Research on the Evaluation of College Curriculum Teaching Effect Based on Association Rules

Qiancheng Chen(✉)

Shenyang Normal University, Shenyang, China
touming821106@163.com

Abstract. The effect of teaching monitoring and evaluation in colleges and universities is not ideal. External monitoring and internal monitoring indicators are inconsistent, which is unable to achieve the effective use of evaluation data. Part of the evaluation process is more inclined to the evaluation of management and teaching effect and lack of a monitoring index system for the teaching process. Some colleges and universities have carried out network evaluation, which broadens the time and space of evaluation in the form of evaluation. However, they do not make effective use of the large amount of data generated in the evaluation process. Based on the research of data mining algorithms, this paper proposes the effective application of data mining in the internal teaching quality monitoring and evaluation of colleges and universities. It combines with a large number of original data clustering analysis, puts forward a new model construction idea after the weight distribution of the original teaching evaluation model indicators, and applies association rules analysis. The data preparation and data processing are described in detail. The performance of the improved algorithm proposed in this paper is compared by using the same data set and running environment. The effect of this research is proven. By analyzing the results of teaching evaluation, teachers' information, and curriculum characteristics, we can get the ideas and measures to effectively use the results of teaching evaluation and improve teachers' teaching ability.

Keywords: University curriculum · Teaching effect · Association rules · Index evaluation · Data mining

1 Introduction

Under the background of the new curriculum reform, English teachers need to become researchers of teaching behavior. They also need to process and summarize the existing phenomena and laws. However, the current application of teaching behavior in English is mainly to explore teachers' teaching ideas and teaching phenomena. The data collection methods used in the study are mainly questionnaires and classroom observation. Without analyzing the deeper teaching characteristics and teaching modes under the teaching behavior, the data analysis also stays at the level of simple statistical analysis [1]. At

C. F. Peng et al. (Eds.): EIMT 2023, AHSSEH 8, pp. 369–377, 2023.
https://doi.org/10.2991/978-94-6463-192-0_48

present, the importance of teaching process data and teaching process evaluation in teaching research has gradually emerged. The development of data analysis technology provides an effective way to deepen the depth of teaching research. However, the combination of teaching behavior analysis based on data analysis technology and English subject teaching research is still lagging. Therefore, the analysis of English teaching behavior is one of the effective means to deepen the depth of English teaching research [2].

Apriori has also been widely used in the field of education. Researchers often use the Apriori to mine the hidden correlation between various dimensions or factors in a certain research direction. Literature [3] adopts the Apriori algorithm to analyze the access logs of students using the college physics learning system. It also records the frequently accessed knowledge points and infers students' interests and habits. Then, it analyzes their learning trends to provide personalized resource recommendations. Literature [4] uses the Apriori algorithm to analyze the data in the online learning platform. It is found that age, online learning time, the number of online learning modules and the attendance rate of face-to-face teaching are related to students' academic performance. Literature [5] chooses the Apriori algorithm to analyze students' performance data and infers the relationship between basic courses and professional courses of electrical specialty. Literature [6] proposes an Apriori algorithm based on a matrix. The algorithm only needs to scan the database once. Literature [7] proposes the Napriori algorithm, which improves the shortcomings of the traditional Apriori algorithm by compressing the candidate set. Literature [8] proposes the parallel FP-growth algorithm, which does not improve the efficiency optimization problem due to the high communication cost of the algorithm. Literature [9] proposes a distributed frequent item set-mining method based on Map Reduce. There are still many problems in association rule mining. For the limitations of the algorithm itself, in the face of a large number of multi-dimensional data, the complexity of the algorithm running time and space will be increased, which will bring challenges to the research of the algorithm.

This paper proposes to improve the efficiency of mining frequent item-sets. The improved algorithm can be applied to the teaching index association analysis. Besides, the improved algorithm can be applied to the teaching index association analysis, and be extended to other types of data, such as course performance indicators, behavioral indicators and other teaching evaluation indicators to verify the necessity and importance of this study. To efficiently evaluate the level of teaching effectiveness, the data is firstly preprocessed and converted into a data set suitable for teaching effect evaluation. Similarly, according to the characteristics of teaching data, the operation efficiency of the algorithm is improved. The accuracy of the analysis results of the teaching index association classification model based on the improved Apriori is verified. Then, aiming at the situation that there may be a large number of redundant rules in the results, this paper proposes to use the promotion degree to screen out the effective strong association rules. Finally, it verifies the accuracy of the results of the teaching index association classification model through experiments.

2 Data Mining Optimization Algorithm

In this paper, the association rule mining algorithm is applied to the research and analysis of teaching index data in colleges and universities. However, due to the shortcomings of the algorithm, such as occupying a large amount of memory when running and repeatedly scanning data sets, the execution efficiency of the algorithm will be affected. The quality of the mining results will be affected, which requires an efficient association mining algorithm.

In the analysis of teaching behavior, the frequent mining results of the teaching effect evaluation of university courses are selected to assist the analysis of teaching behavior characteristics. After the function of the teaching platform is improved, Apriori can also mine association rules for teachers' teaching evaluation results and basic information. The Apriori can also explore the influencing factors of teaching, predict teachers' teaching characteristics and intervene in time. It can assist teachers' teaching decisions and management's educational decisions [10]. The association rule structure is presented in Fig. 1.

(1) Itemset: a collection of items. For example, {Warm up, Vocabulary learning} is a binomial set, and Warm up and Vocabulary learning are items.
(2) Subject set: a set of subjects. A subject is equivalent to an item set, and a subject set is equivalent to a set of item sets. Let the number of item sets be Data _ count.
(3) Support count (support _ count): the number of subjects in a set of subjects.
(4) Support: It is equivalent to the probability that a certain item set appears in the discipline of the discipline set. The calculation formula is as follows:

$$\lambda = \sqrt{\frac{S}{E}} \tag{1}$$

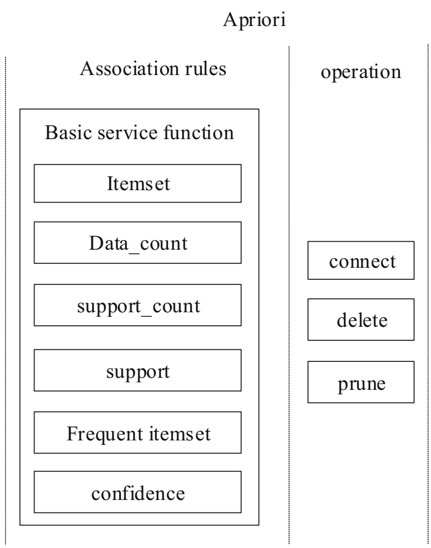

Fig. 1. Association rule structure

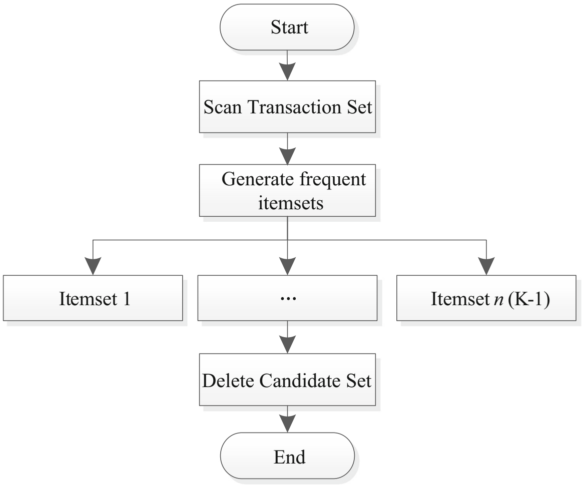

Fig. 2. Apriori algorithm flow

In Eq. 1, λ is support, S is support _ count, and E is Data _ count.

(5) Frequent item set: the minimum support (min _ support) is set in the algorithm to control the output of the result. The frequent item set is the item set whose support is greater than or equal to the set support threshold.

(6) Confidence: the proportion of j item-sets appearing in the discipline with I item-sets. The algorithm can set the minimum confidence (min _ conf) to control the output of the strong association rule results. The calculation formula is as follows:

$$\phi_{(i \to j)} = \sqrt{\frac{\lambda_{(i \to j)}}{\lambda_{(i)}}} \tag{2}$$

where, $\phi_{(i \to j)}$ is the proportional value.

The association rule mining of the Apriori algorithm is divided into two parts, namely, frequent item setgeneration and strong association rule generation. The relevant operation steps involved in the frequent item setstage of the Apriori algorithm are displayed in Fig. 2.

(1) Delete: Scan the subject set to generate the frequent k item set.
(2) Connection: Frequent (k-1) item-sets are self-connected to generate candidate k-item-sets.
(3) Remove it from the candidate set.

3 The Evaluation Model of Teaching Effect in Colleges and Universities

The rating information of teaching quality comes from the evaluation and scoring of each teacher by students in the school system after each semester, which is also the main data set composed of the teaching evaluation system. Different attributes of teaching evaluation are also graded at different levels, for example, classroom driving includes high, medium, and low levels [11]. There are also Boolean quantities. For example, the comprehensive evaluation has two attribute levels of yes and no. In the scoring system, different attribute levels correspond to different score intervals to obtain the teaching evaluation information of teachers, as shown in Table 1.

When setting the impact factor of the algorithm, the impact factors of students' participation in practice and assessment methods that college students are more interested in were increased. Then, we further preprocessed the data in combination with the above information, converted the attribute values into codes, and finally formed the corresponding values. The information that came from a university after preprocessing is displayed in Table 2.

College students generally think that the theoretical knowledge of basic science is too boring. They also think that the traditional face-to-face teaching can no longer meet the needs of them. Thus, we can increase the investment in the practice of the course. We can also make the course more diverse so that students' understanding of theoretical knowledge can be changed from perceptual to rational.

Table 1. Teacher's teaching evaluation information

Index		Subjects	
No.	Items	Teachers	Students
A1	Classroom motivation	High	Medium
A2	Classroom interaction	High	Medium
A3	Course acceptance	-	High
B1	Verbal expression ability	High	Medium
B2	Participation	High	High
…	…	…	…

Table 2. Information of teaching evaluation

No.	Associated item	Weight
A1	A2, A4, B2	0.075
A2	A3, A4, B1	0.048
A3	A1, A4, B3	0.151
B1	A2, A4, B2	0.059
B2	A1, A2, B3	0.099
…	…	…

4 Evaluation Effect Verification

4.1 Experimental Platform Settings

The paper selects university courses and takes the evaluation data of two majors as the experimental object. The experimental data came from the literature [12]. Thus, the data obtained are normative and authentic, but the original data need to be preprocessed by data cleaning, data reduction, data conversion, and other operations to generate Boolean data suitable for improving the Apriori algorithm. Then, it realizes the application of the improved algorithm in the data.

Compared with the traditional ID3 algorithm, the data set is based on 500 subject data and 38 common items. The number of 3-order item sets generated in theory is 950, while the number of 3-order item-sets generated in this algorithm is 690, which is more obvious when the 4-order frequent item-sets are used. The number of 3-order item-sets is 9880, which is reduced to 1135 by combining support and pruning in this algorithm. When generating 3-frequent item-sets, the number is 600 less than that of 2-frequent item-sets. The first 500 pieces of data (40000 pieces in total) extracted from the data set are adopted as the data source for the example analysis. Each record contains 38 attributes. The algorithm in this paper is analyzed by observing the execution time of the algorithm from the fixed minimum support and the change of the time required in different minimum support States. The minimum support min_ sup = 0.05 is unchanged. According to the 38 attributes of each record, a candidate item set M is generated.

4.2 Analysis of Experimental Results

(1) The perform efficiency analysis

The experimental results of the improved Apriori algorithm indicate that the candidate item sets at different stages are formed through the change of the curve. The results are displayed in Fig. 3.

In Fig. 3, with the increase of the candidate set level, the number of candidate item sets and frequent item sets is gradually reduced. The response time is also gradually reduced. Since the candidate item sets of the upper level are reduced, the consumption of the generated item sets of this level is correspondingly reduced. The time difference between the two algorithms is not very large, and the overhead time of the ID3 algorithm is gradually reduced with the pruning and mapping. It takes less time and slowly reveals its advantages.

(2) Comparative analysis of confidence level

First, the minimum confidence m_c = 0.1 is given, and the minimum support is set at 1%, 2%, 3%, 4%, 5%, and 6% respectively. The rules with a boost greater than 1 are retained. The results of comparing the association rules of the Apriori improved model and the ID3 algorithm are displayed in Fig. 4.

Under the same confidence threshold, the number of association rules obtained by the algorithm is the same, that is, the results of association rules are the same. The improved algorithm ensures mining accuracy under the same parameters and has certain

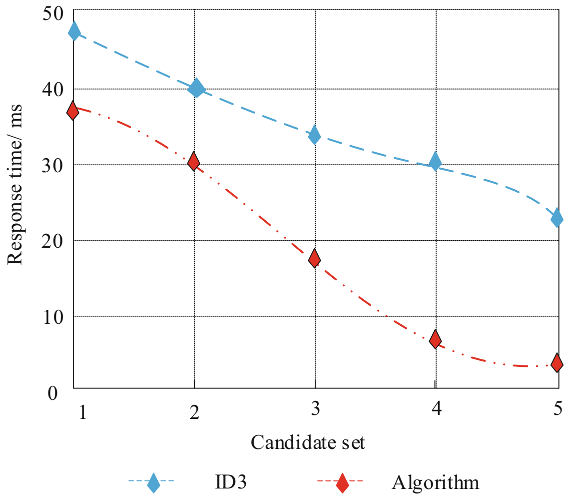

Fig. 3. Execution time of different algorithms

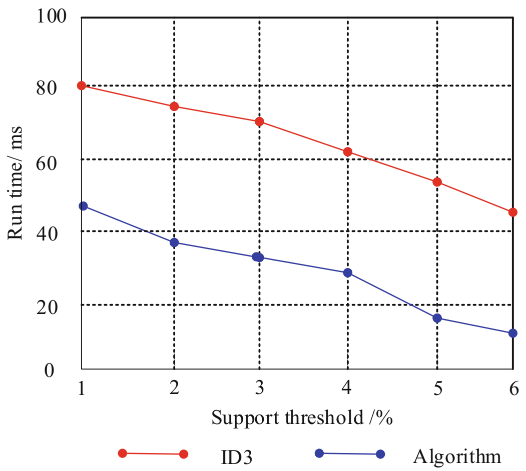

Fig. 4. Confidence comparison results

reliability. The efficiency of the improved algorithm in this paper is obviously better than that of the ID3 algorithm.

5 Conclusion

In view of the deviation between the result of the clustering analysis of teaching evaluation data and the proportion of students' course scores in the original teaching evaluation model, the weights of some indicators in the original teaching evaluation model are adjusted and the weights of indicators are redistributed. Through experimental verification, it is concluded that the new model is easier to reflect the relationship between teachers' teaching evaluation and students' course scores. It is more conducive to the improvement of teaching quality.

Because there are too many elements involved in teaching quality monitoring and evaluation, data collection can not cover all aspects. Through the analysis and research of the collected data by using data mining methods, this paper proposes that data mining methods should be applied to teaching evaluation mechanisms, which is very helpful.

In future work, according to the correlation analysis of teaching evaluation results and curriculum characteristics, it is proposed that according to the curriculum characteristics factors, coefficient factors should be added respectively before comprehensive analysis. The specific factor calculation is the direction of beard ability.

References

1. Zhixiang Zhou, Yuxuan Zhang, De'nan Li. Evaluation of Higher Education Competitiveness—Based on Analytic Hierarchy Process [J]. Journal of Educational Research and Policies, 2021, 3(7): 111–119.
2. Shuiyun Deng, Xiang Que. Research on the teaching assessment of students of science and engineering teachers in a university [J]. Computer Applications in Engineering Education, 2019, 27(1): 5-12.
3. Jesus Silva, Noel Varela, Luz Adriana Borrero López, Rafael Humberto Rojas Millán. Association Rules Extraction for Customer Segmentation in the SMEs Sector Using the Apriori Algorithm [J]. Procedia Computer Science, 2 019, 151: 1207–1212.
4. Wu Chunqiong, Yan Bingwen, Yu Rongrui, Huang Zhangshu, Yu Baoqin, Yu Yanliang, Chen Na, Zhou Xiukao. Improvement of K-Means Algorithm for Accelerated Big Data Clustering [J]. Internation al Journal of Information Technologies and Systems Approach (IJITSA), 2021, 14(2): 99 -119.
5. Afgan Fazri Handoko, Antoni Wibowo. Three-Layer Data Clustering Model for Multi-View Customer Segmentation using K-Means [J]. International Journal of Recent Technology and Engineering (IJRTE), 2020, 8(6): 1840–1846.
6. Zeeshan Akbar, Jun Liu, Zahida Latif. Discovering Knowledge by Comparing Silhouettes Using KMeans Clustering for Customer Segmentation [J]. International Journal of Knowledge Management (IJKM), 2020, 16(3): 70-88.
7. Li YY, Yin SH.Mining algorithm for weighted FP-tree frequent item sets based on two-dimensional table(Conference Paper) [J].Journal of Physics: Conference Series.2020, 1453(1): 1256–1264.
8. Lei Jing, Liu Wei. Association Rule Mining Algorithm in College Students' Quality Evaluation System [J]. Journal of Electrical and Computer Engineering, 2022, 2022.
9. Wang TF, Xiao BR, M WX. Student Behavior Data Analysis Based on Association Rule Mining [J]. International Journal of Computational Intelligence Systems, 2022, 15(1): 56–60.
10. Ejaz Ahmed S. Statistical and Machine-Learning Data Mining: Techniques for Better Predictive Modelling and Analysis of Big Data [J]. Technometrics, 2021, 63(2): 280-280.
11. Fister Iztok, Fister Iztok. Computational Intelligence and Data Mining in Sports [J]. Applied Sciences, 2021, 11(6): 2637-2638.
12. Yang Xiaobo. Algorithm Study of New Association Rules and Classification Rules in Data Mining [J]. Journal of Physics: Conference Series, 2021, 1732(1): 745-755.