Advances in Social Science, Education and Humanities Research, volume 290
2nd International Conference on Economic Development and Education Management (ICEDEM 2018)

Research on the Application of Data Mining in Teaching Based on Statistics

Juan Luo
Yunnan College of Business Management
Kunming, Yunnan, China

***Abstract*—This paper briefly discusses the status quo of teaching evaluation in new undergraduate colleges, and uses the university teaching evaluation system as the data mining platform, and the university teacher evaluation data as the training set, focusing on the application research of data mining technology in college teaching evaluation system. The research shows that the data mining results enable teachers to find out the factors affecting the quality of teaching and the factors affecting the quality of students' learning, which plays a positive role in improving teachers' teaching methods and improving teaching quality. It has certain reference value for college teaching evaluation.**

Keywords—Application; Data mining; Teaching; Statistics

I. INTRODUCTION

With the rapid growth of student achievement records based on the campus network educational management data warehouse, it is difficult for current educational administrations to find out the rules directly according to the student's performance data distribution and make decisions according to this law. Therefore, it is necessary to automatically discover the hidden rules or patterns in the data by means of the corresponding data warehouse analysis tools to support the decision. Data mining technology can be used to find the hidden rules or relationships between data from a large amount of data. It is usually automatically identified by the machine and does not require more manual intervention. Using data mining technology, it can provide intelligent and automated assistance for user decision analysis.

In recent years, with the rapid development of computer technology and network technology, the teaching information management system of colleges and universities has been greatly developed and widely applied. At present, all domestic universities have been equipped with information management systems to varying degrees. Most of these teaching management systems use database technology and network communication technology, which basically include student management, teacher management, course management and performance management. functional module. In the teaching management system, a large number of records and data generated in the teaching process are stored and managed in the database, which improves the shortcomings of the traditional paper recording methods such as easy loss, easy damage and inconvenient query, and saves paper and improves management efficiency. At the same time, it can be economical and environmentally friendly. On the other hand, the application of network technology in the teaching management system enables the transmission, processing and query of teaching information to be completed remotely, which improves the flexibility of teaching management. The emergence of the information-based teaching management system has greatly facilitated the teaching management of colleges and universities, improved the operational efficiency of college management, and reduced the cost of running schools.

However, in the application process of the teaching management system, the system will store a large amount of data, such as basic information of students and teachers, student achievement and so on. If effective and organic use is not available, these massive amounts of data are simply stored in the management system's database, which will likely turn vast amounts of data into useless garbage, causing so-called "data explosions, lack of knowledge." "The phenomenon. In fact, these massive amounts of data often have some potential connections and objective laws within and between each other. Effectively discovering and utilizing these links and laws to analyze and evaluate the quality of teaching, decision support for university management, etc. The help of the teaching management system thus plays a greater role. Data mining technology is a technology that analyzes hidden relationships and laws in massive data and obtains useful information from them.

II. THE CONCEPT OF DATA MINING

With the rapid development and wide application of information technology, the application database system of various industries preserves and manages a large amount of data, but most database systems can only provide some simple data management and processing functions. On the other hand, with the development of society, the importance of data is becoming more and more obvious, and the demand for data analysis and processing is becoming more and more intense. These requirements are difficult to achieve by using traditional, manual data analysis methods and database systems. . With the explosive growth of data in various industries, the phenomenon of “data explosion, lack of knowledge” has become more serious. Therefore, in the face of massive data, people are eager to have a scientific system of technology that can be used to analyze and process this data, thus discovering valuable information contained in massive data to serve decision-making.

Data Mining refers to the process of analyzing and extracting the knowledge of people from massive data or databases. This knowledge is some potentially valuable information, which can generally exist in the form of concepts, rules, regularities, patterns, etc. [1]. For data mining, another more authoritative definition is: data mining refers to the extraction of hidden data from a large number of incomplete, noisy, fuzzy, random, and practical application data. , but is a potentially useful process of information and knowledge [2]. In a nutshell, data mining is the process of analyzing massive amounts of data and mining knowledge from it. “Excavation” vividly represents the process of discovering useful, high-value data from a large amount of unprocessed, low-value data. "Knowledge" can be thought of as concepts, rules, rules and patterns, that is, valuable and interesting information extracted from massive and complicated data. These "knowledge" can be used to discover data rules, provide decision support, etc., and data mining technology is an effective means to achieve this process.

Data mining technology involves the intersection of multiple technologies and fields, including database technology, artificial intelligence, machine learning, artificial neural networks, statistics, pattern recognition, information retrieval, and high-performance computing technologies. With the rapid development of information technology, the amount of data in various industry databases will continue to maintain exponential growth, and the demand for massive data analysis and processing will continue to increase, and the research and application of data mining technology will have greater development. .

III. DATA MINING METHODS

The analysis method of data mining technology, that is, the task classification of data mining, mainly includes the following:

A. Analysis of Association Rules.

Association rule analysis is used to find correlations between data items in large amounts of data. Its general representation is: A => B, that is, the data item that satisfies A is also likely to satisfy B. According to different association rules, different rules of data can be reflected and used to predict the occurrence or development trend of events. The coverage of an association rule is the number of instances that the association rule can correctly predict, called support. Accuracy or confidence is the representation of the number of correctly predicted instances as the proportion it occupies in all instances involved in the association rule application [3], ie the probability that an association rule is predicted to be accurate. Association rules analysis is widely used, such as the famous "diapers and beer" story. When people analyze the sales records of Wal-Mart stores in the United States, they find that male customers often buy beer while purchasing baby diapers, thus making baby diapers and The decision to put the beer together has finally achieved good results. Therefore, finding the association rules from a large amount of data has a significant effect on the decision support of commercial activities such as market sales. At the same time, association rule analysis is the basis of many other data mining methods, such as classification.

B. Classification Analysis

Classification analysis is the classification of data items in a large amount of data based on a given category, that is, constructing a classification function or model, mapping the data items to one of the given categories, and using the classification rules to perform unknown data. The classification is predicted, and the classification function or model is based on a training set that has been classified into categories in the data. Classification analysis is generally divided into two stages: First, a function or model is to be established to describe a known data classification rule, that is, a classification function or model is trained by a known classification of data items, ie, a training set. A classification function or model, which can be expressed as an IF-THEN rule, a decision tree or an artificial neural network, etc.; then, test data is used to verify the accuracy of the model, and if a predetermined criterion is reached, the model can be used to predict the category of the unknown data item, If it is not accurate enough, continue the training process.

C. Cluster Analysis

Cluster analysis is to divide the data items in a large amount of data into natural groups according to some characteristics of their own. The purpose is to reduce the distance between data items of the same category as much as possible, and increase the distance between different types of data items. That is, the intra-class similarity is maximized and the inter-class similarity is minimized. Cluster analysis is used to organize similar data items in the data. The difference from the classification analysis is that the classification analysis must be based on a predefined category, ie a training set is required.

D. Predictive Analysis

Predictive analysis is a method of discovering trends and patterns of data items over time. Regression analysis is a typical predictive analysis method, which uses a large amount of known data and takes time as a variable to obtain a linear or nonlinear regression function, so as to obtain the law of data over time. Usually, the prediction is based on classification, and the predicted result takes time to verify, that is, the accuracy of the prediction must be known after a certain period of time.

IV. THREE DATA MINING PROCESS

Data mining is a technology that finds its regularity from a large amount of data by analyzing and processing large amounts of data. There are five stages in the implementation process: target definition, data preparation, data mining, result representation and knowledge absorption.

A. Target Definition

In the goal definition stage, it is necessary to combine the background knowledge of related fields to define clear and accurate data mining objectives, which is equivalent to demand analysis.

B. Data Preparation

The data preparation phase refers to collecting and selecting data from a data source, and processing and converting the data into a form suitable for data mining. Specifically, the data preparation phase can be divided into three steps: data selection, collecting relevant data from a large number of data sources for data mining; data preprocessing, in order to ensure data integrity and consistency, to collect The data is preprocessed to meet the requirements of data mining; the data transformation, after a series of transformations, transforms the data into a specific format suitable for the data mining method, that is, extracts specific features or dimensions from the data.

C. Data Mining

In the data mining stage, data mining methods are used to mine the implicit rules and knowledge in the data. This stage is the key and core of data mining technology. Specifically, first, determine the type of analysis method, such as association rule analysis, clustering, etc.; then, for a specific analysis method, select a suitable algorithm, such as Apriori algorithm in association rule analysis; finally, in the data Run this algorithm to find out the knowledge contained in the data, that is, to mine the data.

D. Results Are Expressed

The result representation stage is to further convert, extract and interpret the results of data mining, that is, the rules and knowledge of discovery, according to the needs of users, and finally express them into a form that the user can understand and accept.

E. Knowledge Absorption

The knowledge absorption stage combines the requirements of specific fields, applies the results obtained by mining to specific areas, and provides decision support for decision makers, thus completing the ultimate goal of data mining.

V. THE APPLICATION RESEARCH OF DATA MINING TECHNOLOGY

The research of this subject proposes a teaching evaluation system based on data mining technology. In this system, teachers, students and school administrators are evaluated in a variety of ways, especially the application of data mining technology, which can better provide decision makers with Decision help.

A. Algorithm Selection

Data mining technology implements many algorithms, such as decision trees, association rules, clustering, log mining, and neural networks. In the research of this subject, according to the characteristics of teaching evaluation and the relevance of data, in the implementation of mining technology, the association rule algorithm and decision tree algorithm are selected. On the one hand, the association rules can establish a network of relationships between the data through the discovered association knowledge, and launch relevant results by detecting a phenomenon. Its advantage is that it can produce clear and useful results can support indirect data mining, and even the processing of variable length data, its calculated consumption is predictable. On the other hand, the decision tree algorithm is a very effective method for classification, which has a good effect on the evaluation of teachers. Because classification, it can clearly understand what aspects of teachers are insufficient in the teaching process, and what is the impact of teaching. These two algorithms are very effective methods for network-based teaching evaluation systems. This paper only analyzes and discusses the application of decision tree technology.

B. Application of Decision Tree Technology

The decision tree literally means a kind of mining algorithm that reflects the data mining results in a tree structure and gives classification rules according to the branches. It is one of the core algorithms of data mining technology. It finds some potential, decision-making information by purposely categorizing large amounts of data, often used to predict models. There are many algorithms for decision trees, such as the ID3 algorithm proposed by Quinlan in 1986, the CART algorithm proposed by Leo-Breiman et al., and the SLIQ algorithm proposed by Melhaetal. Among many algorithms, the classification mining algorithm of this system uses C4. 5 algorithm, used to construct the decision tree to generate a classification model, in order to make predictions for teachers' basic information data such as teacher quality, teacher responsibilities, teaching ability, etc., for decision makers to use, so that people can do their best to achieve the enthusiasm of teachers and improve The purpose of teaching quality.

VI. APPLICATION OF DATA MINING TECHNOLOGY IN COLLEGE TEACHING MANAGEMENT

A. Teacher Teaching Quality Assessment

The evaluation of teachers' teaching quality is an important aspect to evaluate the teaching effect of colleges and universities, and also an important basis for the evaluation of teachers' titles. It is a major component of the teaching management of colleges and universities. The current assessment methods for teaching quality are mainly based on statistics and simple calculations, that is, statistics, collecting students' evaluations and achievements, and then weighting the teacher's scores as an indicator of their teaching quality. However, the scientific and authoritative nature of this method is not strong, and the large amount of data collected during the teaching process has not been fully explored. A more

scientific method is to fully mine all aspects of data through the method of association rules, and obtain valuable information as the evaluation basis for teachers' teaching quality. Association rule analysis: Association rule analysis is to dig out the interdependence between data items from the data, that is, the form is "X→Y, support = s%, confidence = c%".

Confidence c% of the association rule X→Y: In the whole event set D, c% of the events satisfying X also satisfy the event Y. Confidence indicates the strength of the X→Y association, denoted as confidence (X→Y), and the minimum confidence is denoted as minConf, which is generally given by the user. Supporting degree of association rule X→Y s%: In the whole event set D, there are s% events satisfying both X and Y. The support degree indicates the frequency of X→Y association, which is recorded as Support (X), and the minimum support degree is recorded as minSup, which is generally given by the user. Frequent itemsets: The support level X of the item set X is not less than the minimum support degree minSup given by the user, and X is called the frequent item set. Association rule analysis is generally divided into two steps: (1) Find all frequent itemsets that exist in the overall event set (database). (2) Generate association rules using frequent itemsets. For each frequent item set A, if $B \in A$, $B \neq \phi$, and Support (A) / Support (B) $\geqslant$ minConf, there is an association rule B => (A-B). In the two steps, the second step is relatively simple. The effect and performance of the association rule analysis are mainly determined by the first step. The Apriori algorithm is a relatively classic association rule mining algorithm.

B. Apriori Algorithm

The Apriori algorithm [4] uses a layer-by-layer search, iterative approach, essentially width-first traversal, that is, frequent (k+1)-item sets are obtained from frequent k-item sets. The execution process of the Apriori algorithm is as follows:

In the first step, the frequent 1-item set is first mined as the starting point for the iteration. Then, the iterative method is used to mine the frequent k-item set (k>1). After mining the candidate frequent k-item set (C k), the minimum confidence degree minsup is used to filter and obtain frequent k-terms. set. Finally merge all of the frequent k-term sets (k>0).

In the second step, all the association rules (candidate association rules) are first mined from all the frequent itemsets, and then the frequent association rules are obtained according to minConf, that is, the association rules whose confidence is greater than the minimum confidence minConf.

C. Application of Association Rules Analysis in Teaching Quality Assessment

First, in the data preparation stage, 500 records of the teacher's teaching evaluation are obtained from the database of a university teaching management system, and 6 attributes are selected and extracted: teacher number, gender, education, title, teaching age, evaluation score, And convert these indicators into binary numbers for program calculations, such as converting professors, associate professors, lecturers, etc. in the title into 00, 01, 11, and so on.

Then, using the association rule analysis method in the data mining described above, and using the evaluation score of 90 points or more as the judgment criterion and the search target, that is, as the threshold for determining the high teaching quality, it may be referred to as "excellent". After searching, the final result is 143 records, the minimum confidence level minSup=15%, the minimum support degree minConf =10%, and the obtained association rules are as shown in the following figure, for example, the association rule "Professor → Excellent, Confidence = 82.5%, support Degree = 21%" means that the number of teachers whose teacher's professional title is a professor and whose score is greater than or equal to 90 points accounts for 21% of the total number of students. At the same time, among all professors, the number of people whose score is greater than or equal to 90 points, accounting for 82.5 of the total number of professors. %. Finally, the results of the experiment were evaluated. It can be seen from the above results that the popularity of male teachers and female teachers is basically the same among students; the higher the education of teachers, the better the teaching effect, indicating that the basic skills of teachers with higher education are more solid and the level of scientific research is higher; The quality of the teaching of long old teachers is also high; in addition, it can be seen from the degree of support that there are more highly educated teachers and professors in the university, and there is a certain level of education.

D. Assessment of Student Achievement

In the teaching management of colleges and universities, the students' academic performance is also an important indicator to evaluate the level of running schools and the quality of teaching. Most of the existing evaluation methods are artificial and simple calculation methods. This method is difficult to comprehensively and comprehensively analyze the performance data. Data mining methods, such as decision tree-based classification methods, can be used to mine useful information from the performance data to provide effective decision support for school administrators to improve the level of running a university.

VII. CONCLUSION

As an important part of education informatization, a large amount of teaching information is collected in the teaching management system of colleges and universities, but most of them have not been well explored and studied. Therefore, the application of data mining technology in college teaching management system has practical significance. This paper applies the association rule analysis and decision tree method in data mining technology to evaluate and mine the teacher's teaching quality and student's academic achievement data in college teaching management system, find some valuable rules, and teach for colleges and universities. Management provides decision support and has made new explorations in the teaching reform and information construction of colleges and universities.

REFERENCES

[1] Huang Jiejun, Pan Heping, Wan Youchuan. Application research of data mining technology [J] . Computer Engineering and Applications, 2003, (2) : 45- 47.

[2] Jia Caiyan, Ni Xianjun. A Review of Research on Association Rules Mining [J] . Computer Science, 2003, 30(4) : 145- 148.

[3] Ma Tinghuai, Zhang Haisheng, Zeng Zhenshou. Mining of Association Rules with Conclusion Domain[J] . Computer Engineering, 2003, 29(5) : 16- 17.

[4] Zhang Hongyun, Liu Xiangdong, Duan Xiaodong, et al. Comparative study of clustering algorithms in data mining[J] . Computer Science, 2003, 30(9) : 5- 6.

[5] Ma Guangzhi, Long Shuozhu. Self-learning system model based on clustering and classification[J] . Computer Engineering and Applications, 2003, 10: 83- 84.