Design and Research on Learning Early Warning Model for Online Education Based on Big Data in Education

Jianqiang Xu[1,a,*], Simin Zhou[1,b], Yuting Shi[1,c]

[1]Beijing Institute of Technology, Educational Technology Research Center, Beijing, China

[a] xujq@bit.edu.cn, [b] 3120211960@bit.edu.cn, [c] 3120231961@bit.edu.cn

*Corresponding author

Keywords: Big Data in Education, Online Education, Early-Warning for Learning, Early Warning Model, System Architecture

Abstract: In the era of big data, various school data information systems can be utilized to deeply explore the relationship between students' learning behavior and learning effect, so as to dig deeper into educational problems. Based on the analysis of existing research problems in educational big data and learning warning, and combined with the technical framework of educational big data, this paper adopts attribute analysis, cluster analysis, text mining, classification prediction and other technologies to construct a logical model of learning warning system based on educational big data. The system is composed of five layers: data collection layer, data integration layer, data mining layer, learning analysis layer and application service layer. According to different functions, the system is divided into five modules: data collection module, data integration module, attribute selection module, analysis and evaluation module and warning output module. In addition, the article also designs the specific process of the learning warning system and a terminal-based warning effect display interface, and finally puts forward practical suggestions and reflections to provide more effective services for online education users and educational administrators.

1. Introduction

In recent years, the scale of online education users has significantly increased due to the combined effect of national policy support, prosperous development of the internet, and an increase in people's willingness to receive education. However, the drawbacks of online education have gradually emerged. For example, the online environment restricts deep communication, and the virtual nature of online learning results in one-way, low-frequency, random, and delayed weak interactions between teachers and students as well as among students. There is also a contradiction between learners' autonomy in online courses and their need for objective organization[1]. According to a study conducted by Harvard University and MIT (Massachusetts Institute of Technology), out of 840 thousand registered participants in 17 MOOCs (Massive Open Online Courses), only 5% successfully completed the course and obtained certificates. Even at Peking University, less than 4% of MOOC learners pass[2].

Learning warning is an effective measure to address existing problems in online education. By mining and analyzing a large amount of data generated during the learning process, it assesses overall learning performance to timely identify struggling students and implement effective assistance measures with the aim to improve learning achievement rates and retention rates for online education students. Since the 1990s, foreign countries have started developing learning warning research aimed at helping students complete their studies on time. Typical examples include Purdue University's Course Signals system; Desire2Learn's Student Success System; Khan Academy's Learning Dashboard; all achieving certain effectiveness in preventing dropouts[3]. In comparison with these efforts abroad， relevant research on this topic started relatively late domestically，with more emphasis on theoretical studies and less on applied research. Moreover, the majority of research has been focused on traditional higher education [4], lacking in-depth investigation into online education warning systems.Therefore,it is urgent to explore the learning

  DOI: 10.25236/meeit.2024.001

warning model suitable for online education,and use the technical advantages of education big data to improve the warning effect.

Big data in education, as a technology system applied to and serving education, has a strong driving force in promoting teaching reform, accurate learning, personalized learning and other aspects, and is an important support for the development of education modernization. The research and application of educational big data will continue to be an important driving force for educational innovation and social development for a long time [5]. How to make full use of the technical advantages of educational big data and improve the effect of online education learning early warning is an urgent problem to be solved. Therefore, this study designed an learning early warning model based on education big data for online education users. By collecting academically related data, big data mining technology is applied to analyze the correlation and dependency among the data, find the key factors affecting the study, accurately discover the problems existing in the study, and provide corresponding early warning and intervention strategies. This model can provide strong technical support for online education, improve students' learning level and retention rate, and promote the development of online education.

2. Current status of learning early warning based on big data in education

At present, there are numerous research achievements on learning early warning at home and abroad, with slightly different focuses. In foreign countries, the emphasis is mainly on the development and application of warning systems, analysis methods, visualization tools, and other fields [6]; The Course Signals system at Purdue University and the visual learning dashboard system at Khan Academy are typical online learning warning system platforms[3]. The former collects and analyzes information such as student characteristics, course performance, and course effort through SSA algorithms, and presents different learning grades through red, green, and blue color signals, while the latter tracks and analyzes student learning behavior data and learning process data through information tracking technology and mirroring technology, and provides visual output in the form of charts and other charts for learners' and teachers' learning analysis. [7]The E2Coach system designed by Mckay T et al. provides personalized warning push services for students by analyzing learners' historical learning data as well as their past learning performances. [8]José A. Ruipérez-Valiente et al. developed the ALAS-KA Learning Analytics Extension Tool based on the Khan Academy platform to further analyze learners' learning styles and visualize the results in order to help teachers choose appropriate teaching strategies and optimize learners' learning process. These learning early warning systems have achieved certain effectiveness in reducing dropout rates.

Domestically, research mainly focuses on the areas of learning early warning model design and risk prediction[9] Xiao Jun et al. (2019) proposed the "ODAS" open learning framework, which uses big data technology to analyze students' online learning data in order to build learner profiles, help identify high-risk learners, and provide personalized learning alerts. [10]Shu Ying et al. (2019) analyzed the data of non-intervention behaviors in online learning using data mining and analytics techniques to determine warning factors for online learning crisis, and constructed an accurate early warning model using plain Bayes to promote personalized teaching and student management.[11] Ding Yongguo et al. (2019) collected data related to students' social and demographic characteristics, students' personal characteristics and students' inputs, and constructed a prediction model for college students' performance in laboratory classes. [12] Wu Xuguo et al. (2020) used association rule analysis to find out the implied relationship between failed courses, constructed a research framework for college students' learning early warning, and provided students with early warning hints for course failure. [13] Li Xiongping et al. (2020) designed learning early warning work outcomes and needs based on teaching-related functional departments, and tried to explore the educational information and value behind students' learning difficulties.

In summary, the application of big data technology in learning early warning is of great significance in promoting the improvement of education and teaching and the optimization of learning quality, but the following problems remain: ① The early warning results are not

personalized, the universality is low, the prediction results are difficult to be interpreted by non-technical people, and the early warning content is not comprehensive enough. ② The target of early warning is mainly higher education learners, and there are fewer studies on learning early warning in online education environment. ③ There is a single source of data, and the influencing factors of learning early warning are not deeply explored from a global perspective. This is also the research focus of this study. In future research, it is necessary to strengthen the exploration of learning early warning carried out on the whole learning process of online education students in order to understand the inner mechanism of learning early warning more comprehensively and deeply.

3. Modeling of learning early warning systems

3.1. Logic model of the learning early warning system

The logical model of the learning early warning system based on big data in education is a five-level architecture, formed by the data collection layer, data integration layer, data mining layer, learning analysis layer and application service layer, as shown in Figure 1.

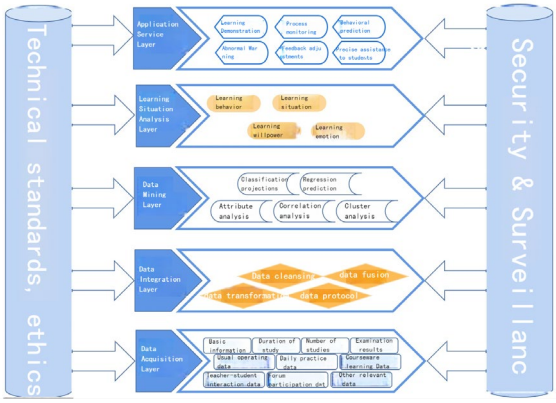

Figure 1 Logical model of the learning early warning system

To realize the drive for learning diagnostic development, big data in education needs to solve the problem of data collection [14]. The data collection layer is mainly responsible for collecting all the data generated during the whole learning process such as students' enrolment, course selection, cultivation plan, assessment, learning practice, defence, etc., including basic information, study time, number of study times, examination results, usual homework data, daily practice data, teacher-student interaction data, forum participation data, and other parameters that are closely related to the learning process, which have an important impact on the final learning effect. The data collection layer focuses on the data collection process, which has an important impact on the final learning effect. The focus of the data collection layer is to ensure the completeness and accuracy of data collection, so the data should come from a wide range of sources, including student management platforms, network information centre management platforms, e-learning systems, faculty systems, network course resource management platforms and other types of information systems.

In order to solve the problem of difficult data comparison and analysis caused by different formats in most school systems, it is necessary to focus on the results of data integration during the construction process at the data integration layer. By using big data mining techniques such as data transformation, data cleaning, data reduction, and data fusion, different types of scattered heterogeneous data can be transformed into unified and feedback-generating data to address the issues of inconsistent and difficult-to-integrate data formats. This will better facilitate the construction of the data integration layer and ensure that the information stored in subsequent educational databases is highly relevant to learners' learning performance, with high validity and easy-to-analyze high-quality massive datasets.

The Data Mining Layer is a key link in data processing, which is based on the data integration layer, classifies and cleans the various types of data collected, and uses appropriate data mining techniques and algorithms for in-depth exploration and analysis.In the mining process, it is first

necessary to preprocess the raw datasets, including data cleaning, de-duplication, missing value filling and other operations, to guarantee the quality and accuracy of the data. Then, according to the different types of data, choose the appropriate data analysis methods for mining, such as cluster analysis, classification analysis, correlation analysis, anomaly detection and other methods, to find out the key factors affecting students' learning performance, and form an early warning indicator system, which assists students in effectively monitoring and managing their own learning situation .

The function of the learning analysis layer include learning analysis assessment and determination of warning levels. This layer integrates and dynamically analyses data from the first three layers to provide data support for the application service layer. The core task of the learning analysis layer is to extract key learning factors and conduct comprehensive, multimodal analysis of students' learning situations throughout their entire learning process in order to make an integrated judgment on whether a warning is necessary, and determine appropriate warning strategies. Threshold setting for learning assessment and warning level definition usually involves statistical analysis of indicator data using methods such as median principle, mean principle, majority principle, parameter principle, and correlation principle. The work of this tier provides valuable data support for students' studies, providing more personalised learning assistance.

The application service layer is the highest layer of the Learning Early Warning System, which mainly provides users with process supervision, learning situation display, behaviour prediction, abnormal warning, feedback adjustment and precise study assistance services. The system monitors the whole learning process of learners and provides users with clear feedback by displaying the results of learning analysis in a visual way. At the same time, the application service layer is also capable of predicting subsequent learning behaviours, focusing on abnormal learning situations that have occurred or are likely to occur, and providing learners with personalized early warning strategies, with the main purpose of providing learners with precise assistance services to improve learning effects and learning quality.

Each layer above needs to follow the relevant technical standards and ethics, and is subject to global monitoring and security protection to ensure the standard, interoperability and security of educational data in all links, and protect the legitimate rights and interests of participants.

3.2. Functional model of the learning early warning system

The functional model of the learning warning system is shown in Figure 2, which shows the main process of the online education learning warning system, including five modules: data collection, data integration, attribute filtering, analysis and evaluation, and warning output.From the learners' learning status, learning behavior, learning willpower and learning emotion, this paper constructs an online education learning warning function model, which provides comprehensive warning for online education students from the four aspects of knowledge, emotion, intention and behavior.

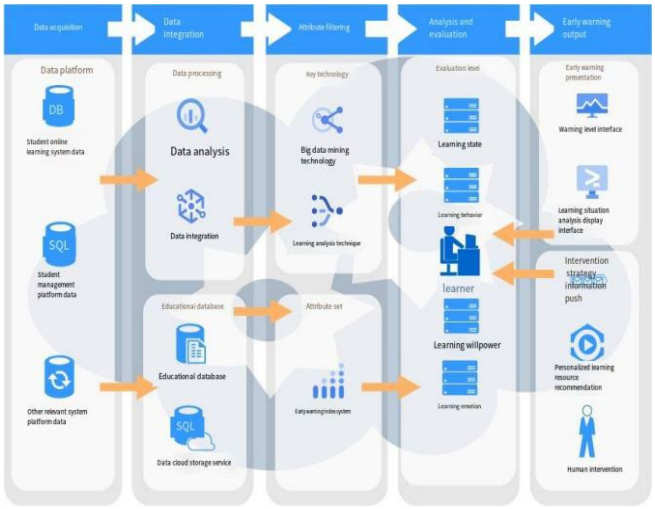

Figure 2 Functional model of learning early warning system

3.2.1. Data collection module

The data collection module uses data collection technology to collect learners' basic information data, learning emotion data, learning interaction data, learning behavior data, learning performance data and other relevant data from various educational data platforms in the context of online education. Among them, basic information data includes personal information, learning starting point, learning style, habits, etc.; learning emotion data includes emotional state, learning motivation, learning pressure, learning willpower, etc.; learning behavior data includes course selection, teacher-student interaction, forum participation, courseware learning, daily exercise, etc.; learning interaction data includes learning start time, learning end time, online duration, learning times, download times, note times, etc.; learning performance data includes grades, ranking, and obtaining credits, etc.; other data includes learning terminals, learning environment, network conditions, etc. The initial data collected provides support for online education learning early warning from two aspects: on the one hand, massive data of the whole process of online education learners' learning work is collected to provide data support for the formation of a large education database; at the same time, through real-time monitoring and collection of learners' learning process data, a data foundation is laid for the selection of subsequent learning early warning intervention strategies.

3.2.2. Data integration module

This module mainly constructs an educational big database, i.e. a learner-centred and dynamically changing big data collection. The collection of educational big data is formed through the collection of the whole process data of different types of learners, and the data are cleaned, fused, transformed, statute and other processes are carried out, so that the characteristic values of various types of data, which are standardised and consistent with the needs of learning early warning, are stored in the educational big database, in order to identify the overall status of the learner's academics and to improve the accuracy of the learning early warning.

3.2.3. Attribute filtering module

According to different warning purposes, key attributes are screened by combining big data mining techniques and learning analysis. Specifically, attribute screening is carried out on the massive data in the educational big database using technical means such as principal component analysis, attribute scoring based on information gain, etc. According to different early warning needs, algorithms such as correlation analysis are used to analyse the relationship between each attribute and find the correlation patterns between attributes, and algorithms such as classification analysis, cluster analysis, and regression analysis are used to rank the importance of attributes, and ultimately, to find the key attributes of the demand and form the early warning indicator system.

3.2.4. Analytical assessment module

The analysis and evaluation module primarily assesses the academic status of learners across four dimensions: learning status, learning behavior, learning willpower, and learning emotion. It utilizes methods such as social network analysis, content analysis, discourse analysis, text analysis, collaborative filtering, and various clustering algorithms to analyze and evaluate learning status and behavior.The main technologies for recognizing learning willpower and learning emotion include emotion recognition and facial expression recognition technologies. Additionally, text sentiment analysis technology can be employed to analyze and extract emotions expressed in learners' texts, such as happiness, sadness, and anger. This analysis helps to infer whether there are any abnormalities in the learners' learning state, behavior, willpower, and emotions.Consequently, it determines the outcomes of the learners' learning situation analysis and judges the warning level according to the four-level signal system of learning early warning. This process aims to identify learners at risk who require an academic warning. The discriminant table of the four-level signal system of learning early warning is presented in Table 1.

Table 1 Discriminant table of the four-level signal system of learning early warning

Level	Color	Meaning	Early warning countermeasures
level one	blue	Just OK	Recommend personalized learning resources
level two	yellow	Minor issue	Recommend personalized learning resources, sends early warning reminders and learning suggestions
level three	orange	Serious issue	Recommend personalized learning resources, sends early warning reminders and learning suggestions, teacher intervention
level four	red	Immediate action required	Recommend personalized learning resources, sends early warning reminders and learning suggestions, implements intervention immediately by multiple educational subjects, and sends notice of severe learning difficulties, etc.

The learning early warning model adopts a four-tier signalling system, which is formulated with reference to the warning signals for sudden meteorological disasters. The signal system adopts four colours, namely blue, yellow, orange and red, which are classified according to the severity of the abnormality and visually represent the four abnormal learning states. Among them, the blue signal indicates that the learning performance is basically normal, the yellow signal indicates that there is some problem, and the orange signal indicates that the problem is more serious and requires manual intervention. The red signal is the highest level of warning signals, which means that the learning problems are extremely serious and require immediate intervention by multiple educational actors and the adoption of appropriate intervention strategies according to the actual situation.

3.2.5. Warning output module

The main function of the Warning Output Module is to provide appropriate intervention strategies for learners who need early warning. To this end, the module will make full use of the superiority of big data data analysis, adopt the concomitant evaluation strategy, accurately portray the learner's learning situation, and present the diagnostic results of the early warning level definition and the learning situation analysis to the learner in the form of visualisation. The early warning presentation part adopts the forms of traffic lights, dashboard charts and radar charts to visually present learners' warning levels and learning conditions. The early warning strategy part is mainly through system intervention and human involvement. The system uses an analysis and evaluation engine to automatically generate personalised learning resources with outstanding tendencies and appropriate warning messages in conjunction with the results of the warning level to provide personalised interventions to learners. Teachers can use the Early Warning System to keep track of their classmates' learning status and implement targeted interventions, such as communicating with students to extend their learning horizons. The two-tier intervention mechanism of the Early Warning Output Module helps learners to change from passive content reception to active participation in learning activities, and to gradually improve their independent learning ability and motivation, and ultimately realize independent learning and self-development.

3.3. Learning Early Warning System Process Model

The learning early warning system model is shown in Figure 3. First of all, the system reads the detailed parameters of the learner's online education situation from various data platforms, uses big data analysis technology to transform, clean up, statute, fusion and other online pre-processing to integrate into the education database, and sends the processed learner data to the learning analysis and assessment platform for learning analysis and early warning level judgement through the dedicated information transmission interface. The assessment results and the output warning information are then sent back to the system in different ways through the special information transmission interface to present to different end users. After the early warning is implemented, the system collects the resource data generated during the learner's learning process again to achieve

dynamic updating and tracking analysis of the learning trajectory, and carries out secondary intervention for learners with poor warning results until the learning status is back to normal.

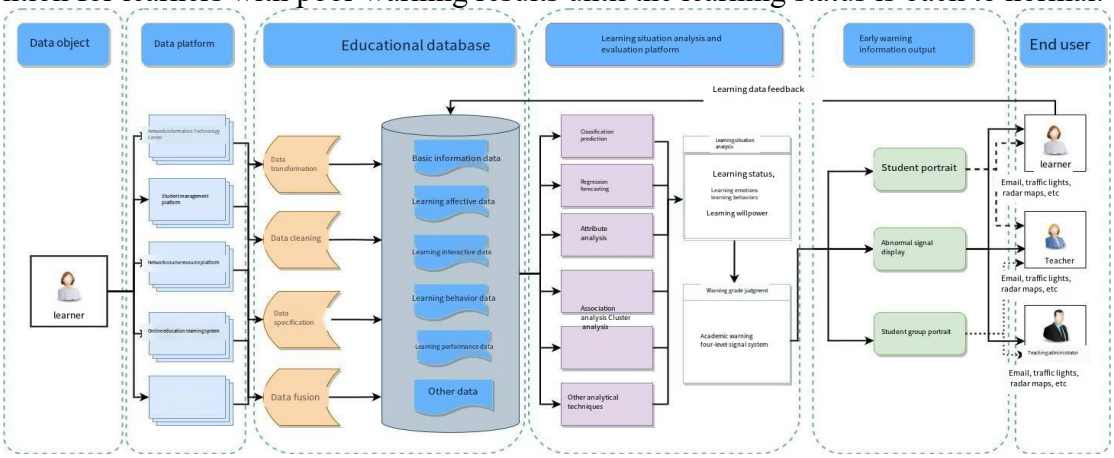

Figure 3 Learning Early Warning System Process Model

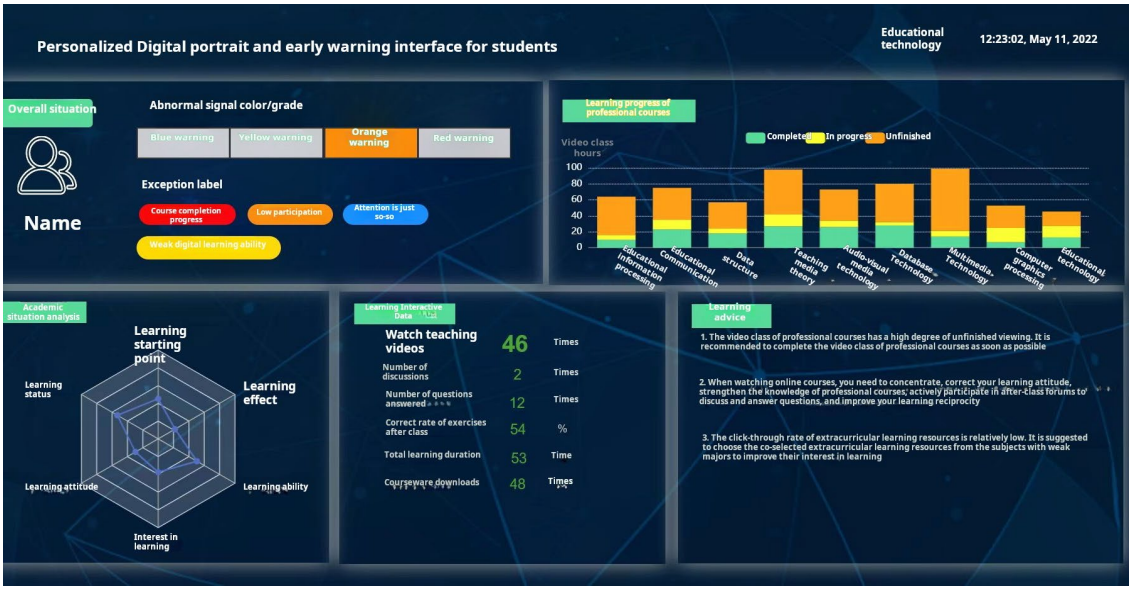

Figure 4 Early warning system display effect diagram

In this study, a learning early warning system for mobile terminal devices was developed and designed using the Digital Vectors data visualisation APaaS platform. The display effect of the early warning system is shown in Figure 4, which mainly includes modules such as overall situation, learning progress of professional courses, radar chart of learning situation analysis, learning interaction data, learning suggestions and learning resources recommendation. Among them, the overall situation module reflects the learner's current warning level status and specific abnormal labels, providing the learner's current status information in an intuitive form; the professional course learning progress module displays the learner's progress in various courses, where the learner can learn about his/her own course learning progress in order to better plan the learning time and adjust the learning strategy; the learning situation analysis radar chart presents the learner's learning status, learning attitude, and other multi-dimensional information. The learning situation analysis radar chart presents learners' learning status, learning attitude and other multi-dimensional learning situation indicators, so that learners can intuitively understand their own performance in the learning process and adjust their learning status and attitude in a timely manner; the learning interaction data module characterizes learners' interaction level on the online platform to help learners understand their own habits and behavioural patterns in online learning; the learning suggestion module provides personalized learning guidance, which will provide personalized learning guidance to learners according to learners' learning status and learning characteristics. The Learning Advice module provides personalised learning guidance based on the learner's learning

status and learning characteristics, and provides personalised learning advice to learners so that they can complete their learning tasks more effectively. The Learning Resources Recommendation Module provides learning resources with high relevance and timeliness by exploring learners' most urgent intelligent needs and synthesising learners' weak points. This module can help learners better find the learning resources they need.

4. System Implementation Conditions and Application Reflection

4.1. System implementation conditions

In order to ensure the effective implementation of an learning early warning system, there are two key points to consider: firstly, school departments need to collaborate with each other to ensure the integration and flow of data. School application systems have different formats, which makes it difficult to compare and analyse data and hinders the interoperability and processing of early warning data. Therefore, there is a need for departments to collaborate with each other to establish consistent data standards and to reduce barriers to data interoperability between systems caused by different data standards to facilitate the effective operation of the early warning model. Second, the implementation of the learning early warning system requires the establishment of relevant guarantee mechanisms. ① Establish an early warning guarantee mechanism. The early warning system requires the participation of multiple subjects, such as teachers, administrators and students. At the same time, it is necessary to ensure the normal operation of a variety of early warning resources, such as information security, real-time visualisation of the output of the resource facilities to ensure and ensure the normal operation of the early warning system of the resource technology to ensure. In addition, the school should incorporate the learning early warning work into the daily teaching management, set up personal files for the early warning targets, pay close attention to the early warning process and the effect of early warning, and summarize the experience and deficiencies in a timely manner, in order to continuously optimize the early warning method and improve the effect of early warning. ② It is necessary to establish a big data ethical regulation mechanism to guarantee information security, in order to effectively protect the personal privacy of educational big data users. When implementing the learning early warning system, the data collected, generated, and used in the early warning process should be strictly protected to improve the security of the early warning model. At the same time, it is necessary to strengthen the moral education and constraint management of big data users to avoid the ethical dilemma of privacy protection in the process of big data use. Third, after the implementation of the learning early warning system, the early warning system model needs to be continuously optimized to improve the early warning effect. In terms of technology, more perfect data mining technology and artificial intelligence technology can be used to improve the success rate of learning early warning; in terms of manual intervention, it is necessary to enrich the early warning resources from multiple dimensions, strengthen the means of intervention, and improve the efficiency of early warning. Ultimately, the universality of early warning should be increased to lay a solid foundation for realizing online accurate learning assistance, creating a personalized learning environment, and improving the overall quality of online education.

4.2. Reflections on the application of the system

Currently, most learning alerting research has focused on the exploration of data mining techniques, but less on the theoretical frameworks that underpin the techniques. However, the purpose of educational big data applications is to facilitate learners' growth and enhance their learning outcomes to support human development. Therefore, the use of big data technologies should always be guided by educational theories.

In practice, we need to balance the relationship between the application of big data technology and human beings as the subject of the application, to avoid excessive blind use of the technology that will adversely affect the data generator and the subject of the use. Behind the big data is a summary of the emotional state, psychological state and physical health state of each learner, which

has a very high emotional value and is not just a simple number. Although the application of big data in education facilitates the improvement of learning efficiency and effectiveness, human development has its own characteristics, and therefore the scope of authority to use the technology needs to be regulated on the basis of ethical standards. Therefore, in practical applications, it is necessary to strengthen the ethical constraints on data users, improve data security, and reasonably use educational big data under the premise of safeguarding data privacy and security, balancing the relationship between human development and technology applications.

5. Conclusion

learning early warning, as one of the important measures to improve students' learning quality, can provide a scientific basis for learning interventions. However, the existing research on learning early warning is mainly oriented to traditional offline education forms, and the research on learning early warning for users of online learning forms has a high practical value. Therefore, this study analyses and discusses the current problems of online education and learning early warning, and constructs a five-level architecture of online education learning early warning system based on educational big data, designs the functional model and process model of the learning early warning system, and finally puts forward suggestions on the implementation and application of the learning early warning system, with a view to providing a new way of thinking and methodology for the learning early warning of online education. At a later stage, we will continue to integrate learning data from different learning scenarios, continuously improve the warning indicator system, explore smarter warning algorithms to enhance the system application effect, and continue to iteratively optimise the user experience.

Acknowledgements

This paper is the phased research achievement of the general project "Research on Online Education Learning Analysis and Early Warning Model Based on Educational Big Data" (No. CHDB21172) of the 14th Five-Year Plan for Beijing Education Science in 2021.

References

[1] LIU Qian,LI Ying.Active Choice:A Positive Understanding of“High Dropout Rate”of Online Courses: Based on the Perspective of Learner Investment Theory[J].e-Education Research, 2021,42(04):45-52.

[2] Jiang Zhuoxuan, Zhang Yan, Li Xiaoming. Learning Behavior Analysis and Prediction Based on MOOC Data[J]. Journal of Computer Research and Development, 2015, 52(3): 614-628.

[3] WANG Lin-li, YE Yang, YANG Xian-min.Design of Online Learning Early-warning Model based on Big Data ——The Learning Early-warning of“Research and Practice Column about Big Data in Education”[J].Modern Educational Technology,2016,26(07):5-11.

[4] WANG Lin-li. Design of Online Learning Early-warning Mechanism Based on Big Data [D].Jiangsu Normal University,2017.

[5] Yang Yong,Duan Chao,Li Kaiqi,Yi Xueling.Study on the impact of big data in education on social development[J].Chinese Journal of ICT in Education,2021(09):40-42+80.

[6] YANG Feng-yu,NIE Wei,ZHENG Wei,QIAO Guang-chao.Survey of Learning and Early Warning Research Based on Big Data[J].Modern Computer,2021(04):8-17.

[7] Mckay T, Miller K, Tritz J. What to do with actionable intelligence: E2Coach as an intervention engine[EB/OL]LAK’ 12 Proceedings of the 2nd International Conference on Learning Analytics and Knowledge.2012:88-91.

[8] José A. Ruipérez-Valiente et al. ALAS-KA: A learning analytics extension for better

understanding the learning process in the Khan Academy platform[J]. Computers in Human Behavior, 2015, 47 : 139-148.

[9] XIAO Jun, QIAO Hui, LI Xuejiao.Construction and Empirical Study of Online Learners' Persona in the Big Data Environment[J].Open Education Research,2019,25(04):111-120.

[10] Ying Shu, Qiang Jiang, Wei Zhao.Accurate alerting and prevention of online learning crisis: an empirical study of a model[J].Chinese Journal of Distance Education,2019(08):27-34+58+93.

[11] Ding Guoyong,Qin Xinguo,Wang Xue.Research on predicting college students' performance in laboratory courses based on educational data mining[J].Chinese Journal of ICT in Education,2019 (13):41-44.

[12] Wu Xiuguo,Sun Tao.Research on Academic Early Warning for College Students Based on Educational Big Data Mining[J].Chinese Journal of ICT in Education,2020(07):55-57+62.

[13] Li Xiongping,Wang Li,Tan Xiongsu.Design of an Academic Early Warning System Based on Problem Diagnosis[J].Chinese Journal of ICT in Education,2020(19):50-52.

[14] Sun Jia,Li Wei,Wang Feifei,Wei Xu,Sheng Yejuan.Reflections and Explorations on Big Data-Driven Academic Diagnostics in Education[J].Chinese Journal of ICT in Education,2019(15):48-52.