The Application of Association Rules Mining in the Analysis of Students' Test Scores

Jigang Zheng[1, a*] and Jingmei Zhang[2, b]

[1]Department of Mathmatic, Baoshan College, Baoshan, Yunnan, 678000, China

[2]Library of Baoshan College, Baoshan, Yunnan, 678000, China

[a]6913641@qq.com, [b]279619568@qq.com

* The corresponding author

Keywords: Data mining; Association rules; Test scores; Teaching quality

Abstract. Data mining from large amounts of data to extract efficient, implicit, potential useful knowledge and information, it has in many other areas has been successfully applied. However, the application in the field of education is relatively less. The central work of institutions of higher learning is teaching, the focus is to improve the quality of education, but students are precisely the basis for measuring the quality of teaching. The characteristic of this paper is to apply association rule mining method of university students' grades, previously unknown effects on student achievement factors of mining, to provide some valuable reference for teachers and administrators, to provide the necessary decision support for teaching and student management, to better carry out teaching, in order to improve the quality of teaching in Colleges and universities.

Introduction

As a college, its primary task is to train high-level talents through teaching and education, so the teaching work is always the center of the school. The key and the key of higher education is to improve the quality of education, and students' performance is an important basis to assess the quality of teaching, and it is important to analyze the data of students' performance [1]. Student achievement is an important basis for evaluating the quality of teaching, and the important sign of whether students master the knowledge they have learned. At the same time, efforts to improve the students' academic achievement is the goal of every university. In the school teaching work, the examination is one of the important part, pass the exam, teachers can understand the student's learning, and like to take this opportunity to know teaching effect, improve teaching methods, improve teaching quality, students can understand the pass the exam his mastery of the content, purpose for learning. The results of information mining and utilization in the quality management of the modern university examination basic links and routine work, timely and effective examination analysis and information feedback to improve and perfect the examination and improve the examination quality and the quality of teaching is of great significance and role.

Apriori Algorithm

Apriori algorithm is a classical algorithm of association rule mining, first, find all frequent itemsets, and then generate strong association rules from frequent itemsets. The first step, from the transaction database d to find all the support degree is not less than the user specified minimum support threshold) (frequent item sets; the second step, using frequent itemsets generated the desired association rules, the basic principles of association rules is confidence must be not less than the user specified minimum confidence degree threshold. The second step is easy and intuitive, the first step is to find all frequent itemsets is the core of the algorithm, most occupy the entire computation, so sometimes only consider the efficiency of mining frequent itemsets [2].

Definition1 Set $I = \{I_1, I_2, \cdots, I_m\}$ is a collection of data items, D transaction is a collection of all [5], A transaction T has a unique identifier TID. If items, transaction support items claimed T set A, also known as T transaction that contains the item set A.

Definition2 Association rules are shaped like $A \Rightarrow B$ type of implication, among them $A \subset I$, $B \subset I$, and $A \bigcap B = \Phi$.

$A \Rightarrow B$ supports the association rule is defined as:

$$\sup port(A \Rightarrow B) = \frac{\sup port(A \bigcup B)}{N} \times 100\%$$

,credibility is defined as:

$$confidenc\!e(A \Rightarrow B) = \frac{\sup port(A \bigcup B)}{\sup port(A)} \times 100\%$$

.

Definition 3 Support and confidence required to be greater than the threshold set by the user (ie, minimum support threshold and minimum confidence threshold), that:

$\sup port(A \Rightarrow B) \geq \min\_sup$, $confidenc\!e(A \Rightarrow B) \geq \min\_conf$.

Data Preparation

Existing in the teaching of mathematical modeling problems and contradictions have become increasingly prominent, mainly in the mathematical modeling teaching team strength is not strong, the old teaching material, mathematical modeling course opened is not wide enough, the number and strength of participating students needs to be improved.

Table 1 Student information

ID	Gender	Degree of tension	Difficult subject	Learning attitude	Achievement	Class seating
1	Male	No	Hard	Correct	96	2
2	Female	No	Commonly	Very well	87	1
3	Male	Yes	Hard	Not correct	47	8
4	Female	No	Easy	Correct	85	3
5	Female	No	Commonly	Correct	84	7
6	Male	Yes	Commonly	Not correct	47	7
7	Male	Yes	Commonly	Not correct	37	6
8	Female	No	Commonly	Very well	87	1
9	Female	Yes	Commonly	Not correct	26	7
10	Female	Yes	Commonly	Not correct	34	6
11	Female	No	Commonly	Correct	81	3
12	Female	No	Commonly	Very well	88	1

13	Male	No	Commonly	Not correct	64	4
14	Female	Yes	Commonly	Not correct	32	5
15	Male	No	Hard	Correct	15	8
16	Female	No	Hard	Not correct	42	7
17	Male	Yes	Commonly	Correct	70	8
18	Male	Yes	Commonly	Correct	72	4
19	Female	Yes	Commonly	Not correct	32	6
20	Female	Yes	Commonly	Very well	80	1
21	Male	Yes	Commonly	Correct	46	7

Data Conversion

Get the student achievement data is continuous numeric data, and mining in the treatment must be Boolean and discrete data, which requires to the original data of a data set grouping, the discretized and transformed into Boolean type. First put on the table a gender, exam stress degree, subject, learning attitude and achievement, seating in order packet, sex we put it is divided into: A1 male, female A2; tension in the exam is divided into: B1 tension, B2 is not tight; difficulty of the subject is divided into: C1 hard, C2 commonly,C3 easy; learning attitude is: D1 very well,D2: correct,D3: is not correct; scores are divided into: E1 [0 58-59], E2 [60, 69, E3 [70, 79]. E4 [80,89,E5[90100].Seating is divided into the first three rows of F1[1, 2, 3]. The middle three rows F2 [4, 5, 6]; reciprocal two rows of F3 [7, 8].

A record in Table 1, is regarded as a transaction, the field value in the record is converted to the corresponding data item, and the relation is converted to the corresponding transaction database. For example, the first record can be converted to a transaction: {A1, B2, C1, D2, E5, F1}, can be obtained in Table 2, as shown below.

Table 2　Conversion view

ID	Gender	Degree of tension	Difficult subject	Learning attitude	Achievement	Class seating
1	A1	B2	C1	D2	E5	F1
2	A2	B2	C2	D1	E4	F1
3	A1	B1	C1	D3	E1	F3
4	A2	B2	C3	D2	E4	F1
5	A2	B2	C2	D2	E4	F3
6	A1	B1	C2	D3	E1	F3
7	A1	B1	C2	D3	E1	F2
8	A2	B2	C2	D1	E4	F1
9	A2	B1	C2	D3	E1	F3
10	A2	B1	C2	D3	E1	F2

11	A2	B2	C2	D2	E4	F1
12	A2	B2	C2	D1	E4	F1
13	A1	B2	C2	D3	E2	F2
14	A2	B1	C2	D3	E1	F2
15	A1	B2	C1	D2	E1	F3
16	A2	B2	C1	D3	E1	F3
17	A1	B1	C2	D2	E3	F3
18	A1	B1	C2	D2	E3	F2
19	A2	B1	C2	D3	E1	F2
20	A2	B1	C2	D1	E4	F1
21	A1	B1	C2	D2	E1	F3

Mining Association Rules

Minimum confidence set is 10%, minimum support is set for 50%, adopt Apriori algorithm of association rules mining, mining student test scores is more than or equal to 80 of the following rules, as shown in Table 3.

Table 3　Association rules

rule	Degree of tension	Difficult subject	Learning attitude	Class seating	Support	Confidence
1	B1				23.8%	45.3%
2	B2				47.6%	42.1%
3		C1			14.3%	68.6%
4		C2			57.1%	35.7%
5		C3			12.5%	23.5%
6			D1		24.0%	70.8%
7			D2		53.3%	63.3%
8			D3		18.1%	40.8%
9				F1	33.3%	54.1%
10				F2	28.6%	42.5%
11				F3	22.1%	32.6%

In Table 3, can see Rule 2 show that tests the student not nervous let student test scores is more than or equal to 80 degree of support was 47.6%,confidence is 42.1%; Rule 7 show that peacetime learning attitude test scores greater than or equal to 80 percent of the support rate is 53.3%,confidence is 63.3%; 9 rules that usually class seats in the first three rows of the test scores of students is more than or equal to 80 degree of support is 33.3%.confidence is 54.1%; other rules can also do a similar understanding.

Analytical Association Rules

The main factors of the above rules including the examination anxiety, difficulty of the subject, the usual learning attitude and class seating. Usually students learning attitude is very correct, correct, is not correct, classes usually has a great attitude and correct student achievement is outstanding, the teacher should take the method encourages students to learn, improve the students' enthusiasm, to

obtain the progress of students to appropriate rewards and praise, let the student active learning, love of learning. In addition, class seating on student achievement also have certain effect, in the first few rows of excellent student rate is higher, better grades, teachers should pay attention to the seat arrangement, appropriate sub low and low vision students as much as possible arrangements in front and regular of seat of wheel change. Maximum because of the reduced seating to student achievement to bring the negative influence as far as possible. In addition to the examination before the students to give a certain encouragement, to establish the confidence of success, so that they have a correct attitude towards the examination, both to have the pressure and more to have the power to adjust the mentality.

References

[1] Zhang Yanli. Application of data mining technology in the educational administration system of Digital Campus [D].Northeast University, 2005.

[2] Jiawei Han, Micheline Kamber. Data Mining Concepts and Techniques [M]. Beijing: Machinery Industry Press, 2007:146-168.