Research on the Application of Educational Big Data

Huiying Cai

School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan, China

caihuiying323@163.com

Abstract. With the advancement of social informatization, a large amount of teaching and learning data, namely educational big data, has emerged in the process of education. Educational big data is known as an important natural resource. The research on educational big data started relatively late in China, and the research is not sufficient to fully play its role. This paper mainly explores the concept, research status, and application scenarios of educational big data, so that it can be reasonably applied in reasonable applications and maximize the role of educational big data.

Keywords: Educational big data; Diversified evaluation; Adaptive learning; Comprehensive literacy.

1 Introduction

With the innovation and development of information and communication technology, society has ushered in the era of big data. People can record larger and diverse types of data through sensors and store them using information technology. These data have the characteristics of large quantity, high speed, diversity, low value density, and authenticity (5V), known as "big data", which contains rich internal information and core value. Through the analysis and processing of these data, people can obtain the basic laws of things' movement, which has a strong driving force for industrial upgrading, economic development, and human civilization improvement. Big data technology is widely used in various fields from life to production. Among them, the education field is one of the important application fields of big data. In a broad sense, due to the large number of students, numerous courses, and long duration in the education process, the data generated conforms to the characteristics of 5V and can be referred to as educational big data.

2 Types of Educational Big Data

Educational big data refers to the data generated during the education process, which can be classified in various ways.

S. Zhu et al. (eds.), *Proceedings of the 3rd International Conference on Education, Language and Art (ICELA 2023)*, Advances in Social Science, Education and Humanities Research 831,
https://doi.org/10.2991/978-2-38476-214-9_72

2.1 According to the source of data

Big data can be divided into course data and non- course data. Course data refers to live and recorded videos, courseware, assignments, and different types of test questions related to the course being studied, and is the carrier of knowledge transmission. Non-course data refers to behavioral data generated by students during the learning process or data accumulated in daily life.

There are two main types of behavioral data. One is the learning state of students during class, such as facial expressions, whether they are looking at their phones, whether they are sleeping, and learning postures. Another type is for students to rely on different online education platforms to watch the traces left by live or recorded videos, homework, completion time and scores of various test questions, and interaction information during classes. By analyzing behavioral data, it is usually possible to determine students' learning habits, attitudes, and progress. The data accumulated in daily life includes student activity trajectories, internet browsing records, campus card consumption records, communication records, and other life data. Both course data and non-course data are important resources for guiding students to actively and healthily develop and improve their comprehensive literacy.

2.2 According to data types

Big data can be divided into various forms of data such as video, audio, images, files, logs, etc. Video data not only refers to live or recorded videos, but also includes videos collected through monitoring devices such as classrooms and schools, as well as video materials uploaded by classrooms or students. This type of data sometimes exists in the form of audio or images to save memory. Files refer to various types of documents, such as PDF, PPT, Word, and other forms of information, which include not only learning materials uploaded by teachers, but also learning materials uploaded by students, such as assignments, papers, and exams. Logs are traces left during the learning process, such as records of clicks or replays while watching videos, completion time and duration of assignments, and records of completing test questions.

2.3 According to the degree of data structure

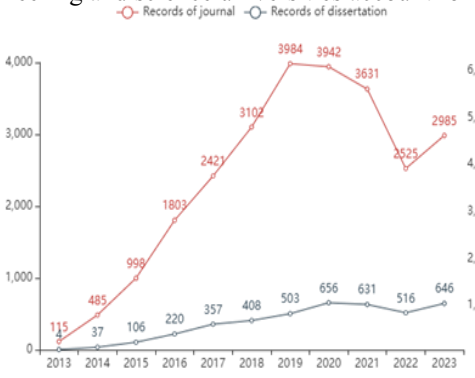
Big data is divided into structured data, unstructured data, and semi-structured data. Structured data refers to data with relatively complete entities and relationships, such as data already defined by various online platforms and stored in databases, such as personal information of students and teachers, scores, etc. Unstructured data refers to free form data without defined and standardized constraints, such as videos, audio, images, etc. collected by various monitoring devices. Semi-structured data refers to some data that is structured, but there is a large amount of less structured data, such as courseware, interactive area data, etc.

3 Current Status of Educational Big Data Research

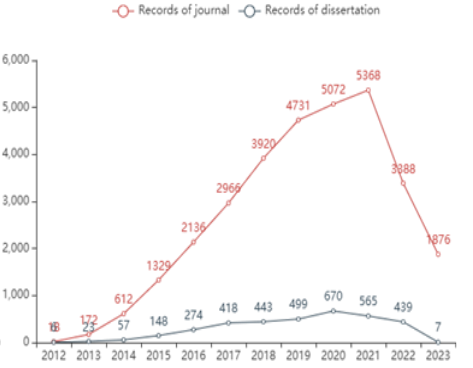
Based on the current mainstream large-scale literature databases such as CNKI, Wanfang, and Web of Science, the application of obtaining educational big data was organized using literature research methods.

3.1 Retrieve using CNKI and Wanfang

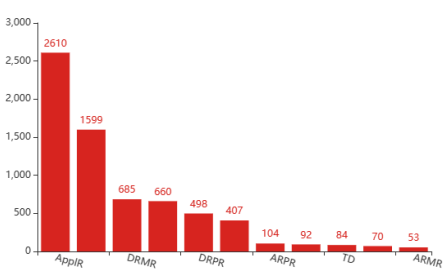
Using the advanced search functions of CNKI and Wanfang, with "big data" and "education" as the main search keywords, 27.2 thousand and 35.2 thousand search results were returned, respectively. The number of search records for academic journals and dissertations per year is shown in Fig. 1 (a) and Fig. 1 (b), with the 2023 data being the predicted value. From the results returned by CNKI and Wanfang, it can be seen that the curve shape is basically the same, from only a few dozen in 2012 to over 3000 in 2022, reaching its highest value in 2019. According to discipline classification, there are 10.3 thousand records in the discipline of "higher education", accounting for 37.9% of the total. This is because students in high school, middle school, and primary school still rely mainly on offline education, and due to their low academic years, there is relatively less accumulated data. In the higher education stage, students' sense of autonomy is enhanced, the scope of activities is wider, and the forms of educational activities involved are more diverse. According to the classification of research levels, the retrieval results of different research levels are shown in Fig. 1 (c), among which the number of applied research and development research is the highest, accounting for 10.02%, indicating that the research focuses relatively more on the practical application and mining analysis of data. According to the classification of universities, the number of research records returned from the top eleven universities with search results is shown in Fig. 1 (d). Among them, seven universities are normal universities, indicating that research on education big data is concentrated in normal universities, while engineering and science universities account for a relatively small proportion.



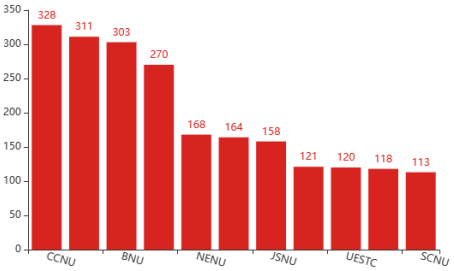
(a) Results from CNKI



(b) Results from Wanfang



(c) Divided by research level



(d) Divided by college

Fig. 1. Results from CNKI and WanFang

3.2 Retrieve using web of science database

Relying on the Web of Knowledge literature database for retrieval and analysis. Using "topic" as the search criteria and selecting "education" and "big data" as the search keywords, a total of 11388 records were obtained. Among them, there are 7657 records for journal papers, 1641 records for conference papers, and 1542 records for degree papers. The total quantity distribution for each year from 1990 to 2023 is shown in Fig. 2, with 2023 being incomplete data. The quantity roughly develops in a ladder like manner and can be divided into three stages. The first stage is before 2006, which can be called the enlightenment period. During this stage, the technology and equipment performance are difficult to meet the requirements of analyzing big data, so there are fewer researchers; The second stage is from 2006 to 2016, which can be called the period of rapid development. During this stage, the performance of devices has developed to the level of distributed and parallel computing. At the same time, the development of some data mining algorithms, such as deep learning and big data technology, has laid a good theoretical research foundation for the application of educational big data.

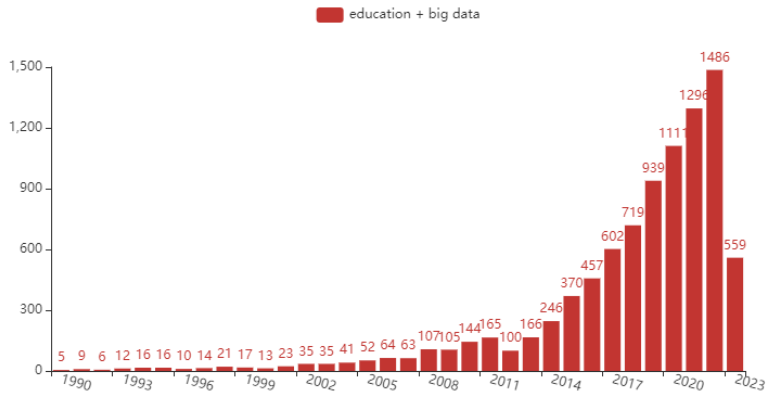
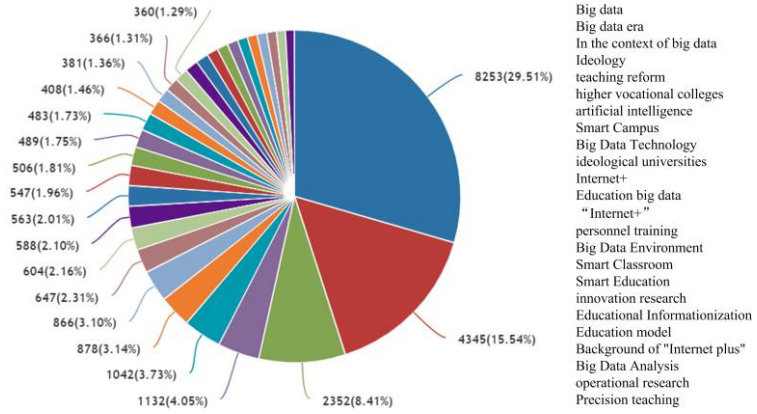


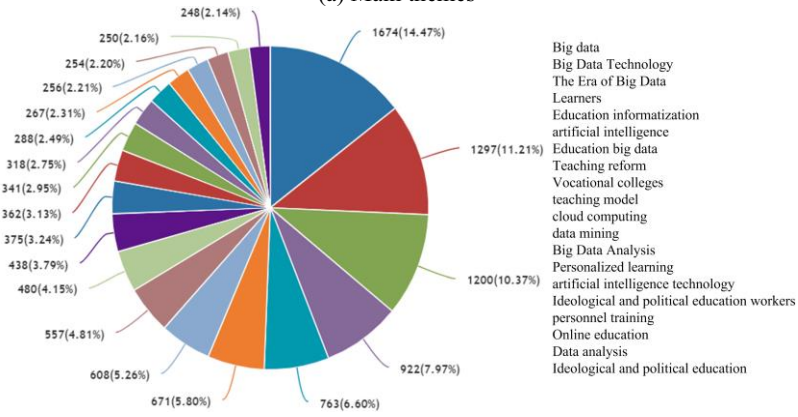
Fig. 2. Results from Web of science

4 Analysis of educational big data application scenarios

Through the advanced search function of CNKI, set the themes as "education" and "big data" as search keywords for retrieval. The results of extracting the main and secondary themes from the returned results are shown in Fig. 3(a) and Fig. 3(b). By analyzing the main and secondary themes, it was found that the application scenarios of educational big data can be divided into five categories: teaching reform, smart campus, personalized learning path, evaluation model construction, and educational decision-making.



(a) Main themes



(b) Secondary theme

Fig. 3. Distribution of Main and Secondary Themes

4.1 Teaching reform

The deep integration of big data in the field of education has promoted reforms in teaching models, teaching methods, curriculum systems, teaching content, and other aspects. In terms of teaching mode, searching with the theme of "education", "big data", and "teaching reform" as search keywords can obtain 2914 search results, accounting for 13.12% of the total education big data. Through the analysis of education big data, it

was found that the teaching mode has gradually shifted from traditional offline teaching to a combination of offline and online teaching. For example, relying on teaching platforms such as Wisdom Tree, Chaoxing, Rain Classroom, and MOOC, the teaching mode of flipped classroom, precision teaching, blended learning, and information based teaching has been transformed [1]. In terms of teaching methods, there has been a gradual shift from traditional cramming teaching to case driven teaching, interest driven teaching, personalized training teaching, individualized teaching, etc., which can better improve teaching quality and student learning efficiency [2]. In terms of teaching content, analyzing educational big data can help teachers and students identify and fill in gaps, and promote teachers' teaching of knowledge points and students' understanding of knowledge points at a reasonable pace [3]. For teaching reform, the country has proposed four new construction projects from a strategic perspective to promote "engineering, new medical, new agricultural, and new humanities". In order to promote the construction of the new disciplines, it is necessary to carry out teaching reform. Combining big data with education can encourage researchers to propose deeper and more suitable teaching reform plans for student development.

4.2 Smart Campus

With the development of the Internet of Things, big data, and artificial intelligence technology, building a digital smart campus that can ensure students' learning and life has become an important aspect of school construction [4]. Smart campuses usually rely on the construction of virtual big data platforms, allowing schools to showcase their educational achievements through data analysis. They promote the improvement of students' comprehensive literacy from various aspects of students and life, such as analyzing students' learning and life data, mental health, abnormal behavior, etc. The purpose of a smart campus is usually to build an open, diverse, humane, efficient, safe, and harmonious campus environment, achieve comprehensive intelligence in teaching and research, management and services, and enhance the comprehensive literacy of students and teachers, and enhance the core competitiveness of the school.

4.3 Personalized learning path

With the continuous accumulation of educational big data and the continuous optimization and improvement of big data processing technology, students' learning process is easily recorded through videos, audio, images, logs, etc. By analyzing and mining student learning data, unique characteristics such as learning characteristics, learning habits, interests, and areas of expertise of each learner can be obtained [5], laying a solid foundation for the construction of personalized learning paths and precise teaching for students, truly achieving the personalized training task of "teaching according to their aptitude", and enabling students to grow into "top talents" in their fields of expertise.

4.4 Evaluation model construction

By analyzing and mining educational big data, the reproduction of teaching and learning processes can be achieved. Therefore, based on educational big data, education quality evaluation models and student comprehensive evaluation models can be constructed [6]. Based on data-driven evaluation models, imperfect links and influencing factors in learning process can be identified, effectively improving the quality of education and teaching, and promoting the improvement comprehensive literacy.

4.5 Educational decision-making

In the era of big data, emerging technologies such as the Internet of Things, artificial intelligence, and cloud computing have effectively supported the continuous collection, deep mining, and systematic analysis of campus data. They can be presented to education management personnel in a dynamic manner through data fusion and interaction, and can effectively provide good decision support for multi-entity value judgment, scientific management, and continuous improvement [7]. Based on educational big data, analyzing the learning attitude, teaching attitude, and degree of achievement of teaching goals of students and teachers can provide accurate and reasonable education plans for education management personnel, and formulate reasonable development plans for improving students' comprehensive quality from a macro perspective.

5 Summary

Educational big data, in response to the rapid development of information technology, can provide students, teachers, and education management personnel with corresponding support to promote their reasonable participation in the education and teaching process. For students, evaluation models can be used to obtain evaluation indicators such as their learning attitude, learning ability, and mental health level, promoting their personalized growth. For teachers, they can provide teaching reform, teaching evaluation models, and judge their teaching attitude and degree of teaching rationalization. At the same time, it can provide reasonable educational decision-making for educational management personnel. How to reasonably and effectively utilize education big data to achieve the above application scenarios is our focus in the next stage of work.

Acknowledgement

The paper is a part of research result of 2022 Undergraduate Teaching Reform Research General Project of Shandong Province “Research and Practice on Curriculum Construction Based on the ‘OBE Concept’ and "Multi driven" - Taking <Computer Vision> as an Example”.

References

1. Prahani, B. K., Jatmiko, B., Hariadi, B., Sunarto, D., Sagirani, T., Amelia, T., Lemantara, J. (2020) Blended web mobile learning (BWML) model to improve students' higher order thinking skills. Int. J. Emerg. Technol., 15(11): 42–55. https: // doi. org/ 10. 3991 / IJET. V15I11.12853.
2. Liu, Q., Zeng, L. (2020) Design and application of experimental teaching system driven by big data technology in economics and management majors. Int. J. Emerg. Technol Learn., 15(11): 56–66. https://doi.org/10.3991/IJET.V15I11.12997.
3. Li, B. (2023) The innovation and development of university international trade teaching reform under the background of big data. J. Comput Methods Sci., 23(2): 1043–1051. https: // doi.org/10.3233/JCM-226618.
4. Villegas-Ch, W., Palacios-Pacheco, X., Luján-Mora, S. (2019) Application of a smart city model to a traditional university campus with a big data architecture: a sustainable smart campus. Sustainability, 11(10): 2857. https://doi.org/10.3390/su11102857.
5. Huan, S., Yang, C. (2022) Learners' autonomous learning behavior in distance reading based on big data. Int. J. Emerg. Technol., 17(9): 273–287. https: //doi. org/ 10. 3991/ ijet. v17i09.31373.
6. Cai, X., Feng, X., Yu, H. (2023) Broad learning algorithm of cascaded enhancement nodes based on phase space reconstruction. Appl. Intell., 53(2): 2321–2331. https: // doi. org/ 10. 1007/s10489-022-03513-4.
7. Ghasemaghaei, M., Turel, O. (2023) The duality of big data in explaining decision-making quality. J. Comput. Inform. Syst., 63(5): 1093–1111. https: //doi. org/ 10. 1080/ 08874417. 2022.2125103.