Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 15, No. 3, 2024

CiteSpace-based Big Data Processing Technology Research Hot Spots and Frontier Analysis

Huiqin Sun, Gaorui Zhang, Xiaohong Wang *

School of Information Engineering, Wuhan Business University, Wuhan 430056, China
* **Corresponding author**: Xiaohong Wang (Email: 241544488@qq.com)

Abstract: The objective of this study is to present a visual representation of the literature on big data processing technology (2020-2024) in the China Knowledge Network (CNKI) and to conduct a detailed analysis thereof. The analysis is based on the community structure analysis software CiteSpace, which allows for a comprehensive examination of the institution that publishes the relevant literature, the authors of the literature, and the keywords. The results of the analysis demonstrate that the primary research foci in the field of big data processing technology are the analysis of data utilization in societal contexts and the development of techniques for processing and mitigating potential risks. In the future, the advancement of big data will witness a diversification of trends. As artificial intelligence, cloud computing, and other technologies continue to evolve, big data processing technology will increasingly prioritize intelligence and automation to enhance the efficiency and precision of data processing.

Keywords: Big data technologies; Data visualisation; Research hotspots; CiteSpace

1. Introduction

In the context of the accelerated development of science and technology, the processing of big data has emerged as a pivotal driver of scientific and technological advancement [1−2]. This has led to a heightened focus on big data technology across a range of sectors. However, the current research landscape surrounding big data technology and its applications remains complex and challenging. Data's sheer volume, velocity, diversity, and low-value density present significant challenges in effectively managing and interpreting them.

CiteSpace is founded upon the tenets of co-citation analysis and path-finding network algorithms, which are employed to quantify the corpus of literature within a specific field, thereby facilitating the creation of a visual representation of the knowledge domains [5−6]. CiteSpace possesses a distinctive advantage in the processing of large data sets [7]. In particular, the multi-database support, deep mining capabilities, and other aspects of the software's functionality assist researchers in identifying the evolution of scientific knowledge and the current state of research. The utilization of CiteSpace for the analysis of big data processing technology offers a novel perspective and methodology. The core functionality of the CiteSpace visualization software is its capacity to analyze literature co-citation [8]. This enables the researcher to identify the connecting relationships between research topics, locate key literature, and explore frontier hotspots. This approach is particularly suited to the analysis of big data technology.

This paper is based on CNKI's corresponding journals and a systematic review of China's big data technology literature. The literature was searched using the keyword "big data processing technology," which was then analyzed using CiteSpace to identify technology hotspots. To facilitate a comprehensive analysis of the community, a variety of techniques, including graphical representations and other methods, are employed to examine the scientific and intuitive aspects of big data processing technology. This encompasses data visualization, machine learning, and other methods for systematic analysis. The objective is to provide a foundation for subsequent technological advancement and scientific and technological innovation.

2. Data processing

The data employed in this study were sourced from the China Knowledge Network (CNKI) database, which is a more authoritative indexing tool for scientific and technological literature in China, and its data sources are widely used in bibliometric research. The statistical year is set as 2020-2024, with 'big data processing technology' as the theme, and a total of 1,353 articles were obtained through manual screening.

2.1. Data processing

Table 1. CNKI Literature Search Refinement Table

Retrieve key items	CNKI
search term	Big data processing technology
periodicals	SCI,CSSCI,Beida Core,CSCD,AMI
Search method	Topic Search
search term	2020-2024
Search Field	data science
Number of valid papers (articles)	1353

The retrieved CNKI documents should be exported as Refworks, with the file name comprising the string "download_xx+serial number." The retrieved CNKI documents should then be imported into CiteSpace 6.3. R1. The 'TimeSlicing' option should be set to '2020-2024', the 'YearPerSlice' option to '1', and the 'Pathfinder', 'Pruning', and 'Pruningslicednetworks' options to their default values. The remaining options and parameters should be left as they are. To obtain a variety of knowledge graphs and identify relevant nodes for the analysis, the WoS literature in the NodeTypes was selected for visualization. This included the author,

institution, and keywords.

3. Analysis of results

3.1. Analysis of cooperation with issuing bodies

Figure 1 illustrates the network diagram of the relationship between research institutions in China. It can be observed that communication and cooperation between domestic research institutions is more frequent, which plays an indispensable role in the development of big data technology. Among the institutions, the University of Chinese Academy of Sciences, Peking University, and the Chinese Academy of Sciences National Astronomical Observatory have published papers on big data technology, indicating that these universities are engaged in national scientific research on big data technology. Furthermore, the influence of big data in colleges and universities is worthy of note. Conversely, given the relatively limited number of institutions, the potential for big data technology is considerable. It is therefore recommended that universities expand their provision of big data technology to facilitate the training of talent [9−10]. To further advance the development of big data technology, it is advised to reinforce the collaboration between universities and research institutes and to establish interdisciplinary and inter-institutional joint research teams [11].

3.2. Network analysis of co-author relationships

To gain an understanding of the core authors in a given research field, it is necessary to consider both the number of academic papers they have published and the frequency of citations to their work. As illustrated in Figure 2, the number of research teams in this field is relatively limited. For instance, the team comprising Ji Qingchang, Zhang Dongmei, Yin Zhaowu, and Qiu Yingjie has published two papers per member. About the number and distribution of authors' partnership networks, the authors are more widely dispersed, except for the teams and small network partnerships that have been established. Scholars in China's related data field have been encouraged to increase their investment in this scientific research in recent years. They should facilitate timely and relevant exchanges between peers, align their work with technological development trends, and continuously innovate and improve in order to meet the future needs of society for big data [12−14].

Fig. 1 Mapping of cooperation networks of issuing institutions

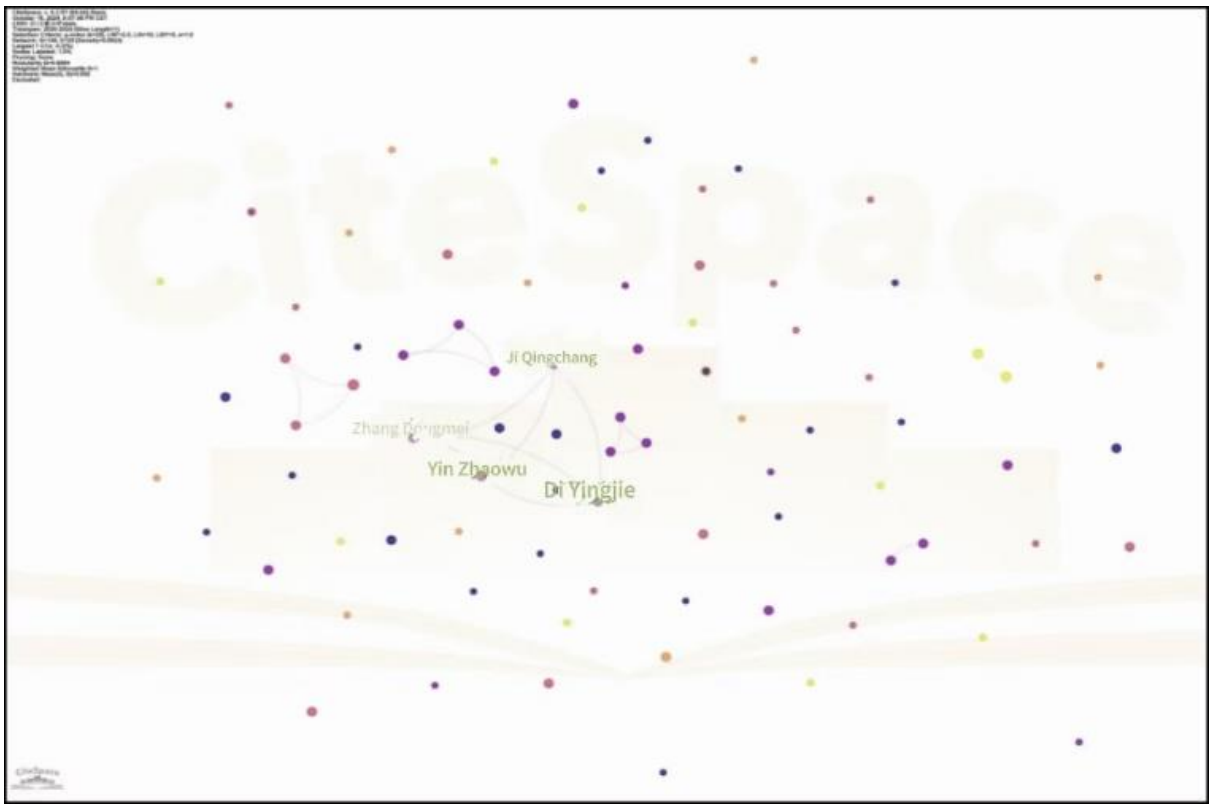

Fig. 2 Mapping of author collaboration networks

3.3. Analysis of Research Hot Spots and Research Frontiers

Keywords are a comprehensive summary of the research content of the literature, and to a certain extent, can represent the primary research content of a piece of literature. Network node centrality[15] is a significant indicator of the core node in the Chengdu network. Through the utilization of CiteSpace for the visualization and analysis of literature related to big data technology in the CNKI database, Figure 3 illustrates that, concerning the front-end of big data processing technology, the computer remains the primary focus, followed by data collection, data processing, cloud computing, machine learning and blockchain and other related data processing technologies. In terms of machine learning technology's application to data processing, the primary objective is to undertake data processing, construct models, and evaluate these models. This is done to predict potential data scenarios that may emerge in the market[16-17]. A cloud computing data center is capable of carrying out The processing of a substantial quantity of data, the utilization of data platforms and services, software, and services, and other related elements, enabling the virtualization of data to a high degree[18-19]. This facilitates the advancement of enterprises and the optimization of enterprise management and accounting decisions[20]. In addition to the aforementioned technology areas, emerging technologies such as artificial intelligence, the Internet of Things, and edge computing are gradually becoming research focal points. The advancement of these technologies offers novel concepts and methodologies for big data processing. For instance, the deployment of artificial intelligence technology[21-22] in data mining and pattern recognition, and the utilization of IoT technology[23-24] in data acquisition and real-time processing. Furthermore, edge computing[25] can be employed to reduce the latency of data transmission and enhance the efficiency of data processing. Additionally, edge computing can be integrated with various aspects of life, such as the power distribution network[26]. A comparison of Figure 4 (which illustrates the historical evolution of each keyword) with the preceding analysis reveals that the research focus has consistently been on big data, data collection, data processing and cloud computing. The extensive use of big data in everyday life will undoubtedly lead to further in-depth development of these data processing technologies, which will become a prominent research area in the community. It is likely that there will be significant advances in data processing technology as a result.

Figure 5 presents the keyword prominence mapping of CNKI, which has been analyzed by combining it with the color change of each keyword node in Figure 4. The period under consideration is from 2020 to 2024. The keywords "application analysis," "processing," and "countermeasures," along with other keywords, have been the subject of frequent searches, with the highest search degree reaching 0.76. This indicates that big data processing technology has been a significant focus in both academic and industrial contexts. Extensive attention has been devoted to the application of big data in business intelligence[27-28], the application of big data in image processing, the analysis and mining of big data, the analysis of image processing principles, and the solving of image processing work with advantages[29-31]. Additionally, big data technology in data visualization has been shown to possess notable advantages. The visualization of urban flooding disaster data[32] is one example of the application of data visualization in various fields. The research preface is distinct from the research hotspot in that it reflects the

evolving research frontier of a field. This is a set of dynamic concepts and potential research problems that are highlighted and appear with increasing frequency in specialized terminology.

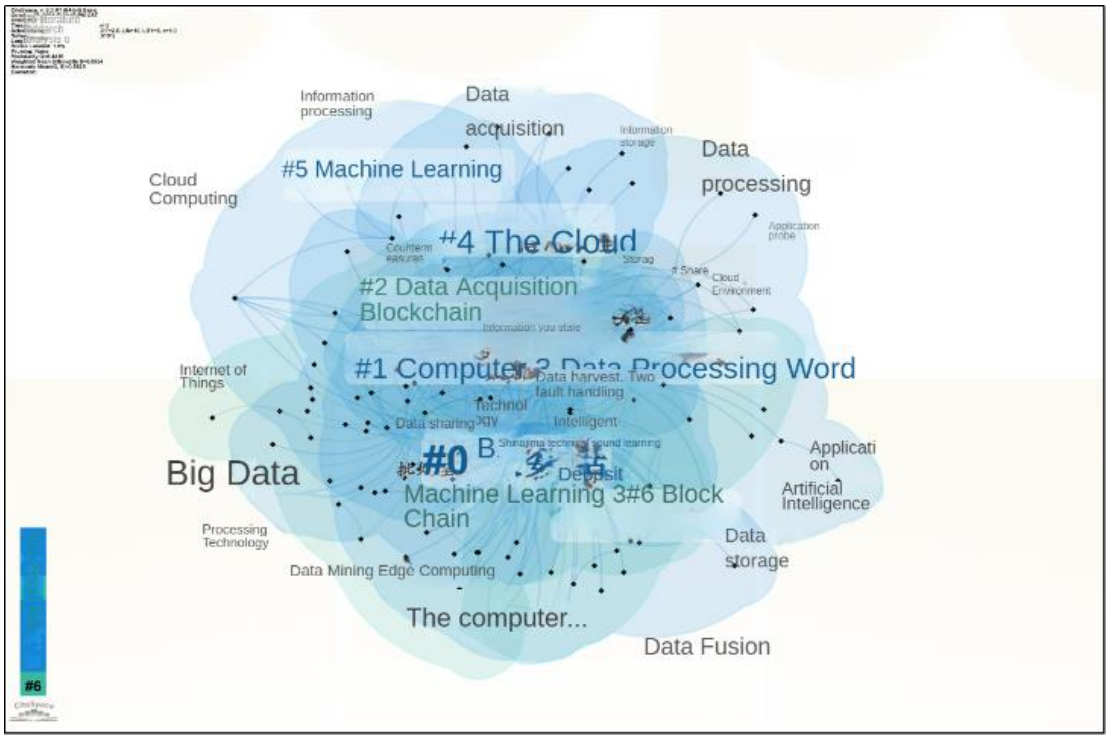

Fig. 3 CNKI database keyword co-occurrence network clustering map

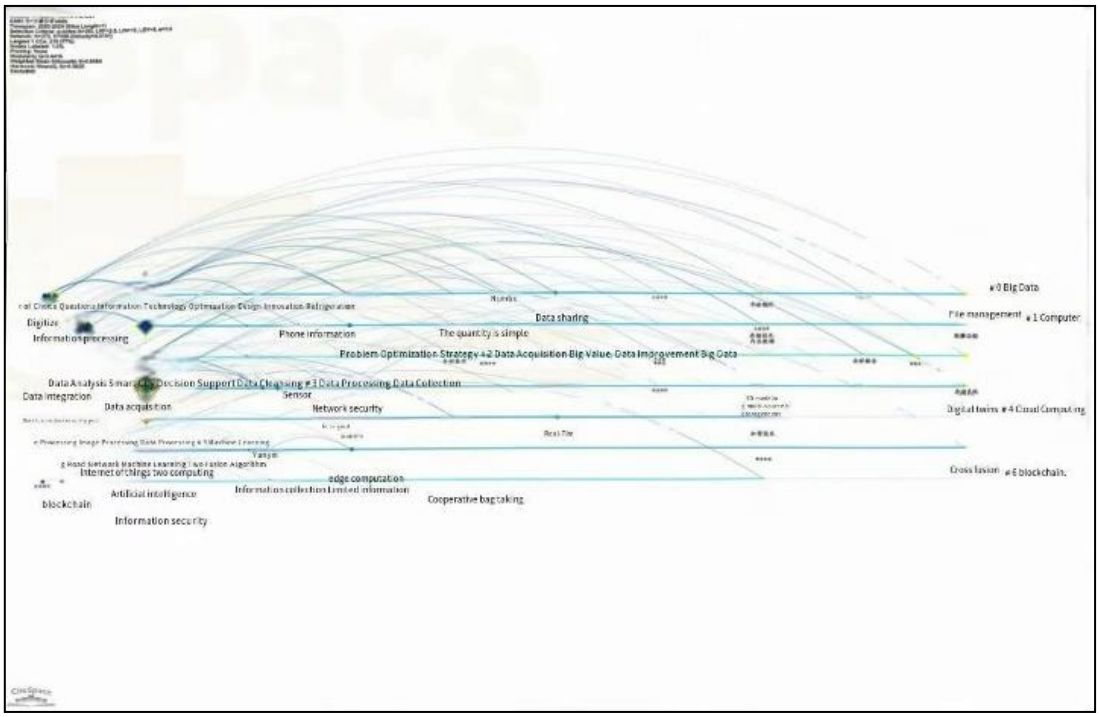

Fig. 4 CNKI keyword evolution mapping

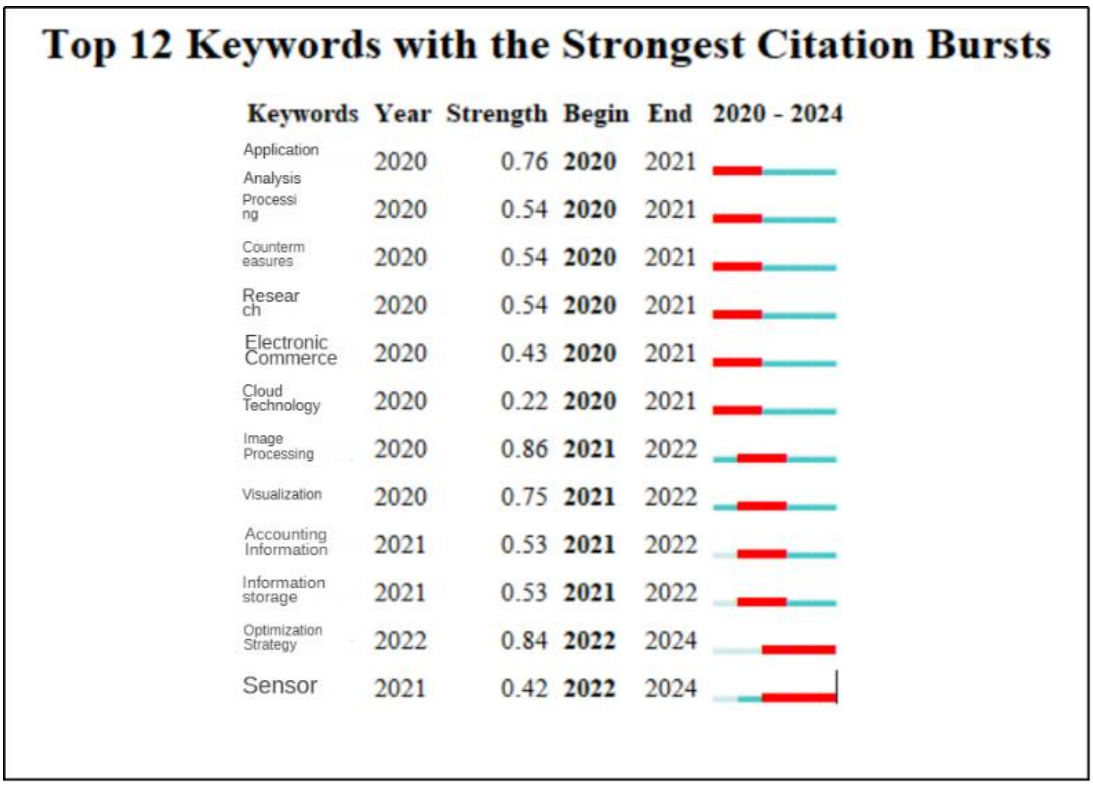

Fig. 5 CNKI keyword emergence mapping

4. Conclusion

In this paper, we conducted a comprehensive literature review of the latest research on big data processing technology in the CNKI database, spanning 2020 to 2024. We employed the CiteSpace community discovery approach to analyze the literature and identify key insights. Our analysis yielded the following findings:

(1) China's research institutions are primarily universities and research institutes, notably the National Academy of Sciences, Peking University, and the National Astronomical Observatory of the Chinese Academy of Sciences (NAO). However, these institutions are relatively few, as are the research projects and related literature they publish. Furthermore, the researchers and research groups affiliated with these institutions are geographically dispersed.

(2) The research areas of greatest interest in the field of big data-related processing technologies include the following: big data, data collection, data analytics, cloud computing, machine learning, and blockchain. The research focus is on the analysis of applications, data processing, and the development of appropriate countermeasures.

(3) The frontiers of big data application areas encompass a range of fields, including big data and analytics, big data technologies in business and enterprise management, the financial industry, and commercial banking. Additionally, research on big data itself represents a significant area of exploration within this field.

The above research demonstrates that big data processing technology is still in the developmental stage of application in many fields, exhibiting considerable potential and value. Future research should prioritize interdisciplinary integration and facilitate the integration of theory and practice to address the complex challenges inherent to big data processing. This study elucidates the research priorities and cutting-edge analyses of big data processing technologies, offering scholars a robust foundation for further research on big data technologies.

Acknowledgements

This paper was funded and supported by the 2024 School-level Innovation and Entrepreneurship Training Program of Wuhan Business University (No.202411654184).

References

[1] Mao M. Hot spots and frontiers of research on domestic and international big data technology applications - A CiteSpace visualization analysis based on literature from 2013-2018 [J]. Operation and Management, 2020, (03): 10-13. DOI: 10.16517 / j.cnki.cn12-1034/f.2020.03.004.

[2] LIU Zhen, ZHANG Lirong. Will big data technology bring the end of capitalism? [J]. Economy, 2024, (06):40-50. DOI: 10.16528 / j.cnki.22-1054/f.202406040.

[3] Ning Chao. Application of big data technology in smart city[J]. New City Construction Technology,2024,33(06):22-24.

[4] Yang Xinyi, Su Xinning. Domain Knowledge Structure Cognition--Applicability Analysis Based on Big Data Environment[J/OL]. Library and Intelligence Work,1-13[2024-10-23].

https://doi.org/10.13266/j.issn.0252-3116.2024.00.000.

[5] YANG Rongmei,NIU Saifei,TENG Dongmei. Knowledge mapping of carbon disclosure research in China - a CiteSpace-based analysis[J]. Journal of Heilongjiang Bayi Agricultural Reclamation University,2024,36(05):128-136.

[6] Jiang Peng. CiteSpace-based knowledge mapping analysis of He Luo cultural research[J/OL]. Journal of Henan University of Science and Technology (Social Science Edition), 2024, (05):100-105[2024-10-23].

https://doi.org/10.15926/j.cnki.hkdsk.2024.05.014.

[7] TIAN Dongling, ZHAI Yan. Thematic evolution, hotspots and frontiers of management accounting research in China analysis

based on CiteSpace knowledge mapping[J]. Foreign Economic and Trade,2024,(10):146-152.

DOI:10.20216/j.cnki.fert1987.2024.10.024.

[8] ZHOU Zhumeng,LI Danyang,WU Huashan,et al. Analysis of research hotspots and cutting-edge situation of waste biomass hydrothermal conversion technology: CiteSpace-based knowledge mapping analysis of big data [J]. Journal of Ecology and Rural Environment, 2021, 37 (04): 409-420.

DOI:10.19741/j.issn.1673-4831.2020.0204.

[9] Xu Kele. Research and Practice on Talent Cultivation Mode of Big Data Technology under the View of Industry-Teaching Integration[J]. Office Automation,2024,29(19):22-24.

[10] QIN Weina, WANG Jingzhai,SUN Mo. Research on the Path of Cultivating Intelligent Accounting Talents in Colleges and Universities under the Perspective of New Liberal Arts[J]. Foreign Trade and Economics,2024,(10):137-141.

DOI:10.20216/j.cnki.fert1987.2024.10.012.

[11] Huang Shang Shang. Research and Prospect of the Application of Big Data Technology in Agriculture[J]. Henan Agriculture, 2024, (19):13-14.

DOI:10.15904/j.cnki.hnny.2024.19.017.

[12] Kang Meng. Research on Innovation of Big Data Auditing Talent Cultivation in Higher Vocational Colleges and Universities under the Background of "New Era, New Demand, New Liberal Arts"[J]. Shanghai Business, 2024, (08):190-192.

[13] Wu Hongjiao. Research on the application of big data technology in 5G communication network[J]. Yangtze River Information and Communication,2021,34(03):219-221.

[14] Zhang Yifan. Innovation of college education management path based on big data technology[J]. Public Relations World, 2024, (20):97-99.

[15] FU Lidong,AI Xiaotong,DU Zengfa. A key node identification method based on hierarchical partitioning and node characterization[J/OL].Computer Engineering, 1-10[2024-10-18].

https://doi.org/10.19678/j.issn.1000-3428.0069040.

[16] Zhang, Huaqiong. Application of big data and artificial intelligence in e-commerce[J]. Market Outlook,2024, (12):76-78.

[17] WANG Wei, ZHOU Xiangyu. Research on data-driven optimization model of real estate operation decision-making[J]. Vitality,2024,42(20):172-174.

[18] HANG Ning,QIAO Kaicheng,HANG Zhuyong. New energy applications in cloud computing data centers[J]. Energy and Energy Conservation,2024,(10):46-48.

DOI:10.16643/j.cnki.14-1360/td.2024.10.092.

[19] LIAO Fangfang,CAO Zikun. Optimizing the comprehensive quality improvement of students in higher vocational colleges and universities using big data and cloud computing technology[J]. Information System Engineering, 2024,(10):69-72.

[20] Equuschick W.G.,C.H. Chen,X.M. Ding,et al. Exploration of hybrid course construction for big data and machine learning courses in geotechnical engineering for graduate students[J]. Higher Construction Education,2024,33(05):64-69.

[21] Wang Na. Analysis of Artificial Intelligence Application in the Context of Big Data Era[J]. Information Record Material,2024,25(09):147-149.

DOI:10.16009/j.cnki.cn13-1295/tq.2024.09.053.

[22] Huo Lijuan. Methods and practices of using big data technology to optimize enterprise management accounting decision-making[J]. Today's wealth,2024,(30):68-70.

[23] LIU Qi, GUO Shengyue. Power engineering data recognition technology based on Internet of Things technology[J]. Electrical Age,2024,(10):39-42.

[24] Sun Tieniu. Artificial Intelligence Promotes Global Business Productivity [N]. Guangming Daily,2024-10-18(012).

[25] Chen Jiang. A targeted mining method for sensor network sample data based on IoT technology[J]. Automation Technology and Application,2024,43(10):104-107+143.

DOI:10.20033/j.1003-7241.(2024)10-0104-05.

[26] ZHAO Yongzhu. Multi-source data processing and fusion technology for smart distribution networks based on edge computing[J]. Electrical Automation,2024,46(05):79-81.

[27] Liu Chang. Research on the practical application of enterprise business intelligence in the context of big data[J]. Vitality,2024,42(19):181-183.

[28] Wang Xia, Cai Baoyu. Discussion on the application of big data technology in precision marketing of cross-border e-commerce enterprises[J]. Mall Modernization,2024,(22):65-67.DOI:10.14013/j.cnki.scxdh.2024.22.017.

[29] Zhao, Qingqing. Research on the application and practice of big data mining and analysis in image processing[J]. Information and Computer (Theoretical Edition), 2022, 34(17):119-121.

[30] Xu Jun. Exploring the use of big data mining and analysis in image processing[J]. Information and Computer (Theoretical Edition), 2024,36(04):150-152.

[31] Gu Jiale, Xuan Junru. Design of student profiling system integrating big data technology and deep learning[J]. Information and Computer (Theoretical Edition), 2022, 34(18): 83-85.

[32] Wang Jinlong. Research on urban flooding disaster data processing and visualization based on multi-source information [D]. Disaster Prevention Science and Technology Institute, 2024.DOI:10.27899/d.cnki.gfzkj.2024.000095.