Analysis And Implementation Of Key Technologies For Teaching Quality Evaluation System In Universities Based On Data Mining

Li Mingwei, Li Yongfang
Yunnan Open University, Kunming 650223, China

***Abstract*-Contraposing to the problems of professionalization of curriculum design, subjective fairness of teaching evaluation and diversity of importance among itemsets caused by selection factors of employers, an improved Apriori algorithm for mining association rules is proposed by introducing the concept of influence factors and combining with the fast pruning mechanism of address mapping. For the problems of indistinguishability of importance and large candidate itemsets in classical algorithms, a fast pruning of candidate itemsets is accomplished by address mapping, and the pseudo-code description of the algorithm is provided and explained in detail. Then, the performance of the improved algorithm is verified and analyzed, which shows the feasibility and efficiency of the algorithm, to make help in teaching evaluation system of universities.**

Keywords-teaching evaluation; Apriori; itemset; ASP

I. INTRODUCTION

Education system is a complex system, which is influenced and restricted by various factors. During the process of high-quality teaching, there must be scientific and systematic evaluation throughout and guidance. Therefore, teaching evaluation is necessary in all aspects of education. Schools play an important role in the systematic education for social people, mainly through classroom teaching. Whether classroom teaching activities are scientific and normative directly determines the level of education. Institutions of higher learning play an important role in the whole education system, which is a turning point and a connecting point for students to enter society. At the same time, institutions of higher learning also represent the highest level of knowledge, technology and culture in society. Therefore, it is of great significance to evaluate and analyze classroom teaching in colleges and universities.

This article firstly studies the relevant theories of data mining, focuses on the advantages and disadvantages of Apriori algorithm based on iteration and pruning, to analyze the problems of classical algorithms and introduce related improved methods to better discover the frequent itemsets hidden in data. At the same time, aiming at the occupationalization of university curriculum, the subjective fairness of teaching evaluation and the diversification of importance among itemsets brought by the selection factors of employers, with the concept of impact factors is introduced, integrated with the rapid pruning mechanism of address mapping. Then, an improved association rule mining Apriori algorithm is proposed to reduce the number of frequent item set generation, which optimizes the pruning efficiency of candidate sets, and improves the time and space complexity of the algorithm. It is expected to meet the background of current vocational education.

II. DESIGN OF TEACHING QUALITY EVALUATION PLATFORM

A. Principle Function

The teaching quality evaluation system aims to improve the quality of teaching in universities. It provides a comprehensive information platform for student, and achievement information management for teachers, which enhances the understanding of the students in the class, and finally achieves statistical analysis and prediction analysis of the results.

ASP.NET represents dynamic server pages and it can interact with databases and other programs to ensure real-time updates of data. The user opens the browser input command and sends the request over the Internet. The server receives the user command and first calls the script that starts running, reads the command request, and executes the script command. After that, the user browser receives the accurate web request command page. The structure mode of embedding the ASP script into the SQL query statement satisfies the requirement that the campus network area is small and the real-time data is updated. Therefore, we will adopt B/S architecture model, with ASP technology as the platform, and SQL Server design the teaching evaluation system for the back-end database. Figure 1 depicts a schematic diagram of the system structure:

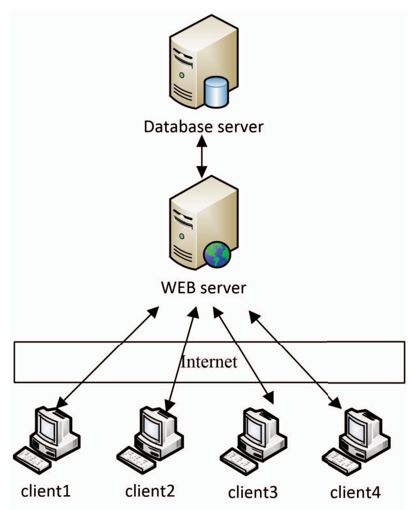

Figure 1. Schematic diagram of system structure

B. System Architecture

The teaching quality tracking platform includes two parts: teaching analysis system and the authority management system. The teaching analysis system is

978-1-7281-4284-5/19/$31.00 ©2019 IEEE
DOI 10.1109/ICICTA49267.2019.00104

mainly for teachers and data administrators. Its functions include data integration of students' curriculum performance, statistical analysis of students' performance in classes, comparative analysis of students' performance in multi-classes, and prediction and analysis of target curriculum performance. Privilege management system is mainly oriented to super administrators of the system. Its functions include user management, role management, privilege management and so on. The overall functional architecture of the system is shown in figure 2.

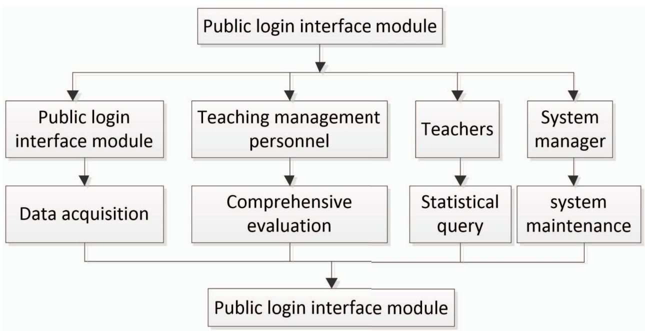

Figure 2. System function structure

III. IMPROVED APRIORI ALGORITHM DESIGN

A. Transactional Data Mapping

For the problem that the candidate set is large in the process of layer-by-layer iteration in Apriori algorithm, before generating the K-order candidate set, the other items of each item set, except the first item, are verified to form a certain sub-item. When the set is a frequent item set of K-1 order, according to the above properties, if it is not in the K-1 order item set, it cannot be a K-order item set, and it is deleted; otherwise, the item set may be K-order. The frequent itemsets of the item set continue to scan the next candidate set. In this way, the size of the K-order candidate set is effectively reduced to some extent, and the time and space overhead is reduced to effectively improve the efficiency of the algorithm.

Mapping sub-function k-1*Status Mapping*(C,L) , where C is the candidate set, L is the frequent item set, K is the order and K is greater than or equal to 2, and the transaction set is processed horizontally by referring to the processing method of AprioriTid. The main steps are depicted as follows:

(1) Firstly, the first set of candidate sets is obtained by scanning the transaction set, and the candidate set is filtered by the support degree parameter. Then the first-order frequent item set L_1 is obtained, and the transaction is added to the corresponding transaction list, and the corresponding K-1 is obtained. The candidate set C_{k-1}.

(2) Acquire the candidate itemset C_{k-1} and generate corresponding transaction list A_{k-1} by the sequence of transactions . According to Apriori, L_k can be obtained by C_{k-1}. Instead of the multiple scanning on database to filter K-order candidate itemset, we use the subset constituted by each itemset in L_k, except the first one, to compare it to the determined sequence of A_{k-1}. If there is not equivalent one, this itemset is deleted from L_k; otherwise, the modified itemset is saved in A_k, and continue to construct the next itemset until all the itemsets in L_k get comparisons, to acquire the filterd L_k.

(3) Using the typical simple transactional database as the base data, the above functions are briefly described. Assuming minimum support minsup=2, regardless of the rule generation, ignoring the minimum confidence parameter. According to the function, A_2 table associated with the second-order frequent itemset L_2 is established. The detailed address mapping after the conversion is shown as figure 3.

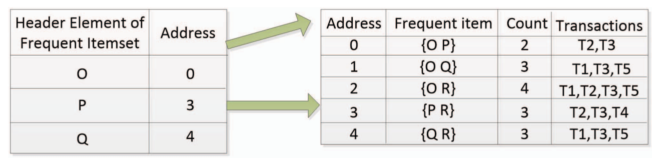

Figure 3. The mapping relation between A_2 and L_2

(4) Then, through L_2 link of above table, a three-order candidate set C_3 is generated. According to the principle of Apriori algorithm, the scanning database compares the candidate sets one by one, and performs pruning. Assuming that the K-order item set is not empty, at least K-1 times need to be scanned from the connection angle. At the same time, the database needs to be scanned again in the process of generating the item set through the candidate set. If the number of items in the database is large, the two combinations are also power. By function *Status Mapping*(C_{k-1},L_k) , we can determine whether the subset constituted by the latter two elements of the item exists in L_2. Using A_2 table as shown in figure 3, the candidate third-order itemsets are deleted to generate C_3 at a faster speed.

Table 3. The result of three-order frequent itemset

Itemset	Support number
{O P R}	2
{O Q R}	3

B. Introduction of User Impact Factor

For Apriori algorithm, only the unique support is set. All the item sets are regarded as the same importance, and there are unscientific problems. In practice, there are often other higher demands with the actual application, which are often different. The project may provide users different levels of attention. If no relevant processing is done, the final result may not be what the user expects. In addition, some projects may only appear in the near future, but they are the projects that users focus on. In order to obtain representative rules, if only recent data is considered, the rules about other products may be lost, and a small amount of data may not be representative. At the same time, there are certain transaction sets with large fluctuations in frequency or probability, and the importance of recording at different time intervals is also different. On the other hand, as the core parameter minimum support and minimum confidence in the research of association rules, there is no uniform standard for the support and confidence settings. Although there are many researchers who have such problem, even there are no guiding rules and references. If the number is too large, the set of options is larger, which increases the time and space complexity of the algorithm, and generates a lot of useless or meaningless rules. Otherwise, if the number of candidates is too small, the set of candidates will be reduced, and frequent itemsets and rules with important practical value may be lost. For special transaction set situations, the user may be required to partially intervene and add some impact factors. To solve above problems, a user influence factor is introduced, and a weighted minimum support degree with a minimum confidence degree are defined. The weighted average principle is combined to effectively solve the above problem and obtain a representative rule.

Definition 1: Supposing itemset $I = \{i_1, i_2, ..., i_n\}$, where i denotes the single item and I is the set of items. Each item is set as a corresponding weight w_j , $\{0 \le w_j \le 1\}$, $\{J \mid 1 \le j \le n\}$. The weighted weight set of corresponding itemset is $\{W_1, W_2, ..., W_n\}$.

Definition 2: For $I = \{i_1, i_2, ..., i_n\}$, each item is set as a corresponding weight w_j . We define $w\sup_w(I)$ as the weighted minimum support degree, recorded as

$$w\sup_w(I) = \frac{1}{k-2}\sum\nolimits_{i \in I} W_i \sup(I) \quad (1)$$

where $w\sup(I)$ is classic support degree, and $\frac{1}{k-2}\sum\nolimits_{i \in I} W_i$ is the weight of W_i that is contained in I .

In the weighted value calculation, K-2 term is used mainly to remove the highest value and the minimum value of the weights. Using the weighted average principle, the individual factors with large influencing factors are removed, and the remaining averaging is satisfied. The processing of extreme values can also solve the influence of personal factors brought by users' influence factors to some extent. On the other hand, as a statistical theory, the setting of support and confidence is also a similar result. The result used to assist decision-making is potential knowledge. There is no absolute right or wrong, and the more is the reference value of the rule. There is still an error with the actual law, which also determines the process of adding the influence factor to adjust the process, and obtaining the rules that users are more interested in. To some extent, it is reasonable to use the weighted average idea to reduce the influence of user factors.

Definition 3: For $I = \{i_1, i_2, ..., i_n\}$ and J , where i and j denote single items and each item is set as a corresponding weight w_j . We define $wconf_w(I \to J)$ as the weighted minimum confidence degree, recorded as

$$wconf(I \to J) = \frac{w\sup(I \cup J)}{w\sup(I)} \quad (2)$$

The weighted minimum confidence is based on the weighted minimum support, which is basically consistent with the definition and concept of traditional confidence. The difference is that the weighted minimum support is used as the evaluation criterion.

C. Algorithm Description

Based on the classical Apriori algorithm, an IFM-Apriori (Impact Factor-based Mapping algorithm for Apriori) algorithm is proposed. The concept of influence factor and the method of user address mapping are introduced to solve the problems of different importance between projects, for too large candidate itemsets and efficiency of generating frequent itemsets. Potential and more realistic rules are discovered, and pseudo-code description of the algorithm is given as follows:

Input: Transaction T, minimum support min sup, minimum confidence min conf, factor $\{W_1, W_2, ..., W_n\}$;   Output: Frequent item set L , rule R

$\forall T \ne Null$
 C_1=Generate_ C_1 (T)
 L_1={c $\in$ C_1| $w\sup \ge \min\sup$ };
 K=2;
 While (L_k>0) do ;
 {K++;
 C_k=Generate_ C_k (L_{k-1});
 Mapping (C_{k-1},L_k) // $K \ge 2$
 L_k={c$\in$ C_k | $w\sup \ge \min\sup$ };
 $L \cup L_k$;
 }
 R=Geuerrrte_Rule(L); // $wconf \ge \min conf$
 End

IV. Application Test of Improved Algorithms in Teaching Evaluation

In empirical analysis, we combine the extracted teacher information file with the student's evaluation information table into a data relation table. As a data mining object using IFM-algorithm, we need to convert this relationship table into a transaction database. After generating the transaction database, the algorithm is used to find the frequent itemsets and generate the association rule base. The minimum support and the minimum confidence are two important parameters of the algorithm. In this application, we assume that the minimum support is 10% and the minimum confidence is 60%. After running the data mining module, a total of 40 rules are mined. We first analyze the teacher's gender, title, education, age and teacher nature, and explain its impact on teaching evaluation.

From table 1, we can see that the support and confidence of the two sets of rules are equal. It can be seen that there is not much relationship between the grade of teaching evaluation and the gender of teachers.

Table.1 The influence of gender for teaching evaluation

Rule NO.	Rule ($X \Rightarrow Y$)	Support degree	Confidence
1	$XB = A1 \Rightarrow PGZF = E4$	34.5%	76.3%
2	$XB = \wedge 2 \Rightarrow PGZF = E4$	36.2%	74.9%

The impact of teacher titles on teaching evaluation is shown as table 2. From table 2, we can see that teachers with professional titles are highly evaluated by students, while teachers with professional titles of lecturers are relatively low. It shows that, to a certain extent, the higher the teaching quality of teachers with higher professional titles, the higher the recognition degree of students.

Table.2 The effect of titles for teachers on teaching evaluation

Rule NO.	Rule ($X \Rightarrow Y$)	Support degree	Confidence
1	$ZC = C2 \Rightarrow PGZF = E3$	2.7%	39.3%
2	$ZC = C2 \Rightarrow PGZF = E4$	3.2%	43.5%
3	$ZC = C3 \Rightarrow PGZF = E4$	4.5%	47.8%
4	$ZC = C4 \Rightarrow PGZF = E5$	6.9%	42.3%

There are many comprehensive factors that need to be considered in the evaluation of teaching quality. We perform data mining on the generated data sets, obtain a large number of frequent itemsets, and analyze these frequent itemsets to generate relevant association rules which are useful for improving the quality of teaching. The code form representation part of the association rules are shown in table 3. These codes cannot be directly stored in the association rule base, and they need to be converted into attribute values again before they can be stored.

Table.3 Part of the association rules table

Rule NO.	Rule ($X \Rightarrow Y$)	Support degree	Confidence
1	$D2 \Rightarrow E4$	14.2%	6838%
2	$C4 \Rightarrow D2$	35.1%	72.3%
3	$D3 \Rightarrow E3$	17.6%	83.1%
4	$C3 \Rightarrow D2$	32.8%	75.2%
5	$C4.AND.C3 \Rightarrow D2$	23.6%	74.5%
6	$C4.AND.D3 \Rightarrow E5$	15.8%	82.1%
7	$C4.AND.D3 \Rightarrow E4$	13.1%	86.9%
8	$C3 \Rightarrow D2.AND.B3$	18.6%	56.1%
9	$B2.AND.E3 \Rightarrow C2$	26.3%	74.1%
10	$C2.AND.E3 \Rightarrow B2$	23.9%	88.3%
...	…	...	...

Although IFM-Apriori specifies stricter minimum support and confidence, since the number of students and subjects in the score data set, some of the results in the algorithm are repeated, and some are wrong because of meaningless reason. Taking into account the principle of truth-seeking, there are the original results and they are analyzed based on the association rule screening method proposed above. A more complete screening of the results to select a better one is useful to us. The results of the preliminary screening are shown in figure 4.

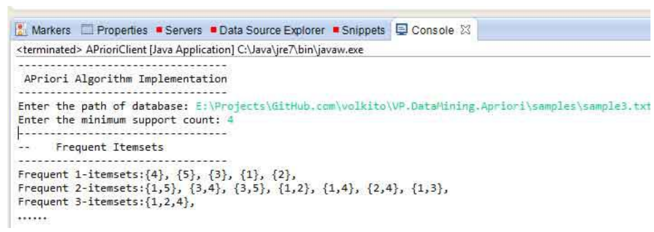

Figure 4. Original mining results

V. CONCLUSIONS

This paper studies the Apriori association rule mining algorithm, designs and implements an improved association rule mining algorithm, to reduce the number of frequent item set generation, optimize the pruning efficiency of candidate sets, and minimize the time complexity of the algorithm. Degree and space complexity are expected to meet the background of current vocational education, which can more effectively, accurately and deeply discover the inner relationship between items, and explore the rules. Finally, the algorithm is applied to teaching evaluation and employment analysis. Through the steps of preprocessing and cleaning the original data, the internal connections are discovered, and the deeper key factors and rules are mined. The curriculum and teaching mode of the school are eEvaluation methods, training models and employment directions and policies provide reasonable optimization and improvement suggestions to support the improvement of teaching management systems and leadership decisions.

REFERENCES

[1] Yuan Y, Hui L I. Research on Teaching Evaluation System Based on Data Mining. Computer & Modernization, 2009, 28(1):2641-2645

[2] Cheng Y, Wang Y L. Design of Network Education Evaluation System Based on Data Mining Technology. Applied Mechanics & Materials, 2013, 241-244:2849-2852

[3] Hung J L, Hsu Y C, Rice K. Integrating Data Mining in Program Evaluation of K-12 Online Education. Journal of Educational Technology & Society, 2012, 15(3):27-42

[4] Dugas M, Trumm C, Stäbler A, et al. Case-oriented computer-based-training in radiology: concept, implementation and evaluation. Bmc Medical Education, 2001, 1(1):1-7

[5] Liu G J, Wang W Y, Deng-Feng W U, et al. The application study of the decision tree AND fuzzy clustering algorithm on the teaching quality evaluation. Journal of Northeast Normal University, 2009, 41(3):36-39

[6] Rieder J, Khan U I, Heo M, et al. Evaluation of a community-based weight management program for predominantly severely obese, difficult-to-reach, inner-city minority adolescents. Childhood Obesity, 2013, 9(4):292

[7] Menzies T, Brady A, Keung J, et al. Learning Project Management Decisions: A Case Study with Case-Based Reasoning versus Data Farming. IEEE Transactions on Software Engineering, 2013, 39(12):1698-1713