A Brief Analysis of the Key Technologies and Applications of Educational Data Mining on Online Learning Platform

B. V. Ramana Reddy

Department of Computer Science and Engineering, Nalanda Institute of Engineering and Technology, Guntur, India. 522438
Email: busireddy100@gmail.com

Abstract

With the rapid development of the Internet and communication technology, online education has drawn more and more attention, online learning platforms, on the other hand, store massive learner behavioral data and educational data. How to effectively analyze and utilize the data to improve the quality of online education has become a key issue urgently needed to be solved in the field of Big Data in Education (BDE). Educational Data Mining (EDM) is exactly an effective and practical method and means of applying BDE. Therefore, EDM is an important academic research hotspot in the field of EDM. Firstly, the paper introduces the basic concepts of BDE, EDM and online learning platform, and then elaborates on the process of how educational data mining transforms raw data into knowledge. Finally, the key technologies of data mining are classified according to their uses, and gives its application in the online education scene. The paper can provide some guidance for the research and application of educational data mining based on online education.

Keywords: Component; EDM; Online Education; BDE; Online learning platforms

I. Introduction

With the explosive development of big data, big data has started a major transformation of the times, and great changes, even substantial changes and developments in economics, politics, sociology and many science categories have taken place [1]. Big data also has a profound impact on the education industry. Online education occupies an extremely important position in the field of education, and produces massive online and observable data at all times [2]. Therefore, the combination of big data and online education has become an inevitable requirement for the development of the times. At present, the most difficult part of the traditional education mode lies in the collection and analysis of data. The online education platform stores vast amounts of data that are real time, monitorable and quantifiable, including learner behavior data and the educational data of the platform [3]. However, the data itself is only data, did not intuitively show any valuable knowledge, how to use these educational data to transform these data into information and knowledge and provide services for educational decision making, teaching optimization and academic improvement is the value of BDE. The difficulties of processing technology of big data mainly focus on the complexity of data, and EDM is an important technical means to realize the application of BDE [4]. Through the analysis and mining of the relevant data in the online learning platforms, the relationship between learners' behaviors and their inner logical relationships are found, and the teaching plan is better formulated based on it. However, big data is in a nascent stage, for educators, the concept of big data and its applications are vague [5]. First, the paper elaborates the related concepts of BDE, EDM and online education platform, so that educators have a deeper understanding of the relationship and development of those concepts. Based on this, this paper mainly includes two aspects: educational data mining process and key mining technology classification: the educational data mining process mainly explains how to extract raw data from the database to provide value for education decision-makers and learners-key mining technology classification is to analyze the technologies used in educational data mining process, the paper not only categorizes these technologies but also attempts to explain the differences and connections between them, namely, when different technologies should be applied to a scenario. Finally, we put forward some prospects of EDM in online education.

II. BDE, EDM and online learning platforms

A. The Concept and Basic Connotation of BDE

Big data refers to a large scale of data which is rich in information in different forms of media, network and other media [5]. BDE is a subset of big data, which refers to data in education. In fact, big data is a nebulous concept that has not yet formed an accepted definition. Even so, there is a difference between big data and past data, and its basic connotation can be summed up in 4V[6] Volume, Variety, Value and Velocity, of which specific meanings are shown in table I.
B. The Concept and Purpose of EDM

Data mining is a process of extracting the information and knowledge from large, incomplete, fuzzy and random data in which people are not known in advance, but potentially useful information and knowledge [7]. In recent years, with the popularization of educational informationization and the rise of new online education model represented by Massive Open Online Course (MOOC), more and more researchers are concerned about the EDM [8]. EDM is actually the application of data mining technology in education and the object of analysis and processing is to BDE, its purpose is to find out and solve the problems in education by using the key technologies of data mining to mine big data in the online learning platform, to obtain the important and valuable information in the teaching process and to improve the quality of teaching and learning.

<table>
<thead>
<tr>
<th>Type</th>
<th>Interpretation</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume</td>
<td>Huge amounts of data</td>
<td>The data volume is generally at the TB level, or even PB level</td>
</tr>
<tr>
<td>Variety</td>
<td>Many types of data</td>
<td>Data sources can be data such as documents, web logs, video and audio, images, and other common data types, and can also be generated from the Internet, Shared cloud, Internet of things, etc.</td>
</tr>
<tr>
<td>Value</td>
<td>Low value density</td>
<td>The availability of useful data on decisions may be one thousandth or less of a huge amount of data.</td>
</tr>
<tr>
<td>Velocity</td>
<td>Quick input and output</td>
<td>Feature data is growing fast, processing speed is fast, and the time efficiency is high, which is the distinguishing feature of big data difference from traditional data mining</td>
</tr>
</tbody>
</table>

C. Online Learning Platforms and Types

Online education is a web-based learning behavior. Through relevant online learning platforms, learners achieve a timely, resource-sharing and zero-distance learning experience and approach [9]. Currently, the education platforms can be divided into three categories. The first is the traditional online education platform, it is mainly distributed in the form of video and audio through television or the Internet, and is widely used in remote areas and areas with poor educational quality. The second is a new type of online education platform represented by MOOC [10], the value orientation of MOOC's initiative is that "anyone can learn high quality educational resources at any time and anywhere". Based on EDM, MOOC has diversified learning tools and rich curriculum resources, integrates technological research and application innovation, and uses efficient precision tests and extensively connected social networks to provide learners with learning experience anytime, anywhere. The third is Small Private Open Course (SPOC), which corresponds to MOOC. SPOC is not an alternative to moocs, but rather the inheritance, perfection and transcendence of moocs [11]. It absorbs and inherits advanced ideas of MOOC, such as high quality resources, accurate testing, timely feedback and personalized recommendation, and so on. In response to the high dropout rate of MOOC and the lack of immersive learning experience, SPOC has changed the conditions for the application of the course and increased the participation rate of the learners and teachers. The word "private" in SPOC enables learners to have a sense of ownership and a sense of urgency. The word "small" in SPOC allows the learner's participation in learning to improve greatly.

III. Data mining on education and its process

Figure 1 shows where EDM transforms data into knowledge during data analysis and processing. From the perspective of generalized data mining, the processing flow includes 3 stages: data preparation, data mining and analysis and evaluation. From an educational point of view, this is a cycle of discovering knowledge from data generated by the educational environment and reusing it to improve the educational environment. From the point of view of data positioning, the data of each stage may transform the data of another stage of other targets.
a. Preparation of Educational Data The foundation of EDM is that data, mass, richness and quality of educational data are crucial to the outcome of EDM. Educational data preparation generally includes data integration, data selection and preprocessing. Online education data exist in the form of structured data (such as student database, etc.), unstructured data (such as teaching videos, English audio and student images, etc.) and semi-structured data (such as advisory mail, Java works and teaching resource repository, etc.). But not all data sources can carry out data mining, but selecting the corresponding data based on data mining purposes as the data source of this educational data mining, which is the process of data selection. The collected data often contain some important information missing, incorrect or containing noise, inconsistency and other issues, data preprocessing technology can improve data quality to meet the requirements of educational data mining, common preprocessing operations can use data cleansing (filling in missing values, removing outliers, etc.), data reduction (reducing high-dimensional data to low-dimensional, etc.) and data transformations (transforming data into another form of expression, etc.). Data preparation is the basis for follow-up work and directly determines its quality and efficiency.

b. Educational Data Mining EDM uses some of the key technologies of data mining to improve the quality of online learning by modeling BDE and finding out the correlation between learner's academic performance and learning behavior, teaching purpose and teaching strategy[12]. In order to achieve the goal, the common key technologies of education data mining are generally divided into association rule, regression and prediction, classification, clustering and diagnosis. For example, Apriori is a typical algorithm for association rule. Using Apriori algorithm, it is found that there is a relationship between the performance of the student and their involvement in doing assignment, internal assessment tests, attendance etc., which would help to identify the average and below average students and to improve their performance to provide good results[13]. Logistics prediction method is a common algorithm for predict mining, teachers can establish achievement prediction model by typical learner behavior indexes, through the prediction model teachers can understand the learning situation of the learners in advance, give learners precise guidance and intervention, let learners return to the normal learning track[14]. The above two examples show that different data mining algorithms should be used in different teaching scenarios. Therefore, according to the types of online education data and the teaching problems to be solved the suitable data mining algorithms are selected. And deciding how to use the algorithm on these data not only increases the efficiency of research but also improve the value of knowledge.

c. Expression and Interpretation of the Results The expression and interpretation of results refer to the analysis and summarization of the results of the excavated data, and put forward valuable rules and restore them to the meaning people can understand. The expression and interpretation of results is the most widely used visualization technology. Although there is some understanding of the data after mining and statistics of BDE, it not intuitive. The results visualized by the form will show the characteristics of data, so that educators can clearly understand the results of digging out and make a precise teaching decision, such as, learners' accuracy rate, time of study, learning curve and so on. The results obtained at this stage are valuable knowledge, for learners, the results can be established learning knowledge model to provide learners with detailed learning feedback and advice; for the education, the relationship can be summed up online teaching strategies and learning results, and then improve their teaching methods.

IV. The key technologies of DME

DME can build models through analyzing and mining education-related data. These models can be student performance prediction models, adaptive curriculum recommendation models, etc. In the field of education, data mining technologies are generally divided into five categories according to the technology data types and problems to be solved, namely, association rule, regression and prediction, classification, clustering and diagnosis. This article will then briefly introduce the 5 technologies and give their application scenarios in online education.
a. Association Rule The goal of association rules is to discover the association or correlation between the data set, and the most famous of them are Apriori algorithm and the FP-Growth algorithm. Association rule mining is widely applied in online platforms. It uses association rules to reveal the internal rules of teaching, student development mode and discipline correlation rules, so as to better organize teaching activities.

b. Regression and Prediction The regression method is a mathematical method to find the correlation between dependent variables and independent variables. Regression analysis is divided into linear regression, multiple regression and nonlinear regression. In linear regression, the data is modeled using a straight line, multivariate regression is an extension of the linear regression, involving multiple variables, nonlinear regression is a nonlinear model with adding polynomial form in basic linear model. Logistic regression algorithm is a widely-used regression and prediction algorithm. For example, when analyzing the relationship between learner's behavior index and learning completion rate on the MOOC platform, the correlation coefficient of the indexes can be calculated by using the logistic regression algorithm to obtain the learning prediction model.

c. Classification Classification is a supervised learning process. Its purpose is to construct a classification function or classification model based on the characteristics of the data set, which can map unknown types of samples into a given category. There are many classification methods, the most common are KNN, neural networks, support vector machines and decision trees. For example, the online learning platforms have a knowledge recommendation system. When learners search for a certain course, the system will automatically recommend similar courses so that learners have more personalized choices. The labeling of these knowledge points is a typical application of classification.

d. Clustering The process of dividing a collection of physical or abstract objects into clusters with similar objects is called clustering. Each object is highly similar to an object in the same cluster and less similar to objects in other clusters. The difference between clustering and classification is that clustering is not grouped according to specific criteria but inherent logic between data. No one knows if the data will be divided into several groups or what kind of group. The common clustering methods include K-means, Hierarchical Clustering, Gaussian mixed clustering and so on. For example, there are differences in the level of many learners in online learning process. Teachers use clustering methods to cluster students who have similar learning backgrounds, providing a more reliable quantitative basis for students' overall level analysis and targeted guidance.

e. Diagnosis In data mining, the goal of the diagnostic method is to find a small number of data objects that are abnormal in the data set, known as outlier. Outliers are usually considered abnormal or noisy data, but sometimes these data objects contain extremely important information. Diagnostic methods can be based on statistical methods, methods based on distance or proximity, and methods based on density. In SPOC, after the diagnosis analysis method, teachers find that a learner is divided into low class, but its mathematical index is extremely high, or one's performance has been stable for a certain period of time, but at a point of time falls sharply: these outliers may hide important information, teachers can launched personalized teaching activities according to the specific situation of individual learners.

V. Conclusion

Big data is of revolutionary significance to online education. With such huge and complicated data on the online learning platforms, it is particularly important to study and apply these BDE. From the perspective of academic research and technology application, the paper elaborates the basic concepts of BDE, EDM and online education, the detailed process of educational data mining, the classification of educational data mining technologies, and the need to pay attention in the process of mining. For example, different algorithms are selected according to different application scenarios, and examples of specific scenarios used by some key technologies are given. The purpose of this paper is to enable other researchers or educators to gain a better understanding of the ways in which educational data mining is used in the education of big data and better utilize educational data mining techniques to improve the quality of online education. The development of big data has brought many opportunities for online education, the emergence of new mining technologies will certainly provide better methods and means for the application of EDM in the field of online education. Using these new technologies will also lead to more scientific analysis and higher-value knowledge, and better service for online education. However, care must be taken to address the managerial, ethical and technical challenges and limitations of educational data mining, all of which need to further strengthen the research on the educational data mining of online education.
References