

# Analysis and Designing of Educational Big Data Mining System for Higher Education

Jing Dong\*

School of Information Engineering, Qujing Normal University, Qujing, China

Dongjing@mails.qjnu.edu.cn

**Keywords:** Big Data; EDM (Educational Data Mining); Cloud Service; Hadoop

**Abstract.** In order to realize deep excavation and application of educational big data in higher education, multi-disciplinary theories such as pedagogy, computer science, statistics and informatics are helpful. Hadoop of big data mining cloud service is utilized to construct big data mining system for higher education. Data mining technology and statistical analysis are used to deal with the massive data collected during the undergraduate teaching evaluation of Qujing normal university. Making full use of the system can achieve more accurate and efficient management for universities, also including intelligent teaching and learning. Besides, it can improve the quality of education and teaching in higher education.

## Introduction

With the popularity of information and communication technology as well as continuous maturity of data mining technology, the era of big data has come [1]. The application of big data in the field of education has also begun to be concerned by much organizations and professional staff. Compared with traditional educational data, educational big data is characterized by high real-time, fine granularity, authenticity and decision making [2,3]. Educational big data is produced in all aspects of teaching and learning, including various types of structured, semi-structured, and unstructured data sets with big data features, such as teaching data, management data and scientific research data, it's possible to utilize data mining and analysis, cloud computing and other information technologies to excavate value of educational big data.

The application of educational big data is mainly divided into educational data mining and analysis based on learner [4]. The educational data mining mainly uses data mining, mathematical statistics and machine learning to process and analyze mass educational data and to establish a scientific data model to discover the relationship between the learning results of the learners and a series of variables, such as the learning content, the learning resources and the teaching behavior, and excavate the students' learning. As a result, teachers can really understand each learner and to provide accurate and targeted teaching according to their personal characteristics such as learning habits, learning interests and learning preferences. As interdisciplinary comprehensive application such as computer science, information science, education, psychology and other disciplines, analysis based on learner deals with and analyzes mass educational data, and then evaluates the learning behavior of the learners, explains the important factors that affect the effect of learning utilizing data models and methods, at last, provides the learners individual adaptive feedback, such as the most suitable courses or textbooks for learners.

There are fewer researches on the application of big data in Higher Education, and some primary researches are lack of relevant data support [5,6]. With the increasing amount of data accumulated in the information system of educational institutions, managers can't obtain effective and high-quality information when choosing resources, which leads to the phenomenon that some effective information can't be exploited and exploited, and "information isolated island" appears [7]. Data mining technology excavates potential as well as useful information from the mass data of educational information to benefit all participants belong to education system, and provides the basis for solving the semi-structured and unstructured decision-making problems in the field of education.

### Key Technology

**Interdisciplinary.** In order to realize deep mining and application of higher educational big data, it's necessary to apply theory of pedagogy, computer science, statistics and informatics comprehensively [8]. Recent years, historical changes have taken place in the field of education and information, and a variety of online learning platforms, smart phone applications, and social networks provide a large number of applications and massive data for researchers [9,10]. Unfortunately, these data haven't been paid enough attention and rational use by many local colleges [11,12]. Many information technologies such as cloud computing, Internet of things, mobile computing and large data technology provide technical support for the analysis and mining of educational big data, which can turn these idle data into a treasure, that is, to realize the effective processing of data technology and excavate hidden values.

Cloud computing service solves the information island phenomenon during informationization construction in Colleges [13]. It can integrate scattered educational resources so that teachers and students can share educational resources stored on the cloud server, in addition, cloud service connects with all kinds of terminal devices, so that all kinds of decentralized relations, structured and unstructured data is analyzed and excavated by data mining technology, data analysis and mathematical statistics[14]. The information obtained from data mining technology can accurately analyze and evaluate the quality of the class, also realize accurate and efficient management of college, which makes the teaching and learning more intelligent. Teachers can even carry out accurate classification and personalized guidance to the students across the college; what's more, students enjoy recommendation of personalized and accurate resources, including personalized mobile phone content push and personalized evaluation of students, which effectively improves teaching efficiency and quality.

**Data Resource.** At present, inefficiency of education funds and the ratio of education input-output contribute problems of backward teaching organization, such as teaching content, teaching method, examination system and talent training mode, which causes students' personality characteristics and cognitive development are ignored. Qujing normal university has just completed the review and evaluation of undergraduate teaching in ordinary universities, and each college including auxiliary departments have collected massive data, covering all aspects of school teaching and service. It is difficult to achieve accurate teaching and learning and quality evaluation for teachers and students. The research is mainly based on the latest teaching state data collected during the undergraduate course evaluation. In addition, the school has set up 10 courses, such as the public relations etiquette practice, and other 9 good college online courses such as the Tang poetry and Song Ci poetry, and so on as many as dozens of MOOC courses. The process data of these teaching can be realized during the class, which can do precise analysis service for our school teachers and students.

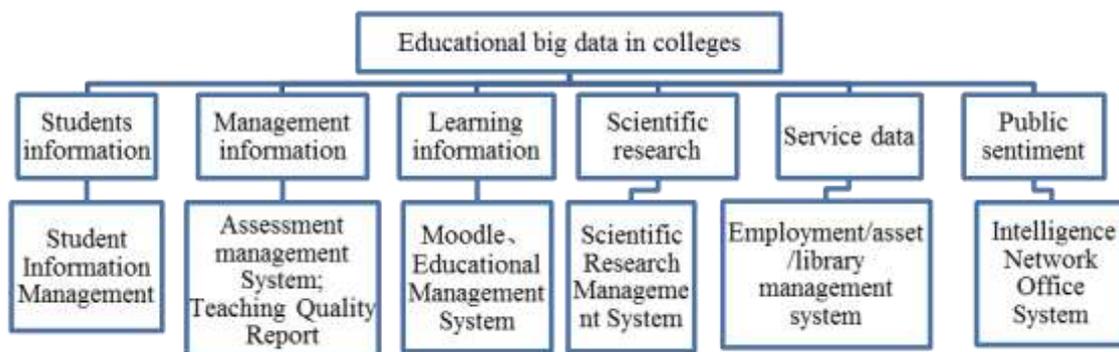


Figure 1 Source of Educational Big Data in Colleges

From Fig.1, students' basic information such as enrollment and source is stored in student information management; management data involves learning effect, teaching quality assurance

system, quality control and teaching evaluation; educational data involves teaching facilities, classroom teaching, teaching reform, curriculum resources, MOOC courses and online teaching platform; teachers' research achievements, practice teaching and college students' innovation and entrepreneurship projects stored in Scientific Research Management System; Service data involves student guidance and service, library service, and student online learning log data; Public sentiment involves UGC data such as WeChat, micro-blog, forum.

**Framework of Educational Big Data Analysis System.** Educational big data analysis system uses Hadoop big data mining cloud services, which is an open source distributed computing platform for analysis and mining massive unstructured data as well as semi-structured data. Hadoop integrates kinds of parallel data preprocessing and mining algorithms such as K-mean, the system is shown as the figure below.

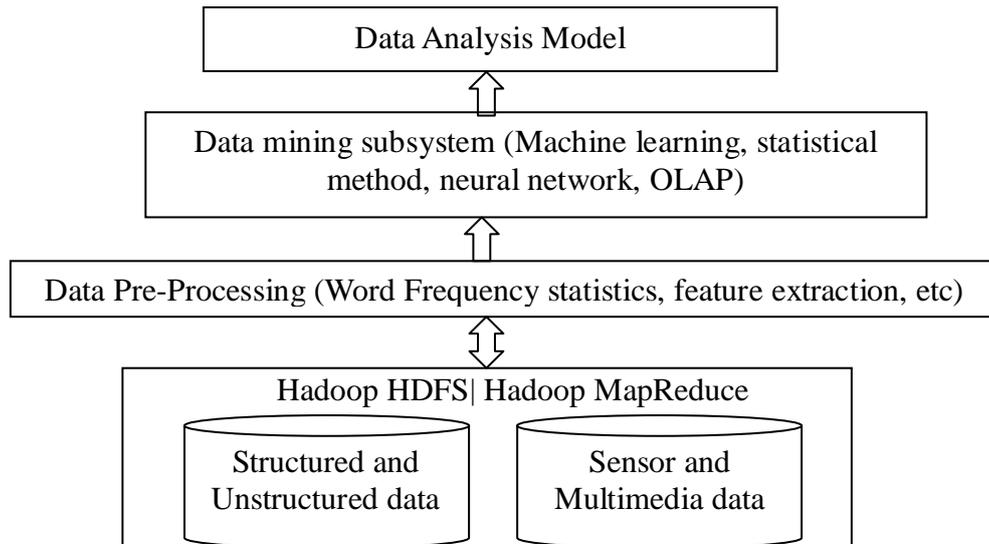


Figure 2 Educational data mining cloud service platform framework

**Data Mining and Mathematical Analysis.** Data mining is a process of excavating the potential useful information that is hidden in large, noisy, incomplete, and random data. Data mining is a process of excavating the potential useful information and knowledge that is hidden in massive, noisy, incomplete, and random data. The original data has characteristics of real and noisy. Data mining technology extracts, transforms, analyzes massive data stored in database, and then shows important information that users are interested in and help the managers to make decisions. The research makes modelling by statistical method of regression analysis, clustering analysis, principal component analysis and association rules mining, machine learning such as support vector machines and biomimetic method of data mining to obtain hidden information hidden in the educational big data, which takes data about qujing normal university as the original data source. SPSS Clementine data mining software is used to construct the students' characteristic behavior analysis module, student employment factor analysis module as well as learning result prediction module, besides, teaching quality monitoring and evaluation, the analysis module of influence factors of university management development and the example simulation is constructed as well.

## Conclusions

Making full use of data mining technology to deal with educational big data in higher education can integrate information and eliminate information isolated islands; what's more, researchers can realize horizontal correlation comparison and longitudinal historical analysis to provide accurate service and support scientific decision-making. Utilizing the information obtained from data mining technology can accurately analyze and evaluate the quality of the classroom teaching, and realize the more accurate and efficient management of the universities, which makes the teaching and learning more intelligent. Tracking students' learning process can dynamically excavate

characteristics of their interest, personality and emotion, which help researchers to establish the resource recommendation to users, and finally realize the personalized precision filtering. At last, administrative staff can make comprehensive, dynamic and objective evaluation about teachers' teaching quality and effect. The value of educational big data can be excavated by data mining, statistics analysis and visualization technology, which promotes education development and innovation; in addition, it improves the quality of education and teaching.

### **Acknowledgements**

This work was supported by Philosophy and Social Science Educational and Scientific Planning Office of Yunnan Province of China (Application Research on Educational Big Data to Influence Management Decision Making of Higher Education, Grant No. AD17012)

### **References**

- [1] Information on <http://www.slideshare.net/BemardMarr/a-brief-history-of-big-data>.
- [2] Information on <http://www.Gartner.com/it—glossary/big—data>.
- [3] X.F. Yan, D.X. Zhang: Big Data Research, Computer Technology and Development, Vol.8 (2013) No.4, p.1.
- [4] Y.N. Zhang: A Study of Big Data Applications In the Field of Education Based on the Practice in America (Ph.D., East China Normal University, China 2016), p.28.
- [5] James Manyika: *Big data: The next frontier for innovation competition and productivity* (Insights & Publications, U.S.A 2011), p.56.
- [6] T.S. Chang: Developing an institutional intelligence system: A new trend of institutional reaserach. Journal of Higher Education, Vol.30 (2009) No.10, p.49. (in Chinese)
- [7] S. Pedersen, T. Irby: The VELscience project: Middle schoolers' engagement in student-directed inquiry within a virtual environment for learning, Computers & Education, (2014) No.71, p.33.
- [8] S.J. Liu: Proc. *Information Technology, Computer Engineering and Management Sciences* (Tamil Nadu, India, September 8-10, 2011). Vol. 4, p.254.
- [9] J.A. Lara, D. Lizcano, M.A. Martínez: A system for knowledge discovery in e-learning environments within the European Higher Education Area – Application to student data from Open University of Madrid, UDIMA, Computers & Education, (2014) No72, p.23.
- [10] Y.H. Wang, H.C. Liao: Data mining for adaptive learning in a TESL-based e-learning system. Expert Systems with Applications, Vol.38 (2011) No.6, p.6480.
- [11] S. Natek, M. Zwilling: Student data mining solution–knowledge management system related to higher education institutions. Expert Systems with Applications, Vol.41 (2014) No.14, p.6400.
- [12] M.M. Chang, M.C. Lin: The effect of reflective learning e-journals on reading comprehension and communication in language learning, Computers & Education, (2014) No71, p.124.
- [13] Y.G. Wang, Q. Zhang: MOOC: Characteristics and learning mechanism, Education Research, (2014) No9, p.112. (in Chinese)
- [14] W. He: Examining students' online interaction in a live video streaming environment using data mining and text mining, Computers in Human Behavior, Vol.29 (2013) No1, p.90.