

Research on Model Construction of Learning Analysis under Educational Big Data Environment

Ningguì Duan^{1, a}, Hong Lu^{2, b}

¹Information and Educational Technology Center, Youjiang Medical University for Nationalities, Baise 533000, China.

²School of International Languages and Culture, Youjiang Medical University for Nationalities, Baise 533000, China.

^aduanninggui@ymcn.edu.cn, ^b997491969qq.com

Abstract

Under the environment of education big data, learning analysis technology has been greatly developed, but a rigorous scientific system has not yet been formed. On the basis of analyzing the learning analysis technology at home and abroad, this paper puts forward six contents of the research and construction of the learning analysis model, and finally puts forward the research ideas and methods on the problems that need to be solved urgently during the construction.

Keywords

Learning analysis, big data, analysis model.

1. INTRODUCTION

Big Data is a kind of data set with a large scale that greatly exceeds the capability range of traditional database software tools in terms of acquisition, storage, management and analysis. It has four characteristics: massive data scale, fast data flow, diverse data types and low value density. In August 2015, the Chinese government issued the "Action Plan for Promoting the Development of Big Data" to promote the development of the big data industry and deepen the innovative application of big data in various industries. Education big data refers specifically to the data set in the field of education. Big data will be used in two specific fields of education: education data mining and learning analysis. The application of big data in the field of education is the premise of learning analysis, which is the application and change of big data in the field of education.

Learning Analysis (LA) is to collect, analyze and creatively apply data from educational management and service, teaching and learning by using technologies in the fields of information science, sociology, psychology, statistics, machine learning and data mining, so as to better improve educational practice. Therefore, how to make good use of education big data and explore the educational laws behind it has become the focus of researchers' attention recently. Since 2010, learning analysis has gradually become independent from the analysis field, absorbing various analysis methods such as data mining, social network analysis and statistical analysis, forming an independent emerging field (Wu Yonghe, Wu Qing, etc., 2015). Learning analysis, as the application and development of analysis technology in the field of education, has attracted more and more researchers' attention and attention, but has not yet formed a mature theoretical system, so this research has certain theoretical value.

2. SIGNIFICANCE OF LEARNING ANALYSIS AND RESEARCH

With the advent of the big data era, more and more data appear in learning, teaching and school management. It is necessary to use learning analysis technology to analyze the significance behind massive learning data. Faced with massive data, learning analysis technology processes and analyzes educational data from different perspectives, turning real-time decision-making of education and personalized learning of learners into reality (Zheng Yanlin, Zhang Hongmeng, etc., 2015).

Learning analysis is an applied technique, It is a technical "bridge" connecting improving learners' academic performance, predicting scores, digging potential problems, recommending personalized resources and improving practical teaching. It is also a link between learners, teachers and educational organizers. Different key technologies can be applied to recommend learning materials and learning paths suitable for personal characteristics to learners, feedback students' learning behavior analysis results and evaluation and diagnosis results for teachers, help teaching organizations and management agencies to improve the quality of educational decision-making and optimize the curriculum evaluation process (Gu Xiaoqing, 2014).

In short, learning analysis is applied to all levels of the education system, including micro, meso and macro levels, providing services to stakeholders (students, teachers, researchers, managers, policy makers and legislators). Among them, the micro-level is mainly aimed at the process data of individual learners, aiming at identifying risk learners and providing intervention, at the same time providing learners with the opportunity to observe their learning habits and giving suggestions for improvement. Stakeholders at the micro level mainly include students, teachers and researchers. The medium level focuses on the educational data at the system level, which is convenient to better understand the difference of variable influence, optimize the allocation of educational resources and improve the efficiency of organizational work. Stakeholders at the medium level mainly include education managers. Macro-level analysis aims to carry out cross-institutional analysis, merge medium and micro-level data, compare different teaching modes and management modes, and improve educational administrative decision-making. Stakeholders at the macro level mainly include policy makers and legislators.

3. RESEARCH STATUS AT HOME AND ABROAD

Judging from the papers related to learning analysis and research at home and abroad, The theory and application results of learning analysis research show an increasing trend year by year. Its researchers' nationalities involve the United States, Britain, Canada, Germany, China, the Netherlands, Australia, etc., covering all continents. The research field has received extensive attention from researchers around the world. Researchers from all over the world form a virtual academic group through network cooperation. In terms of the number of researchers, the United States has the largest number of researchers, which has been leading and promoting research and development in this field (Wu Yonghe, 2014)

The vast majority of documents focus on the micro-level, while the macro-level documents are the least. Students and teachers are the main body of learning analysis research, mainly focusing on how to use the results of learning analysis to enhance the effectiveness of teaching practice. However, with the annual growth of macro and meso studies, the number of documents shows an upward trend year by year. Education administrators, policy makers and legislators began to realize the importance and practical value of learning analysis in education data management, and released relevant research results mainly in the form of research reports.

However, the development of learning analysis is limited by relevant data acquisition and processing technologies and faces many challenges. First, data storage technology, data

processing and analysis technology are limited by the development level of computer hardware and computing power. Secondly, data acquisition technology and analysis technology are still in continuous development, and there is still a big gap between intelligence, automation and systematization. Third, the problem of data compatibility makes it difficult to realize cross-platform operation and data sharing of data. In addition, the current analysis of learning process data is difficult to realize three-dimensional analysis, and its analysis results and areas of play are still very limited (Gu Xiaoqing, Wu Fati, Guo Jiong, etc., 2016).

Most of the existing learning behavior data come from online (such as learning management system, browser data, educational APP, etc.). With the help of the learning behavior data extracted from these systems, learner models can be constructed to carry out research on performance prediction and learning resource recommendation. Huang Ronghuai of Beijing Normal University and Zhu Zhiting of East China Normal University regard learning analysis as one of the supporting technologies for building an intelligent learning environment, which is used to understand and optimize the learning process and learning situation. Gu Xiaoqing of East China Normal University applied learning analysis to the "Next Generation Internet Education Innovation Support System" project to monitor and evaluate the learning process. Wei Shunping of the National Open University selected the massive learning data stored in the school's learning management system to analyze the current situation and influencing factors of adult learners' online learning. Tsinghua University Zhang Yu applies learning analysis to MOOCs to promote the quality of education (Wu Qing et al., 2015).

At present, the most widely used learning analysis technology is "Dashboard" technology, which visualizes the data of teaching and learning. Arizona State University not only applies learning analysis technology and dashboard technology to teaching and learning, but also applies it to administration. It has established a dashboard system for intelligent analysis of the overall education and teaching business of the school.

China has constructed a comprehensive evaluation model based on learning analysis from the levels of teachers, learners, institutions and curriculum, and has carried out practical application. However, the continuous development of technology and methods require new adjustments to the evaluation framework. Therefore, the evaluation framework should be continuously adjusted according to the actual situation. The reference model for comprehensive evaluation of learners constructed by Zheng Qinhuo and others is not suitable for all courses and does not consider the acquisition of practical skills and other aspects. In future research, the adaptability and stability of the model should be strengthened (Li Ting et al., 2019).

The EU Learning Analysis Exchange Project has started research on learning analysis standards in 2014. A series of standards proposed by Chinese scholar Wu Yonghe and others on "Learning and Analyzing Interoperability Concepts", Combining with the process of learning analysis, From the perspective of interoperability related to data sharing, privacy, trust and data control, service quality, etc., it provides targeted terms and reference models for learners, teachers and educational institutions, and provides system requirements for a series of learning analysis processes such as collection, storage and processing, analysis and visualization of learning analysis abstract data streams.

Many online learning platforms and application software in China have accumulated a large amount of structured, semi-structured and unstructured data, but different systems and platforms are easy to form isolated islands of data. How to establish a unified data format standard and learning technical specification based on new technologies so that various platforms and systems can communicate, connect and integrate, thus providing standardized processes and system construction for data acquisition, screening, mining, analysis and presentation has become an urgent problem to be solved.

4. THE MAIN CONTENTS INVOLVED IN THE CONSTRUCTION OF LEARNING ANALYSIS MODEL

Through previous studies, Further comb and deepen the learning analysis theory and related theories, understand the current situation and development trend of learning analysis research, master the data collation and mining in learning analysis, conduct in-depth research on descriptive and predictive learning analysis research, and try to promote and apply the new data structure standard. At the same time, learning analysis is regarded as a whole system, which is no longer limited to the process of learning analysis, but on the basis of clarifying the internal elements and links of learning analysis, the theoretical framework of learning analysis and its interaction with the external environment are further discussed. Through a large number of empirical studies, the original learning analysis model is further revised and innovated, thus proposing a learning analysis model suitable for teachers and students in domestic universities. In order to systematically monitor and evaluate learners' learning process in a mixed learning environment, innovative technologies and learning tools are embedded into the teaching process through instructional design to provide academic early warning mechanism, thus improving teaching practice and providing educational reference for managers and decision makers at all levels.

4.1. Promote and Popularize the Common Standards for Multi-Online Platform Data to Enhance the Interoperability of Data

The integration of data distributed on various information islands, Carry out international comparative research on technical specifications of learning and analysis in due course, Fully grasp the characteristics and shortcomings of the existing norms, And then develop our own standards, This paper focuses on the analysis of the learning analysis technical specifications formulated by ADL, IMS and ISO/IEC. It compares the current international mainstream Caliper Analytics and xAPI learning analysis technical specifications from four aspects: data model, service interface and safe transmission mechanism, mapping relationship between standards and application system category, and analyzes the differences and internal relations between the two. At the technical level, the interoperability problem of different data sets is solved, and mature technical solutions are embedded into corresponding software systems.

4.2. Revise Data Mining Related Algorithms and Learning Analysis Models

Based on the analysis of the existing data mining and learning analysis model prototypes, since most of them are made by relevant research groups in European and American countries and applied by domestic scholars, it is found that they are not fully suitable for Asian learners, and some models will produce large deviations in practical application. This paper attempts to redesign or transform it with the relevant theories of mathematics and statistics, or to construct a new learning analysis framework model for learners with different characteristics. In addition, some unstructured data are modeled mathematically, and a learning analysis model is established to try to establish holographic data.

4.3. Carry Out Information-Based Teaching Reform and Conduct Multi-Source Data Aggregation Research

Actively carry out in-depth integration of modern information technology and education and teaching in colleges and universities. From the perspective of data sources, platforms for recording behavioral data include Moodle, Blackboard, MOOCs platform, education cloud service platform, etc. Judging from the types of data sources, the collected data include learning behavior data, emotional state, expression characteristics, brain cognitive activities, attention level, etc. Centralized learning environment data from student information system and online learning system are still the main analysis objects of learning analysis. These platform data are

gathered and then the teaching behavior of teachers and students is studied to obtain more comprehensive data.

4.4. Pay Attention to the Meso and Macro Research of Learning Analysis

At present, learning analysis focuses on micro-research, usually aiming at specific objects or activities, platforms or systems. It is difficult to transfer research results and explore the learning rules implied in the online learning process, thus appearing that the research topic is relatively narrow. Therefore, we should pay more attention to the research from the medium and macro aspects, try to make educational decisions, adjust teaching activities and improve the learning environment through data integration, and put forward suggestions to management departments and educational policy makers through a comprehensive evaluation of students' learning situation.

4.5. Establish A Standard for the Construction of Big Data Centers and Use Learning Analysis Tools to Carry Out Empirical Research

Propose the construction plan and learning of the big data center to the school, and the allocation of the corresponding learning analysis software system and technical tools also requires the investment of funds and resources to establish a data framework stationed at the school level. Some commonly used learning analysis tools at home and abroad, including WEKA, SNAPP, ELLment, etc., are selected. Combined with the actual situation of online teaching in China, 1-2 kinds of tools are used in the daily teaching process to carry out quantitative experimental research, which not only tests the applicability of learning analysis tools, but also compares the obtained data with offline teaching effect to judge the applicability of learning tools.

4.6. Establish A Learning Analysis and Prediction Model to Give Early Warning to Learners

Learning analysis technology is used to face the learning behavior process of learners, and evaluation, prediction and intervention are used as means to track, intervene and improve them. Through modeling, evaluation and intervention perspectives, this paper puts forward the idea of improving teaching practice with learning analysis. On the one hand, the research uses association rules, collaborative filtering and text mining methods to model and deeply analyze the data, so as to provide decision support opinions for curriculum developers, teachers and managers, so as to organize teaching resources more effectively and improve students' learning efficiency. On the other hand, based on sequential pattern mining, association rules, decision trees and other methods, the research can recommend personalized learning resources to students according to their historical behavior data and learning objectives, and provide individualized help and guidance.

5. PROBLEMS THAT MUST BE SOLVED IN THE CONSTRUCTION OF LEARNING ANALYSIS MODEL

5.1. Construction of Universal Learning Analysis Model

Although the current data model has certain predictability, it is difficult to predict students' academic performance more accurately due to the limitations of data collection. In the process of educational data mining, decision tree algorithm, classification, clustering, multiple linear regression and other data mining technologies are used in learning analysis to achieve various analysis purposes, such as predicting online learning performance, understanding the correlation between various learning influencing factors, finding problem students, and providing online evaluation tools for teachers. In addition, various mathematical and statistical methods are often applied together with data mining in the research of summarizing students'

characteristics, establishing scoring system, analyzing students' problem-solving behavior and improving intelligent tutor system. In terms of statistical methods, including descriptive statistics, ranking, variance, correlation analysis, regression analysis and other learning analysis. Based on various data, learning is described. Typical evaluation indicators include login times, time spent, click rate, type and number of accessed resources, personal works and opinions, test results, etc., but these can only reflect the local information of learners' online learning after all. How to correct the existing learning model according to our students' online learning habits is the key content of the research.

5.2. Empirical Research on Learning Analysis

This paper attempts to carry out learning analysis research with evaluation, prediction and intervention orientation as cases, trying to present the problem raising, research design and preliminary research findings of learning analysis research. The above learning analysis and research cases are all aimed at improving specific teaching practice. In the presentation of these cases, due to the different foothold of teaching improvement, the learning analysis methods and technologies adopted in the research design, as well as the analysis results and their application forms, show diversity accordingly. Combined with the case study, we try to further discuss the practical application of learning analysis technology to teaching improvement, including in-depth empirical research on the application of data analysis content and key technologies in the education big data environment.

5.3. Solve the Problem of Data Source Convergence

Educational data contains great analysis value. The establishment of student learning database and behavior database to store the process of interaction between students and learning content, as well as user behavior and personal characteristic information, is the premise of learning analysis. However, due to the complexity of the educational process and the diversity of learning and teaching situations, obtaining complete and procedural educational data has become one of the major bottlenecks in learning analysis technology. Facing the endless educational software, educational APP and the application of various management information systems in education, useful information is stored in databases of different platforms in isolation and cannot be obtained and used across platforms. This leads to major challenges at the technical level, including how to process big data, interoperate different data systems, and ask the right questions. How to quickly obtain these data and gather them on a larger data platform for analysis has greater value, but this will be a major challenge.

5.4. Learn to Analyze Technical Problems

How to mine the meaning "behind" of semi-structured and unstructured data and extract valuable information from education big data by using analysis technology has become another bottleneck in learning analysis technology. Due to the bad structure and complexity of data, it is necessary to transform multi-source heterogeneous data into interoperable data, and through complex and dynamic processing and analysis processes, these multi-source heterogeneous data show readability and semantics. This process requires continuous adjustment and optimization of data processing technologies and strategies, which is a major challenge for learning analysis technology. Moreover, the traditional SCORM data structure can only record the learning experience data based on the network session. The learning experience recorded by SCORM has certain limitations, lacks continuous and environmental information, and is not suitable for the collection and recording of mobile terminal data. XAPI defines a simple, lightweight technical specification that allows any participant to store and retrieve extensible learning experiences regardless of the specific implementation platform. It provides an interoperable data model for storing students' learning experience data, and realizes the standardized encapsulation and intercommunication of data between formal learning

environment and informal learning environment, online and offline learning environment, and different learning devices and platforms. This new standard still needs to be vigorously promoted and applied.

6. RESEARCH IDEAS AND RESEARCH METHODS FOR THE CONSTRUCTION OF LEARNING ANALYSIS MODEL

We need to take the research on learning analysis technology under the environment of educational big data as the breakthrough point, analyze the history and current situation, development stage and research object of learning analysis, and use learning analysis tools on the network platform to reveal many problems in the current learning analysis research field. After that, it tries to explore the data structure standard, data collection and convergence, analysis model construction and prediction analysis model of learning analysis. This paper puts forward planning requirements for the construction of big data centers in colleges and universities, and further proves relevant theoretical results in empirical research, from which the general rules and successful experiences of learning analysis research are explored, so as to provide possible references and references for learning stakeholders. Using the methods of literature analysis, comparative analysis, empirical analysis, interview, etc., starting from the perspective of learning analysis technology, under the guidance of machine philosophy, combined with the theories of educational technology, computer science, mathematical statistics and other related disciplines, this paper makes a quantitative and qualitative comprehensive research.

7. CONCLUSIONS

By studying the above-mentioned model construction of learning analysis, we hope to further improve the theory and technical system of learning analysis, provide theoretical basis and technical means for further teaching research under the big data environment, and realize the great goal of education informatization.

ACKNOWLEDGEMENTS

His paper was supported by the 2020 Guangxi Young and Middle-aged Ability Improvement Research Project "Research on Academic Early Warning Mechanism for Foreign Students Based on Learning Analysis Technology" (No:2020KY13001); 2018 key teacher training plan of Youjiang Medical University for Nationalities (YYYYZ [2018] No. 45)

REFERENCES

- [1] Adjei,S.A., Botelho,A.F.,&Heffernan,N.T.(2017). Sequencing content in an adaptive testing system: The role of choice[A].Proceedings of the Seventh International Learning Analytics & Knowledge Conference[C].Vancouver, British Columbia,Canada,ACM: 178-182.
- [2] Aghababayan,A.,Lewkow,N.,& Baker, R. (2017). Exploring the asymmetry of metacognition [A]. Proceedings of the Seventh International Learning Analytics & Knowledge Conference[C]. Vancouver, British Columbia, Canada, ACM: 115-119.
- [3] Allen,L.,Perret,C.,Likens,A.,& Mcnamara, D. (2017). What'd you say again?:Recurrence quantification analysis as a method for analyzing the dynamics of discourse in a reading strategy tutor [A]. Proceedings of the Seventh International Learning Analytics & Knowledge Conference[C].Vancouver, British Columbia, Canada, ACM: 373-382.
- [4] Swan M.Sensor mania! The internet of things,wearable computing, objective metrics,and the quantified self 2.0[J].Journal of sensor and actuator networks,2012,1(3):217-253.

- [5] Vanlehn K,Zhang L,BNurlson W,et al.Can an non-cognitive learning companion increase the effectiveness of a metacognitive learning strategy?[[]. *EEE transactions on learning technologies*,2017,10(3):277-289.
- [6] Catrysse L, Gijbels D,Donche V,et al.How are learning strategies reflected in the eyes? Combining results from self reports and eye-tracking[[]. *British journal of educational psychology*, 2017, 88 (5): 118-137.
- [7] MOU Zhijia, WUFATI, George Simmons. Research Status and Trend Analysis in the Field of Learning Analysis Abroad [[]. *Research on Audio-visual Education*, 2016 (4): 18-25.
- [8] WU Qing, Luo Ruguo. Learning Analysis: From Origin to Practice and Research [[]. *Open Education Research*, 2015 (2): 71-79.
- [9] Wu Yonghe, Chen Dan, Ma Xiaoling, et al. Learning Analysis: The New Wave of Educational Informatization [[]. *Journal of Distance Education*, 2013 (4): 11-19.
- [10] Gu Xiaoqing, Zhang Jinliang, Cai Huiying. Learning Analysis: Emerging Data Technology [[]. *Journal of Distance Education*, 2012 (1): 18-25.
- [11] Li Xiangyong, Zuo Mingzhang, Wang Zhifeng. Research status and future prospects of learning analysis-review of the 2016 International Conference on Learning Analysis and Knowledge [[]. *Open Education Research*, 2017 (1): 46-55.
- [12] Gu Xiaoqing, Liu Yan, Hu Yiling. Application of Learning Analysis Technology: Learning Improvement Scheme for Seeking Data Support [[]. *Open Education Research*. 2016 (5): 34-45.
- [13] Wu Yonghe, Li Ruochen, Wang Haonan. Current Situation and Future Development of Learning Analysis Research-Comments on 2017 International Conference on Learning Analysis and Knowledge [[]. *Journal of Distance Education*, 2017 (5): 42-56.
- [14] MOU Zhijia, YU Xian. Academic Groups and Hot Spots Tracking of Learning Analysis from the Perspective of Knowledge Mapping Analysis-Meta-analysis of International Conferences on Learning Analysis and Knowledge over the Years [[]. *Journal of Distance Education*, 2016 (2): 54-63.
- [15] Li Qing, Ren Yuan, Huang Tiantian. Study on the application of learning analysis based on sensor data [[]. *Research on Audio-visual Education*, 2019 (5): 64-71.