Research on the Application of Data Mining Technology in Educational Management System

Ma Jinglin
Administrative Department, Jiangsu Open University, 399 North Jiangdong Road, Nanjing,210036, China

Abstract
Data mining technology has changed the way people understand data and analyze data to a large extent. The technology belongs to a new kind of information processing technology, which can provide information support to the decision makers and make for promotion of integrated development of educational reform. At present, the educational management system in colleges and universities mainly centers on the handling of daily routines, and the mass data merely realized the storage instead of exerting its function, besides, the significance hidden in the mass data often has not been discovered. Therefore, starting from the definition of data mining, the paper analyzed the necessity of application of data mining technology in the educational management system, and then studied the application of data mining technology in the educational management system.
Keywords: Data Mining; Colleges and Universities; Educational Management System; Application

1.Background and Definition of Data Mining

With the increasingly widespread application of educational management system in colleges and universities, the information quantity of historical data becomes more and more massive. So we can take full advantage of these data information and extract useful information from the massive amounts of information by the data mining technology, so as to discover the hidden laws, thus better utilizing these laws to direct the teaching work. The objective of application of data mining technology in the educational management system is to discover the laws through collection of external information and analysis of historical data in the educational management system, so as to realize prediction on the trend of future teaching affairs management and formulation of educational management strategies. Therefore, how to use informationalized means to offer decision support services for teaching management personnel has become a problem that urgently needs to be addressed. We will focus on the application of data mining technology in the educational management system below.

Data mining technology can be traced back to the study of artificial intelligence, and began to develop gradually in the late 1980s. In the 90s, researches on data mining and knowledge discovery were carried out in a large scale, thus generating a good deal of new methods and tools for data mining. The technology refers to the nontrivial process of automatically extracting the useful information hidden in the data from the data set. It assists the decision makers to analyze historical data and current data by summary of modes of the inquired contents and search of inherent laws, as well as discovers the hidden relations and patterns to further predict the possible future behaviors, so as to provide effective supports for the decision-making behaviors.

2. Necessity of Application of Data Mining Technology in the Educational Management System

With the scale expanding of higher education and the constantly deepening of teaching reform in colleges and universities, the teaching management gradually adopts the credit system as the basis for construction and operation, and puts more emphasis on individuation management specific to individual students. Every student may have a different learning progress, learning outcome or graduation year, accordingly, the organizational form of teaching is relatively more flexible and diverse. The methods and means of traditional teaching affair administration are far from suitable for the requirements of the new circumstances, so we must have advanced, efficient supporting educational system to ensure the smooth progress of the educational administration work. By using a variety of data mining tools, not only the information communication between the leadership and relevant business management personnel is promoted, but also the workload of the information personnel is reduced, so that the information personnel can free themselves from daily data statistics. Hence, the relevant business personnel and leadership can timely, conveniently carry out check and analysis of complex data, so as to master the current situation of teaching. Meanwhile, with the application of database management system in the educational administration work, more and more data has been accumulated in the database. Large amount of significant information is hidden behind these data, so a new data analysis technique is needed to extract the desired information quickly and accurately from the vast resources of educational administration, as well as to extract valuable potential knowledge. The method of data mining can be adopted to carry out pattern extraction on the realistic problems of teaching, so as to improve the data analytical ability.

3. Application of Data Mining Technology in Educational Management System

3.1 Architecture of educational data warehouse system

The data warehouse is the collection of subject-oriented, integrated, stable data of different time, which is used to support the decision-making process in operating management. Before the construction of the data warehouse, it's a vital work to determine the architecture of the data warehouse system. Firstly, we should draw up a system framework of the data warehouse, which is suitable for the characteristics of the institution. The establishment of a truly practical, effective and flexible architecture of the data warehouse system is extremely essential, and the establishment of data warehouse system is to meet the increasing, more and more complicated statistics, analysis and forecast demands raised by managers at all levels. On the basis of maintaining the function of the existing application system, the system makes full use of the various data origins in the school, including faculty and staff data, teaching data, enrollment data, student management information, etc., carries out decision-oriented data reconstitution on them, and then transfers them into the data warehouse. With such data warehouse, the data mining application system and flexible inquiring system will be established facing the ultimate users based on decision support. Through this data warehouse system, the users can conduct deep statistics and analysis on teaching, curriculum provision and situation of student management from different perspectives, in order to direct their decision-making. The key technologies involved in the data warehouse are as follows: data extraction, data storage and management, data representation and data warehouse design. Generally speaking, the architecture of educational data warehouse mainly consists of two parts, i.e., the foreground and the background, which specifically includes data sources, data preprocessing, metadata, data mart, data warehouse, front-end application, etc.. The architecture of the educational data warehouse system is as shown in Figure 1.
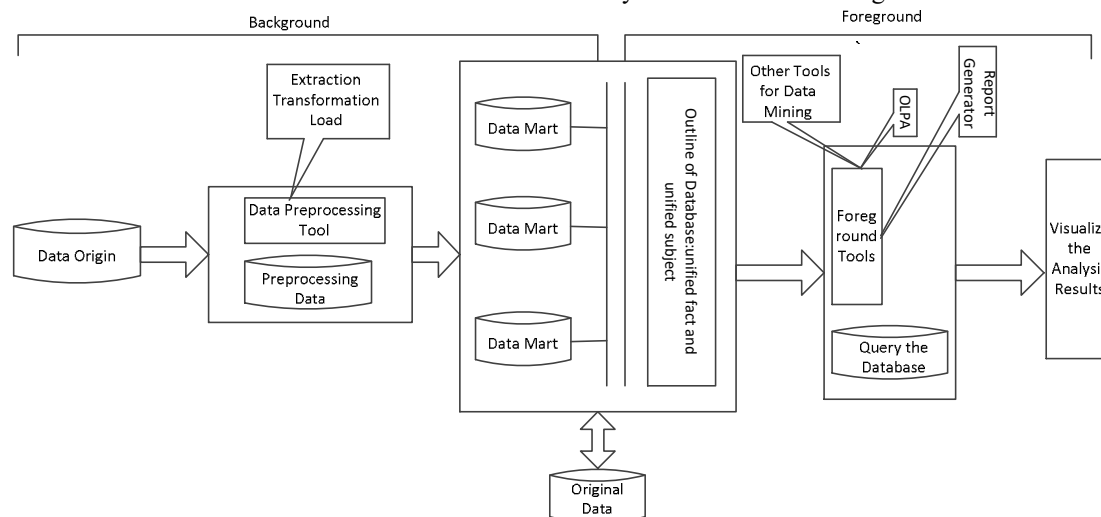

Figure 1: Architecture of the Educational Data Warehouse System

3.2 Association rules mining in the educational administration system

Among the large amount of data mining research techniques, the most popular is the association rules mining. Association rules mainly refer to discovering the inner link between the objects and the basic idea is to determine the minimum support degree, find out all the frequent item sets (referred to as frequency set) larger than the minimum support degree, and then use the discovered frequency set to generate the desired rules. The core of association rules mining is recursion of the frequency set. The main algorithms of association rules mining contain time sequence discovery, sequence pattern discovery, association discovery and so on.

3.2.1 Association rules technology

Data mining can be viewed as a way to discover knowledge in large database. As a new technology, data mining emerges following the development of database technology and it allows users to access or deal with a great deal of information. The purpose of data mining is to generate smart tools for evaluating and extracting information from the database. In recent years, data mining has caused great attentions of the education administrator and the administrative staff play an important role in higher education. Data mining involves a wide range of disciplines and fields and has many commonly used pattern discovery techniques, such as the decision-making tree and clustering technology, which belong to achievements of the field of machine learning and research. The emergence of association rules mining technology has greatly expanded the research of data mining and exerted huge impact on the field of data mining, thus the technology has become a vital branch of data mining research. Association rules can effectively discover the important association relations between data, besides, the expression form of the rules is relatively simple, which is easy to be understood and explained, resulting in the fact that association rules mining from large database has become a popularity in the field of data mining research in recent years.

3.2.2 Apriori algorithm of association rules

Apriori algorithm is one of algorithms of association rules. The first traversal of Apriori algorithm only calculates the number of specific index of each item to determine the large one item set. The subsequent traversal contains two stages. Firstly, candidate item set Ck is generated using the large item set Lk-1 discovered in the (k-1) traversal and the Apriori-gen function. Then the database is scanned to calculate the support degree of the candidate item set in CK, thus getting the order k large item set LK, whose support degree in Ck is no less than the minimum support degree. The above steps will be repeated until the large item set of some order becomes empty, then the algorithm stops.

Description of Apriori algorithm is as follows：

1)1L ={large one item set}；

2) for(k =2; =Lk-1≠ϕ; k ++) do begin

3)Ck=apriori-gen(L k-1);//construct candidate item set

4) for all affairs t =D do begin

5)Ct =subset(Ck , t); //search candidate item set contained in affair t

6) for candidate item sets t=Ct do begin

7) c.count++; // calculate support degree

8) end

9)L k={c =Ck|c.count ≥min sup}；//obtain order k large item set

10) end

11) answer=U k L K

So far, there are a lot of algorithms of association rules, but all of the mining algorithms can be divided into the following steps regardless of the data structure, complexity or efficiency: firstly, carry out preprocessing on data related to the mining task. According to the specific requirements, carry out corresponding operation on the database, so as to form a normalized achievement database; secondly, for the achievement database, obtain all the item sets that meet the minimum support degree, i.e., the large item set. Since under normal circumstances, the database we are faced with is relatively large, this step is the core of the algorithm; thirdly, generate rules that meet the minimum confidence and form the rule set; fourthly, explain and output.

Nevertheless, the second step is the focus and difficulty of mining association rules, so most of the researches focused on this step.

3.3 Application of association rules in reasonable curriculum setting and learning evaluation

Association rules are used to describe the relations between data items in the database, namely, digging out the hidden association or relations between the data according to the appearance of some of the items of an affair. The students' curriculum study is progressive, and certain correlation and sequential order relation exist between the curriculum. We ought to know which courses should be taken before learning a more advanced course, because if the prerequisite course was not well studied, it's bound to affect the study of follow-up courses. The major setup and course system of colleges and universities should be synchronized with the social needs, which requires the schools to revise the teaching program in a timely manner and replace the curriculum unfit for the market requirements. In this way, we can take advantage of the data mining and adopt the correlation analysis, mainly the association rules to analyze the reasonability of curriculum setting. Carrying out evaluation on students is a link of the educational process, therefore, the function of evaluation is consistent with the educational objectives. Highlighting the developmental function of evaluation intensively reflects the educational idea of "all for the development of students". The students are undergoing a process of continuous development and changes, so the significance of education is to guide and promote the development and improvement of students. The development of students needs targets, guidance and incentives. Developmental evaluation determines individualized developmental goals for students. In nowadays, network of teaching and scientific research is universal, so we can use the data mining tools to analyze and process the achievement database and behavior record database to determine the strengths and weaknesses of the students and put forward specific, targeted suggestions on this basis. The topic of association rules is important in the field of data mining. It seeks the frequent patterns, correlation, dependency or causal structure between the item sets or between the object sets in the transaction data, relational data or other information carriers. Through the analysis of relations between data or records, it decides which things will happen together. It's one of the principal modes for current data mining research, it focuses on determining the relations between different fields of the data and finding out the dependencies between multiple domains that satisfy the given support degree and confidence level threshold. The basic model of association rules mining is shown in Figure 2：
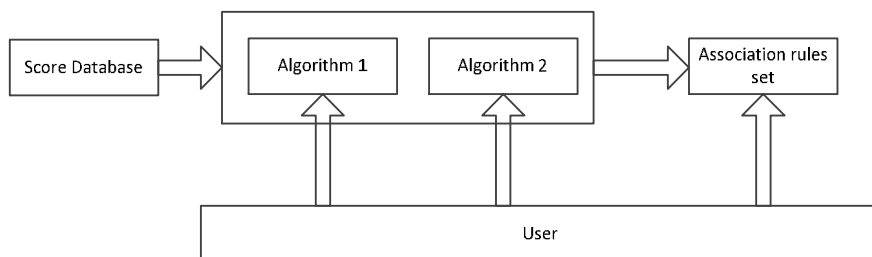
Figure 2 Basic Model of Association Rules Mining

By using different data mining methods in different modules of the educational management system, we can discover the relevant valuable information and knowledge. For example, in the topic of exam results analysis, we would like to analyze the incidence relations between the student's prerequisite course achievements, gender, assignments or classroom performance and the final test achievements, as well as the sequential relations between the courses in the teaching program. Using association rules mining in the student evaluation module can associate the test achievements of the students with the individual information (e.g.,gender, nationality, age, origin of student, etc..), thus applying the discovered rules into decision-making and providing valuable reference for quality-oriented education and teaching modes. Application example of the association rules is as follows:

Assuming that I is a set of m courses,Database D is given as the set of achievement records of the students. Every Record T is a set of attributes in I. T has an exclusive identifier TID. Taking the achievement Record Table 1 of a group of students as an example, Table 2 is further analyzed by association rules.

Table 1 Partial Fields of Student achievement Table

Field Code	Filed Name	Type
ID	Student ID	CHAR（5）
ZY01	Course 1	CHAR（1）
ZY02	Course 2	CHAR（1）
GG01	Course 3	CHAR（1）
GG02	Course 4	CHAR（1）
GG03	Course 5	CHAR（1）

Table 2 Factual Data Table

ID	GG01	GG02	GG03	ZY01	ZY02
001	A	B	A	A	A
002	A	A	A	A	B
003	B	C	B	B	B
004	A	B	A	B	A

According to the above idea of carrying out improvement on the algorithm, the structure of candidate item set contains a list of transaction identifiers, and the length of the list is indeterminate, so the implementation is difficult in the relational database. Therefore, we did the following adjustments in the implementation of the algorithm: in the candidate item set table, each of the <item set, transaction identifier> is regarded as a record, in this way, one item set in the candidate item set can correspond to multiple records. Taking discovery of the association rules for student curriculum for example, part of the data fields of the course achievement table of students major in computer engineering is listed in Table 1, and some data of the table is listed in Table 2. However, the candidate item set obtained by processing data in Table 2 is listed in Table 3.

Table 3 List of Candidate Item Set

Item	Value	ID	Item	Value	ID
GG01	A	001	GG03	B	003
GG01	A	002	GG03	A	004
GG01	B	003	ZY01	A	001
GG01	A	004	ZY01	A	002
GG02	B	001	ZY01	B	003
GG02	A	002	ZY01	B	004
GG02	C	003	ZY02	A	001
GG02	B	004	ZY02	B	002
GG03	A	001	ZY02	B	003
GG03	A	002	ZY02	A	004

Then we carry out statistics on the count of each item set through the candidate item set table and insert the count into frequent item set table. Meanwhile, the infrequent item sets in the candidate set are deleted to facilitate

the formation of the candidate item set of the next level. Next, we use the educational management system-course relevant subsystem to generate frequent item set according to the achievement data. Thus it can be seen that the data warehouse in colleges and universities is a complex and huge project, which involves all kinds of institutions and personnel in the schools, thus needs plenty of investigations and data collections and sorting as well as a longer period for establishment and improvement. However, by building a data warehouse system, colleges and universities can give full play to the roles of all kinds of resources in the schools, so as to enhance the school management and decision-making level.

4. Conclusions

In a word, the data mining technology is a useful tool, which has been applied in multiple fields and obtained satisfactory results. In the field of higher education management, with the ever-increasing data information, the application of DM technology into the educational management system can promote the further reform, improvement and development of the education management. Of course, data mining itself is not almighty and its application can not be separated from the practical background. Only by following the people-oriented principle, will the technology have real vitality. Anyway, the educational management system contains a large amount of useful information to be discovered. In nowadays, Rejuvenate China through Science and Education is vigorously promoted, so data mining technology will definitely have important guiding significance for schools to develop better student training programs.

References

Qiu Yue. Application of Data Mining Method in the Evaluation on Student Achievements [J]. Science & Technology Information, 2010, 12 (9): 76-77

Huo Liang, Yang Liu, Zhang Junzhi. Exploration of Data Mining and Its Development Direction[J]. Estate and Science Tribune, 2011 (20): 49-50

Lin Li. Application of Data Mining Technology in the Educational Management System of Colleges and Universities [J]. Computing and Information Technology, 2012, (4).

Song Xiaomin, Zhang Guofang, Xing Shulan. Application of Data Mining Technology in Curriculum Correlation Analysis [J]. China Out of School Education, 2012, (10): 22-23

Wu Chunqiong. Tentative Analysis of Application of Data Mining Technology in Credit Teaching Administration [N]. Journal of Changchun Education Institute, 2013, 29 (09): 115-116.

Z. LIU. A Proposal of High Performance Data Mining System, Lecture Notes in Computer Science,2003: 34-36

C. RYGIRLSKI, J. C. WANG, David C.Yen. Data Mining Techniques for Customer Relationship Management. Technology in Society, 2002, 24: 483-502.

S. Y. HUNG, D. C. YEN, H.Y. WANG. Applying Data Mining to Telecom Chum Management.Expert Systems with Applications, 2005, 9: 1-10.

S. POMERANTZ, A. WHITE. Re-modeling ILS acquisitions data to financially transition from print to digital formats. Library Collections, Acquisitions, and Technical Services, 2012, 33(1): 42-49.

Author introduction：

Ma Jinglin（1991.09.19-），Female, research direction: educational research, information system, Master of Science in Management ，Research associate ，Jiangsu Open University

Mailing address: 399 North Jiangdong Road, Jiangsu Open University, Nanjing, JiangSu，China

Zip code: 210036