

# The Correlation Analysis of Curriculum Based On Clementine

Xiaoming Du<sup>1, a</sup>, Xinrong Tao<sup>2, b</sup> and Zhen Yang<sup>1, c</sup>

<sup>1</sup>Jiangsu University of Science and Technology, JiangSu 215600, China;

<sup>2</sup>Zhangjiagang Admissions Office JiangSu 215600, China.

<sup>a</sup>260925671@qq.com; <sup>b</sup>190571686@qq.com; <sup>c</sup>1254343698@qq.com

**Abstract.** To further explore the correlation between courses, university culture revision of the scheme provides the basis of quantitative analysis, through the data mining method for correlation analysis of student course grades, the improved Apriori algorithm is performance data in the transaction database is out now is set to 1, did not appear to be set to 0, forming a similar Boolean or binary data, using a binary "and operation" of the to join and prune. Verified by concrete examples show that the method can save the scan time and storage space, and the cultivation scheme formulation has certain directive significance.

**Keywords:** Data mining; Apriori algorithm; Course correlation.

## 1. Introduction

With the progress of science and technology as well as education method continuously improving, courses of University also are in constant change. In cultivating plan revision process, on the one hand, it is through the recommendations of the experts in related fields to analyze, on the other hand, it use the traditional mean, variance and confidence to analyze, which often cannot be found hidden behind the correlation between different courses. In this paper, through the analysis of students' achievements in the past years, it uses Clementine software to analyze the results of the data, so as to obtain the correlation between these disciplines.

Association rules use the improved Apriori algorithm, Boolean valued and binary conversion of data, and calculate the "and" operation. It can save storage space and accelerate the operation speed of the algorithm, and reduce the times of scanning database. The improved Apriori algorithm to mining can satisfy the need of the relationship between the degree and reliability of different courses, which can provide the teacher teaching with appropriate guidance and provides some suggestions for curriculum setting, so as to improve the quality of teaching. To dig out the relationship between curriculum, and intuitive, reflecting the rules, it can provide support for decision-making.

## 2. Data mining model design

Define 1: item and item sets

Hypothesis,  $I = \{i_1, i_2, \dots, i_m\}$  is a collection,  $i_k (k = 1, 2, 3, \dots, m)$  is data items (basic elements in association rules). Data collection  $I$  is called data sets, When  $I$  contains the number of items for  $k (1 \leq k \leq m)$ , it will be called as itemsets. If the data sets contains the data item, it is called the positive item set, and vice versa.

Definition 2: Transaction and transaction database

Hypothesis,  $I = \{i_1, i_2, \dots, i_m\}$  is a collection of all items.  $T = \{t_1, t_2, \dots, t_i\} t_i \in I$ , Transaction  $T$  (the transaction) is a subset of  $I$  sets, namely  $T_i$ , every transaction has only  $TID$  marking with their corresponding, transaction structure of all transaction database  $D$  and its affairs number recorded as  $|D|$ .

### 3. Data mining model design

#### 3.1 Degree of correlation module design

All subjects reflect the students' learning in each semester, in order to study the relationship between courses to set score as 0 or 1. Namely, that the project appeared in data will be set 1, no appeared set to 0. But this article is the second partition. First of all, it analyzes correlation between excellent student achievements and set the value more than 85 as 1, the value between 85 is 0. Then, through using the Clementine software mining and setting the degrees of confidence and support, Apriority algorithm is used to modeling and export it. At the same time, grades with bad correlation reprocessed, namely the value of more than 60 is set to 1 and that below 60 is set to 0 and then analyzed. Algorithm as shown in Figure 1, when  $k \geq 2$ , it can use  $L_{k-1}$  to look for  $L_k$ , then join and prune.

In the Apriority algorithm, firstly, scan the database once and result in frequent 1-itemsets which is denoted as  $L_1$ . secondly, followed by pruning and connection, the project can generate candidate item sets  $C_2$ . Through repeating the first two steps, the  $(k-1)$  - frequent item set generation  $K$  candidate set  $C_k (k \geq 2)$ . The core of the whole algorithm is Apriority-gen sub process, mainly on the frequent set of  $L_{k-1}$  connection and pruning operations, and then generate candidate item sets  $C_k$ .

### 4. Empirical analysis

#### 4.1 Identify mining targets and objectives

In this paper, it refers the a college of information management and information system of student achievement data as mining object, by mining the students of different grades, such as compulsory courses and professional courses, scores between professional courses and professional courses, in order to explore a good programs to the extent of the impact of other courses, or the degree of mining of good courses influencing other courses. This is to get valuable information and rules, so as to find the reasons for the impact on student performance, and then to ask the teacher to make a reasonable arrangement.

This paper regards academics performance of IMIS in certain school as the research object, whose correlation is studied, including the Math, computer based), C++, data structure, e-commerce, database, as shown in Table 1 for the acquired experimental data.

Table 1 experimental data

xh	computer	VC++	structure	business	SQL Server	Java	(OR)	Math
1245793101	62	26	92	63	86	83	61	61
1245793102	65	36	60	61	78	86	36	36
1245793103	64	25	69	70	92	90	75	75
...	...	...	...	...	...	...	...	...

It will generalize the quantitative attributes of the original data. Then the data can be transformed into discrete attributes, thus simplifying the data quantity.

(1) The results to 85 for the boundaries, to be divided, set the data between 85-100 into 1, and the data of less than 85 into 0 to analyze to prove whether the results of the outstanding students of the course has the relevance.

(2) To divide 0-85 into two parts again, one is data between 0-60 will become 0. On the other hand data between more than 60 will be set into 1, so that you can analyze the correlation between students with poor academic performance. And the correlation between medium achievement and above can be analyzed.

(3) Course code

The achievement of the course names will be named as code in the English, such as "advanced mathematics" coding is Math; "C++" number is C++; "electronic commerce" coding is business; "Java" number is Java and so on.

#### (4) Experimental data and mining results

The collected data are processed as shown in Table 2.

Table 2 experimental data after treatment

xh	VC++	structure	business	SQL Server	Java	OR	Math
1245793213	1	1	1	1	1	1	1
1245793204	1	1	1	1	1	0	1
1245793232	1	0	1	1	1	0	1
....	....	....	....	....	....	....	....

This chapter uses the Apriori algorithm to analyze a total of 400 students' related records. In the process, the Min sup is set as 0.2, min conf as 0.8 the experiment only considers the student achievement of excellent courses, such as the scores of 85 and over 85. Finally, the association rules concluded in Table 3 presents; scores greater than 60 average, table 4 lists the related rules.

Table 3 association rule mining results (excellent)

Regulation	Support;	Confidence
Math→Java	17.85%	80%
SQL Server→OR	14.2%	75%
Java→SQL server	10.714%	66.667%
.....	.....	.....

Outstanding students, mathematics support java with 17.85%, the confidence level with 80%, which indicates that credibility is still pretty high. By the Java that SQL Server support degree and confidence degree is not very high, indicating that the correlation between them is not very high.

Table 4 mining results of association rules (medium)

Regulation	Support;	Confidence
Math→Vc++	34.37%	81.8%
Vc++→OR	46.875%	80%
computer→Vc++	37.5%	91.667%
.....	.....	.....

For medium grade students, the support between mathematics and VC++ is 34.37%, confidence was 81.8%. So good mathematics, VC++ will corresponding well. The support degree and confidence of Vc++ and OR are relatively high, VC++ The better computer is, the better operations research is, which explains that relatively high correlation between the two. Similarly, the good computer's base correspond to good Vc++. So, the computer foundation improve, correspond, Vc++ can improve.

#### 4.2 Summary of this chapter

This chapter discusses the improved Apriori association algorithm applied to the student achievement database for data mining, on the condition of setting the minsup for 0.2, minconf for 0.6 cases, to analyze the results of the application. In this paper, we use the number of TID to analyze the correlation between data, and make full use of the mining technology to dig out the doors of the curriculum, such as curriculum first subsequent relations and links between to educational better arrange courses, thereby effectively improving the rate of students' learning efficiency.

From this chapter can be learned and fuse professional opened <VC++>< Higher Mathematics > courses exist different conclusions. It also illustrates the validity and practicability of association rules in the analysis course.

### 5. Conclusion

In this paper, the problems in the field of data mining - based on the curriculum relevance of data mining problems were analyzed and studied. Based on the results of a college of professional students, this paper first export four years of the University of All the Results in the Clementine table into the new excel, and then delete it, Apriori modeling, and mining. Firstly, learn the relationship from the professional course to compulsory courses and students in the process and decide whether there is evidence.

## **Acknowledgements**

This paper is the research findings of Jiangsu province education examination and enrollment (Title: The application of multi-level model in the analysis of the influence factors of the computer grade examination in Jiangsu Province; ID: K-e/2015/29).

## **References**

- [1]. Ho. Application Research of XML technology in the design of basic study system of computer application. [Master Thesis of Zhongshan University] Guangzhou: Zhongshan University, 2007
- [2]. Gu Qingfeng, Song Shunlin. Improvement and application of Apriori algorithm in SQL. Computer engineering and design, 2007, 7 (28): 3061-3063
- [3]. Lan Tian. Research and implementation of data mining method of association rules. Master Thesis of Xi'an University of Science and Technology. Xi'an: Xi'an University of Science And Technology, 2008
- [4]. Zhu Ping. Research and discovery of the common college entrance examination information analysis system based on data mining. [Master Thesis of Xinjiang Agricultural University] Xinjiang: Xinjiang Agricultural University, 2008