ATLANTIS
PRESS

Research on Student Achievement Early Warning Association Rules Mining Technology

Shaoyun Song[1, a *]

[1] School of Mathematics and Information Technology, Yuxi Normal University, Yunnan, China

[a]ssy@yxnu.net

Keywords: Association rules; Data mining; Achievements; Early warning

Abstract. An early warning model of student achievement is established by association rule mining, using the SQL Server association rule mining algorithm. Analyze the association rules of student's first courses exams. Implementation of student achievement warning model. Early warning model is used to find the association rules between courses so as to warn students of their academic achievements. Many years of practice show that this method plays an important role in improving the quality of teaching.

Introduction

Due to the extensive application of the information management system in the educational administration of colleges and universities, the accumulation of a large number of student achievement data, which implied the knowledge and the law of student achievement achievement, discovered the hidden knowledge and the law through the association rules of the achievement data so that Improve the quality of teaching and early warning of student achievement. There are many student achievement early warning systems, but most of them are written in MATLAB and JAVA. The system based on WEB is rare. This paper studies the student achievement mining system based on WEB technology, teachers and students can Internet access to the system, to achieve online student achievement warning.

Basic Concepts of Association Rules

The Basic Concept. Let $I = \{I_1, I_2, \cdots, I_n\}$ be a collection of n different items, the transaction database is $D = \{T_1, T_2, \cdots, T_m\}$, where $T_i = \{I_{i1}, I_{i2}, \cdots, I_{ik}\}$, and $I_{ij} \in I$, the association rule is the implication of the form $X, Y \subset I$, where E, and $X \bigcap Y = \Phi$.

Support Numbers and Support. The total number of transactions in the support project set $X \bigcup Y$ in database D is called the number of support for association rule $X \Rightarrow Y$, denoted as $Count(X \Rightarrow Y)$. The support for rule $X \Rightarrow Y$ is $Sup(X \Rightarrow Y) = P(X \bigcup Y) = Count(X \Rightarrow Y) / |D|$.

Confidence.

$$Conf(X \Rightarrow Y) = P(X \mid Y) = \frac{Sup(X \bigcup Y)}{Sup(X)}$$

The confidence of rule $X \Rightarrow Y$ is defined as .

Item Set. An item set containing k items is called a k-item collection.

Strong Association Rules. Colleagues meet the rules of minimum support and minimum confidence.

Mining of association rules is from the frequent focus found at the same time to meet the user given the minimum support and minimum confidence of the implication, that is, strong association rules. Student achievement online early warning system mainly in the server-side application SQL Server table between the correlation operation, to find out the requirements of the association rules, and then according to the association rules to analyze the knowledge contained among the students. The association rule model is mainly divided into Boolean and numerical. Numerical association rules can be combined with multidimensional association or multi-level association rules to deal with the numerical

field, and then dynamic segmentation, and Boolean-type association rule processing values are discrete, the kind of, it shows the The relationship between.

Model Design

Student Achievement The online early warning system model is based on the SQL Server database. First by the user to determine the model needs the minimum support and confidence, and then model the full application of SQL select function from the database to generate 1-term frequent set of temporary table, and from the table to produce 2-term frequent set from 2- (K ⩾ 3). The association rule mining algorithm (Apriori algorithm) is used to calculate and output the association rule information, and the rule information is stored in the rule base, and the early warning information of the student learning association course is generated based on the association rule base. System users are divided into four levels: students, teachers, educational offices and administrators, different users have different permissions. The system based on the B / S structure design early warning model shown in Fig. 1.
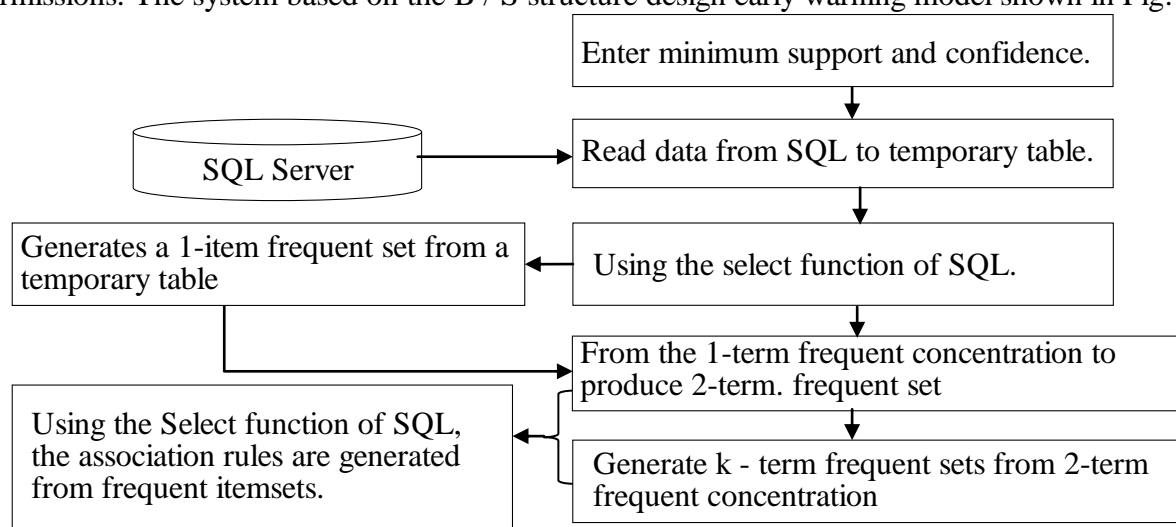

Figure 1.   System model

Model Implementation

System Works. Student achievement online early warning system is based on SQL Server2008 and ASP + VBA development, the full use of SQL Server select statement on the data in the database to find "1-item frequent itemsets" and "2- or more frequent itemsets" And "k-item frequent itemsets" (k> 2). Through the select statement on the database table for complex connection operations and relational operations, to find based on frequent itemsets between the student achievement scores between the association rules.

The system is based on the B / S structure development. The user selects the user type from the Web home page and enters the user name and password of the corresponding user type. The system authenticates the user name and password. If the user name and password are incorrect, the system returns to the home page. If the input is correct, Of the user interface, different users have different permissions, as shown in Fig. 2. System data mining operations in the server side, the client only need IE browser to complete the data mining, so the system is easy to maintain and upgrade, its working principle shown in Fig. 2.

User Classification. Student Achievement The online early warning system divides users into: student users, teachers, academics, and administrator users. Student users have frequent itemsets mining and association rules mining permissions, so that students keep abreast of the relevance of their own course achievement. Teachers can view the import of student achievement, data processing (discretization), frequent itemsets mining and association rules mining, teachers can according to their own needs to partition the data discretization or by discrepancy, for example, the student scores for

discrete (80-59), [90-69], [90-69], [90-79], [90-89], [90-100], can also be discretized into grades A, B, C, D and E Respectively, corresponding to excellent, good, medium, poor, pass, fail. The user of the Office has all the privileges of the system. The post office user has the functions of importing mining data, data processing discretization, frequent itemsets mining and deleting mining data sources. The administrator user is responsible for the management of the system backstage, including the management of the database backup, the appearance of the home page, and the assignment of user rights.
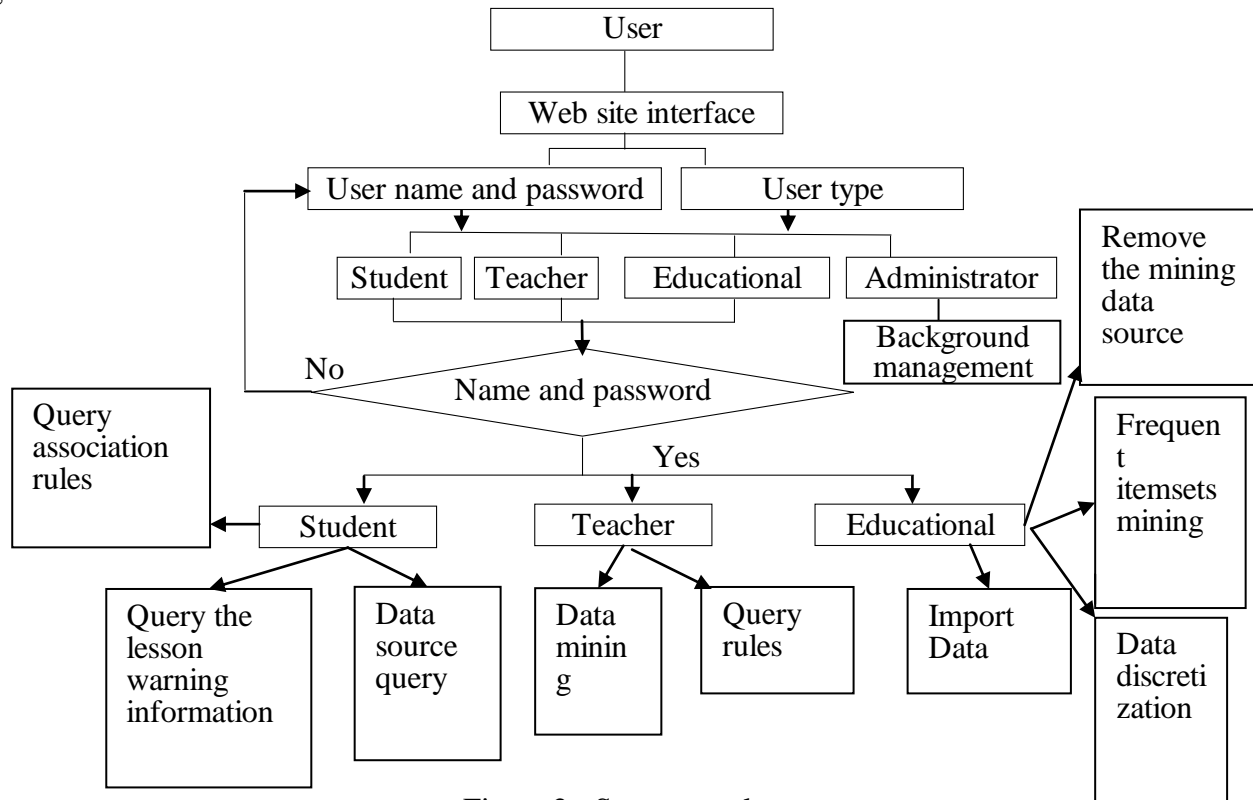
Figure 2. System works

Data Processing

The main function of data processing is based on user needs to partition or sub-level discretization, in order to carry out association rules mining, the specific function is as follows.

Delete the Mining Data Source. the use of the system directly mining association rules, the use of the original database data, if the user needs to use the new data mining, you need to delete the data source, delete the data source before importing new data.

Import Data. Can be prepared from the prepared Excel table, the importation of Excel table data field order can not be the same as the database table field order, but requires the same field name, the system can automatically import Excel table data into SQL Server database in the table.

Data Source Table Query. After the user has finished importing the data, you can use this function to view the imported raw data in order to carry on the knowledge analysis to the mining association rule, get the result that needs it, and the mining association rule Analyze its accuracy and accuracy. The data source table is shown in Fig. 3.

Table 1 Source data table

ID	Name	Math-1	Math-2	Data structure	C language	computer network	Router and switch
1	A	70	65	83	60	91	88
2	B	85	80	77	90	78	75
3	C	92	90	87	95	85	82
4	D	80	77	68	92	82	79

Data Discretization. The function of data discretization is to discretize the original data because the original data is not convenient for mining the association rules. The original data needs to be partitioned or classified into different levels. The partitioning can divide the original data into several regions. For example, Divided into four districts: [0-59], [60-79], [80-89], [90-100], can also be classified, such as A, B, C, D, E, respectively, the corresponding results "" "Good", "good", "medium", "pass", "fail", the way the user is determined by the administrator, but also by the administrator to other teachers to deal with the system default mode is partition. As shown in Fig. 4.

Table 2 Discrete data processing

Math-1	Math-2	Data structure	C language	computer network	Router and switch
70-79	60-69	80-89	60-69	90-100	80-89
80-89	80-89	70-797	90-100	70-79	70-79
90-100	90-100	80-89	90-100	80-89	80-89
80	70-79	60-69	90-100	80-89	70-79

Frequent Item Sets Mining. The user can quickly use this feature to dig out the "1-term frequent set", "2-term frequent set" and "k-item frequent set" (k> 2), the use of frequent itemsets Simply get the percentage of all students from the student achievement table that satisfies the user's support.

The key step of frequent itemsets mining is association rule mining, and its mining quality directly affects the quality of association rules mining. The system uses SQL Server's "Select ... from ... group by ... having ..." statement for frequent itemsets mining, because the statement is a database-specific statement, the calculation of the data fast, easy access to frequent itemsets. For example, the "1-item frequent set" mining statement is as follows.

rs.open "select ["&selefd(n)&"],count(*) as Number of support from studentdm2 group by ["&selefd(n)&"] having count(*)>"&sup&"*(select count(*) from studentdm2)",conn,1,1

Where rs is the record object, studentdm2 is the database table name, the statement using the sub-table nested source table calculation method, you can quickly meet the user to set the minimum support (expressed by the number of support) "1-term frequent items Set "as shown in Figure 5.

Table 3 Results of frequent itemsets with 70% support

Math-1	Math-2	Number of support
70-79	0-59	7896
70-79	60-69	8874
70-79	70-79	4564
80-89	70-79	8965
90-100	80-89	6752
90-100	90-100	5239

(6) Association rules mining: the system automatically from the frequent items to dig out to meet the user set the minimum confidence and minimum support of the association rules. From the association rules, users can analyze the knowledge contained in the actual needs.

The following mining "high number 1, high number 2 => discrete mathematics" association rules. The steps are as follows:

Step 1: Enter the minimum support and minimum confidence (0.05 and 40%) for the access interface entered as any user, as shown in Fig.6. Fig. 6 3-item association rule mining parameter setting

In Fig. 6, click on "association rule mining" to get the mining result of the association rule of Fig. 3, as shown in Fig.
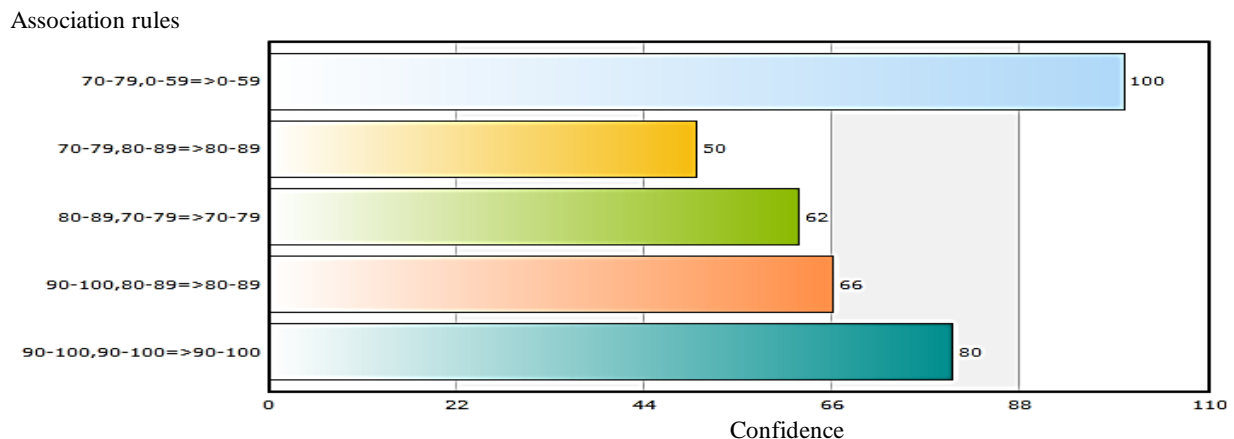
Association rules

Figure 3. Mining results of association rules with support of 75% and confidence of 80%

Fig. 3 The association rule mining results with support degree of 75% and confidence of 90%

From Fig. 7, we can get the knowledge that the scores of "high number 1" and "high number 2" have a great influence on "data structure". The scores of "high number 1" and "high number 2" are between 70-79 and 0-59 between the students, their "data structure" score between 0-59 100% confidence, you can use the knowledge. "High number of 1" and "high number 2" between the 90-100 students, "data structure" in the 90-100 confidence between the degree of 80%, the knowledge can also be used. Teachers need to pay much attention to the teaching of high and high numbers of two courses. Students should continue to strengthen the basic knowledge of high numbers 1 and 2, and lay a solid foundation for better learning of data structure courses.

Summary

In the minimum support degree of 0.15,0.20 and 0.35, respectively, the minimum confidence were 70%, 65%, 40%, the student performance online early warning system experiments, record the number of output rules, the number of rules to collect and deal with the These data are analyzed statistically. The accuracy of the rules under different conditions is shown in Table 4-1. The warning accuracy is P (r), and the set I = R ∩ D is denoted as m. Then P (r) = m / n * 100%. The accuracy of the analysis, from the table can be found: the minimum support and the reliability of the two thresholds set the smaller, the more the number of rules associated with the number of rules. However, there is no such inverse relationship between the early warning accuracy and the minimum support, and the minimum confidence value. It can be seen from the table that when the minimum support is 0.15 and the minimum confidence is 65%, the most accurate, so the next to the back of the data table are set to 65% and 0.15 of the two So that the association rules are relatively accurate.

Table 4 Course Association Rules and Accuracy

Support	Confidence = 70%		Confidence = 65%		Confidence = 40%	
	Number of associations	Accuracy(%)	Number of associations	Accuracy(%)	Number of associations	Accuracy(%)
0.15	3	70	5	82.3	12	72.8
0.20	5	60.52	8	78.6	18	62.4
0.35	7	55.6	12	72.8	36	55.9

The system transplanted to the campus network server, open for the Yuxi Normal College Digital Institute of Computer Science and Technology and professional information management professional use, through nearly four years of application, the structure shows that the quality of teaching to have a greater help, Gradually extended to the use of the school.

References

[1] J. Han, M. Kamber, Data Mining: Concepts and Techniques, 2nd edition, Singapore: Elsevier, 2015.

[2] Z.-H. Zhou Three perspectives of data mining Artificial Intelligence, 2016, 143 (1): 139-146.

[3] P.-N. Tan, M. Steinbach, V. Kumar, Introduction to Data Mining, Reading, MA: Addison-Wesley, 2016.

[4] WU Yong-jun, Chine Microbial Immunology [J], August 2015, Vol23, No.8

[5] AGRAWAL R, IM IEL ISKI T , SWAMI A. Mining ass ociat ion ru les betw een Set s of it ems in large database[C] MPr oc of ACMS IGMOD Int l Conf on Management of Dat a(SIGM OD. 93) , 1993: 207-216.

[6] H OUT SMA M, SWAMI A. Set-Orien ted Min ing f or As sociati on Rules in Relati on al Dat abases[C] MProceedings of the 11th IEEEInt ernational Conf erence on Dat a Engin eering, Taiw an : Taipei, 1995: 25-34.