology and methods have attracted more and more attention in humanities and social sciences, especially in literature, history, economics and Marxist theory, which are in the ascendant and have achieved many important research results [1]. How to mine and apply massive data by big data technology has become one of the focuses of academia, especially in social science.

This paper analyzes the development status and significance of the combination of big data technology and social science research, on the basis of summarizing and combing the concept of big data and its important role. Taking the application of big data method in the research of innovation education as an example, this paper makes a series of visualization analysis with Citespace software on the related literature with the theme of “big data and innovation education” collected by CNKI, such as annual analysis, literature source analysis, co-occurrence analysis of authors, organization analysis, keyword clustering analysis and keyword sequence analysis.

## 2 Concept and function of big data technology

Big data is a major technological breakthrough in research. McKinsey, a well-known global consulting company, first proposes the arrival of the era of big data, in its report *Big data: the next frontier of innovation, competition and productivity* (2011), defines big data as a data set whose size exceeds the ability of conventional database tools to acquire, store, manage and analyze. At the same time, McKinsey emphasizes that the data set does not necessarily exceed a specific TB level in order to be regarded as big data. IDC (2011) defines big data technology as a new generation of architecture and technology that can economically extract value from high frequency, large capacity, different structures and types of data. As for its concept, people have put forward “3V”, “4V” and “5V” successively, and its understanding is consistent [2]. Big data technology methods mainly include data clustering, data mining, spatial analysis, time series analysis, visualization technology and others [3].

Big data technology and methods play an important role in “explaining the past, predicting the future, and making present decisions” [4]. Compared with traditional data analysis methods, big data technology analysis has the following two characteristics or advantages: on the one hand, big data analysis transcends the limitations of sampling analysis and small amount of data. Big data analysis no longer adopts sampling analysis method, but

\*Corresponding author’s e-mail: xzjccc@sina.com

directly analyzes all samples with higher accuracy of analysis results instead; on the other hand, big data analysis can dynamically track data [5]. Social science problems often have the characteristics of real-time and evolution, so it is difficult for traditional data analysis methods to track and feedback in real time. Big data and its technology system provide technical feasibility for the complex adaptive characteristics of effective human social activities, so as to provide strong support for promoting social sciences to learn from natural science achievements and form a new model of social sciences research based on data driving.

Big data technology and methods have developed rapidly in recent years. It is widely used in many fields and industries, such as commerce, manufacturing, transportation, communication, medical treatment, education, finance, especially in social science research. Wal-Mart was the first to analyze sales by big data technology, and found the matching goods suitable for trading together, creating a classic business case of “beer and diapers” [6].

### **3 Situation and significance of the combination of big data and social science research**

The combination of data science and humanities & social sciences, a new research is in the ascendant and has broad prospects. In 2015, the Fifth Plenary Session of the 18th CPC Central Committee adopted the “proposal of the CPC Central Committee on formulating the 13th five-year plan for national economic and social development”, which first proposed the implementation of the national big data strategy and the network power strategy. In the same year, the State Council issued “the action outline for promoting the development of big data”. The theme of the second collective learning conference of the Political Bureau of the CPC Central Committee after the 19th CPC National Congress is the national big data strategy. Since then, there has been a rapid emergence of application findings on the combination of big data technology, philosophy and social sciences in domestic academic circles. The year 2015 is also called “the first year of big data application” by many domestic industry insiders [7].

In recent years, big data technology and methods are more and more widely used in the fields of literature, history, anthropology, sociology and Marxist theoretical research in the multi-dimensional and multi-level aspects of research methods, research problems and research subjects. They have made many extraordinary scientific research achievements, formed increasingly important academic influence, and are profoundly changing people’s production mode and consumption mode and the way of thinking. Some scholars believe that big data is the complementary knowledge of empiricism and rationalism, which not only develops the understanding of causality, but also emphasizes the tolerance of uncertain results, and deepens the cognition of uncertainty [8]. Based on the literature of science & technology and philosophy in the past 60 years, some scholars have introduced the theory and method of scientific knowledge mapping to conduct

scientometrics research on the development of philosophy of science & technology. Through flexible use of various tools to conduct in-depth mining of literature data, they have described the macro, meso and micro prospects of the development of science & technology and philosophy in China [9]. In fact, the past form of “spoon feeding” education has far been unable to meet the humanized demand of people in the big data era nowadays. We must vigorously develop the “internet+ education” mode, make full use of the internet to understand teachers and students mutually and timely, and promote the personalized development of continuing education [10].

## **4 Application of big data method in innovation education research**

In recent years, how to apply big data and give full play to the role of big data analysis methods authentically in the field of innovation education has aroused wide spread concern in academic circles. A series of quantitative and visual knowledge mapping analysis by Citespace software, such as annual literature analysis, literature source analysis, organization analysis, keyword clustering analysis and keyword sequence analysis, are carried out to grasp the research hotspots and development trends, so as to provide scientific reference for the research of big data and innovation education.

### **4.1 Data sources and research tools**

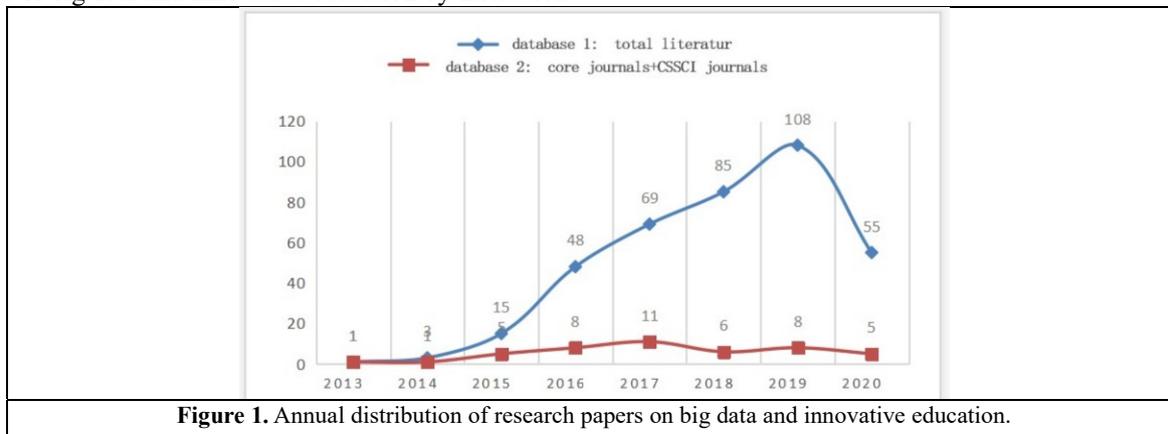
On the CNKI database page, click “advanced search” and take “(subject = innovative education or title = innovative education or v\_subject = Chinese and English extension (innovative education) or title = Chinese and English extension (innovative education)) and (subject = big data or title = big data or v\_subject = Chinese English extension (big data) or title = Chinese English extension (big data))” as the search condition. After eliminating useless literatures, 378 eligible literatures are selected and exported in Refworks format recognized by Citespace, named as database 1. On the CNKI database page, click “advanced search” and take “(core journal = y or CSSCI journal = y) and ((subject = innovative education or title = innovative education) or v\_subject = Chinese and English extension (innovative education) or title = Chinese and English extension (innovative education)) and (subject = big data or title = big data or v\_subject = Chinese and English extension (big data) or title = Chinese and English extension (big data))” as the search condition. After eliminating useless literatures, 45 eligible literatures are selected and exported in Refworks format recognized by Citespace, as in database 2. The retrieved deadline is Mar. 12, 2021.

This paper selects the software of Citespace as research tools [11]. It is suitable for the quantitative and visual analysis of titles, authors, organizations, keywords, word frequency, co-occurrence and citation information of the literature, and then draws the corresponding knowledge mapping.

## 4.2 Visual analysis of research status

This paper analyzes the annual distribution of research literature by Citespace, the status quo of institutional cooperation and the research hotspots and development trends of big data and innovative education by institution

and keyword co-occurrence function, and then draws a visual mapping.



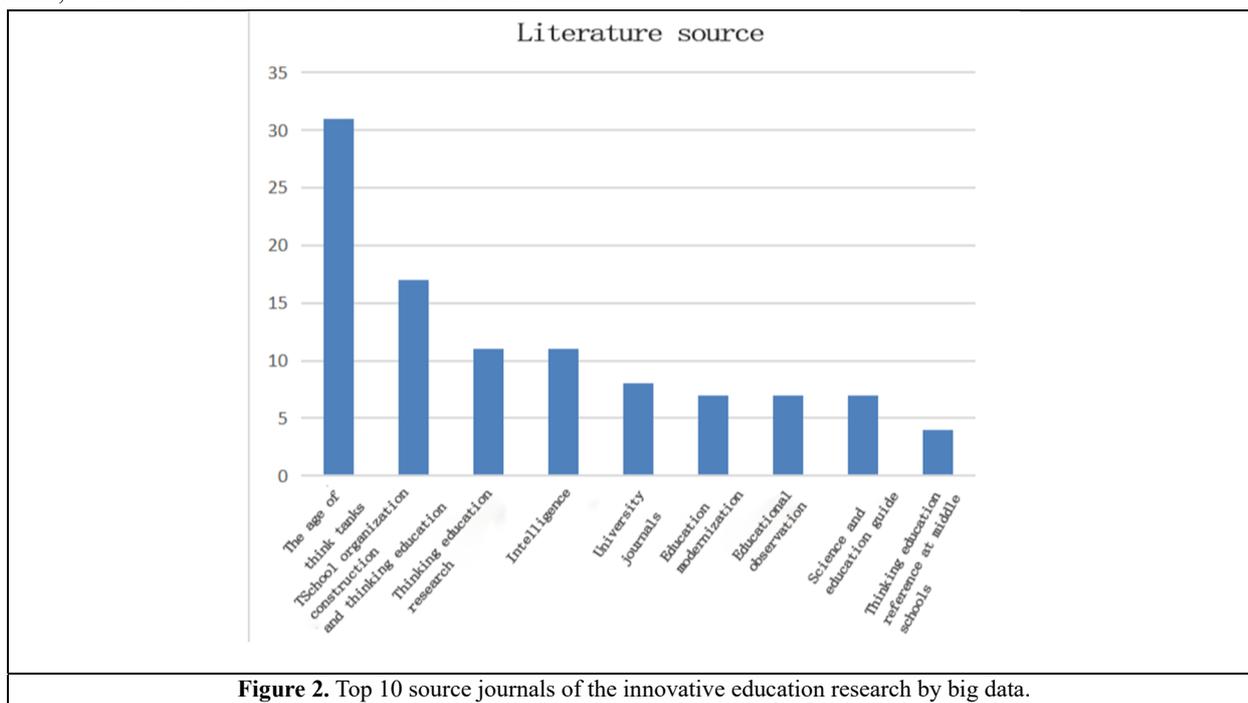
### 4.2.1 Annual analysis of research literature.

According to the year of publication, the literatures in database 1 and database 2 are statistically analyzed, and the publishing trend chart is drawn in Figure 1. We can see that the total literature with the theme of “big data and innovative education” and the literature published in core journals or CSSCI journals began in 2013, only accounting for 1.18%, which indicates that the research quality is not high and it is in the exploratory period. The general trend has been increasing since 2013, a slow growth period from 2013 to 2015 and a fast growing period from 2015 to 2019. In early 2020, the outbreak of Covid-19 in China caused great concern in academic circles, which distracted some scholars’ research efforts.

Consequently, the number of documents on the theme of “big data and innovative education” decreased, but on the whole the research on that theme is still on the rise.

### 4.2.2 Analysis of literature sources.

Based on the statistical analysis of literature sources in database 1, the top ten literature sources are ranked from high to low in Figure 2. The top ten literature sources are mainly in journals which focused on innovative education by big data. Among them there are 31 articles, the largest number, published in Age of Think Tanks. These journals provide academic exchange platform for the innovative education research by big data, and reflect its latest research trends.

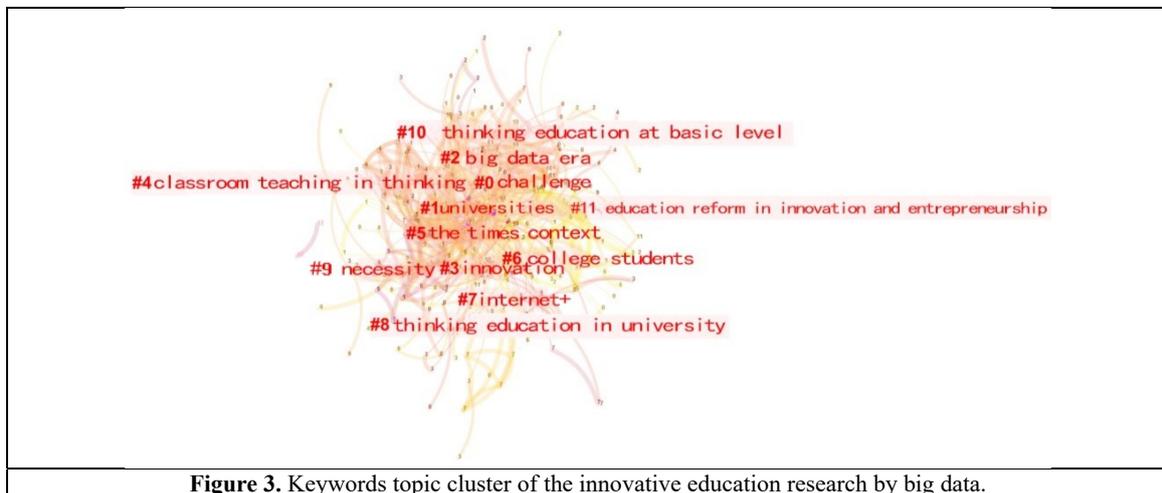


**Table 1.** Institutions of the innovative education research by big data.

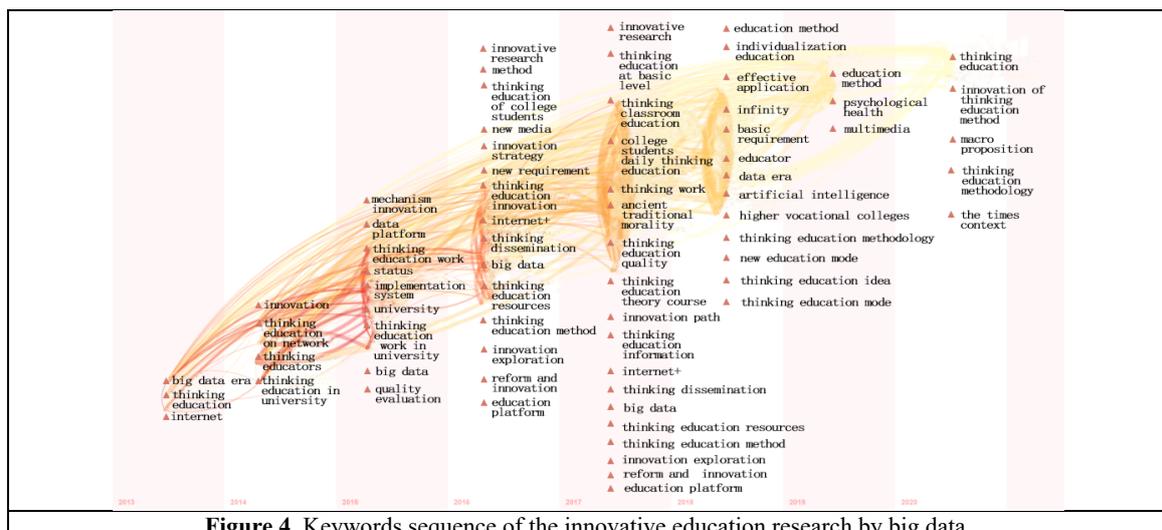
Serial number	Name of organization	Number of papers	Serial number	Name of organization	Number of papers
1	Central China Normal University	12	6	Lanzhou University	4
2	University of Electronic Science & Technology	8	7	University of Defense Science and Technology	4
3	Northeast Normal University	4	8	China West Normal University	4
4	Nanjing Institute of Political Science	4	9	Renmin University of China	3
5	Southwest University	4	10	Harbin Normal University	3

**4.2.3 Institutional Analysis.** According to the distribution of institutions in CNKI group browsing, the top ten institutions are selected, and its paper number of big data and innovative education are drawn in Table 1. Among them Central China Normal University obtains the most scientific research achievements, published 12 articles in big data and innovative education.

**4.2.4 Research hotspots and trend analysis.** In the CX project of Citespace, Time Slicing is selected from Jan. 2013 to Dec. 2020, Years Per Slice for 1 year, Node Types for Keyword, and other parameters remain in the default state. Click Go! and K (Label clusters with indexing terms) to draw the keyword topic clustering of big data and innovative education research literature in Figure 3.



**Figure 3.** Keywords topic cluster of the innovative education research by big data.



**Figure 4.** Keywords sequence of the innovative education research by big data.

According to Figure 3, 12 major clusters of “big challenge”, “big university”, “big data era”, “innovation”, “classroom teaching”, “era context”, “university student”, “internet+” and “innovation & entrepreneurship education

reform” are found in big data and innovative education research.

Each cluster represents a research topic of big data and innovative education. Combined with the top five tag

words and the high-frequency literature in each topic, this paper describes the research topics of big data and innovative education. On the basis of Citespace keyword topic clustering, select Timezone, calculate Node Labels according to By Degree, and select 5 as Threshold. Nodes are adjusted appropriately to ensure the clear display, and the keywords sequence of big data and innovative education is drawn in Figure 4.

According to the sequence chart in Figure 4, big data and innovative education research can be divided into three stages, namely, the budding period in 2013, the development period from 2014 to 2018 and the deepening period from 2019 to 2020. Big data and innovative education research has gone through three stages of combination: “can or cannot - where - how to”. At present, most research of big data and innovative education focuses on multiple topics, and discusses how to carry out innovative education from multiple themes and a macro perspective with big data method.

## 5 Conclusions

The paper analyzes 378 documents on the theme of big data and innovative education collected by CNKI, including a series of visual knowledge mapping of annual analysis, source analysis, organization analysis, keyword clustering analysis and keyword sequence analysis. It provides quantitative, visual and dynamic scientific support to clarify the research status of big data and innovative education in this field. Due to the short time of big data research, there is still a large space for the development of big data and innovative education research. There is less research on how to put the proposed method or model into practice, such as how to predict students' innovative behavior by analyzing students' past behavior habits and other specific issues through big data. We should identify the combination of big data and innovative education, strengthen interdisciplinary exchanges and refine the research content in the future.

In short, big data technology and methods will be more and more widely used in more fields, especially in social science research with the development of economy, technology and society. The research on big data and innovation education needs to strengthen interdisciplinary communication and cooperation, refine and deepen the research theme and content, in order to show broader and stronger application value in the process of “explaining the past, predicting the future, and making present decisions” in big data technology development. This interdisciplinary research has prominent characteristics, obvious advantages, great significance and broad development prospects.

## Acknowledgement

This paper was sponsored by 2020's construction project of university think tanks and social service ability in Capital Normal University “Big data mining and analysis of Marxist symbiosis thought” (SK2020ZK03).

## References

1. Zhang, S., Jiao, D. (2019) 2nd symposium of big data, philosophy and social sciences: let data open the door of good governance and link to a better life. *Guangming Daily*, June 18.
2. Yin, M.Y. (2019) Big data application and analysis. *Technology and Innovation*, 04: 31–39.
3. Manyika, J., Chui, M., Brown, B., et al. (2018) Big data: the next frontier for innovation, competition and productivity. [http://www.mckinsey.com/insights/business\\_technology/bid\\_data\\_the\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/insights/business_technology/bid_data_the_next_frontier_for_innovation).
4. Yu, L. (2013) Big data and big data economics. *China Soft Science*, 07: 23–31.
5. Luo, X.Y., Huang, X.R. (2017) Big data method in social science research. *Journal of Systems Science*, 04: 57–45.
6. Gao, Y. (2008) *Beer and diaper*. Tsinghua University Press, Beijing.
7. Du, X.Y. (2018) How can big data become a national strategy? *Financial Economy*, 11: 17–25.
8. Pan, P., Zheng, H., Lan, L.S. (2015) Essential characteristics of big data system and its philosophical reflection. *Journal of Systems Science*, 03: 13–21.
9. Zhang, D.J. (2017) Big data analysis of the development trend of Chinese philosophy of science and technology. University of Science and Technology of China, Beijing.
10. Yu, H.Y. (2019) Application of management information system in continuing education. *E-commerce*, 04: 51–59.
11. Zhan, P.W., Xie, X.J. (2018) Overview of big data system and key technologies and tools. *Network Security Technology and Application*, 08: 29–37.