

Application of Improved Eclat Algorithm in Students' Evaluation of Teaching

Yuzi Dou^a, Xiwei Fei, Rui Zhu, Tianzhu Gao, Yanbing Wu and Lei Ma

School of Computer and Communication Engineering, Liaoning Shihua University, 113001 Fushun Liaoning, China

Abstract. The evaluation system of students is to find a way to solve the status way according to the exact needs of students and the teaching requirements of teachers, so as to improve the teaching level of teachers and improve the quality of school education. This paper uses the real evaluation sample and uses the data mining association rule algorithm to comprehensively analyze the massive data of the evaluation data and the basic information of the teacher. The purpose is to obtain the association rules between the teacher's comprehensive information and its evaluation results. Using the evaluation data to explore its core issues. In this paper, the Eclat algorithm of association rules improves the problem of insufficient memory and occupying a large amount of time when searching for frequent itemsets in the data. The breadth-first algorithm is added to save operation time and improve the efficiency of the algorithm. The effectiveness of the improved algorithm is verified by comparative experiments and applied to the evaluation system so as to provide suggestions for the professional development of teachers from an objective perspective, and to build a harmonious, "people-oriented" evaluation system for students.

Keywords: Student evaluation of teaching; Association rules; Eclat algorithm; Python.

Introduction

Teaching evaluation is an important means for master and understand teaching information, the evaluation results are closely related to teacher assessment, award promotion and so on. The evaluation of teaching by the teaching management department, the leadership of the department and the teaching supervision group, evaluation of experts, peer evaluation and other methods are common means of monitoring measures [14]. Among them, student evaluation is relatively credible and effective in terms of practicality and operability. The association rule algorithm is used to further deepen the evaluation data, in order to facilitate real-time analysis and tracking of data, and to enhance the visibility effect of students' evaluation work.

In [3], the author's idea of improving the Eclat algorithm is based on the Spark platform, and by parallelizing the data into different computing nodes to achieve parallel computing, to improve operational efficiency, but the disadvantage is that it is not suitable for large data operations. Literature [4] combines the Eclat algorithm with the data programming model MapReduce to speed up data computational efficiency, but still does not implement the a priori nature of association rules. In [7], the author's idea of improving the Eclat algorithm is to divide the items in the database into different modules that are not overlapping, in order to reduce the number of candidates generated by the intersection in the Eclat algorithm and improve the operation efficiency. However, the shortcoming is that the distribution of similar data is uneven and the frequent set loss occurs during the division process. In [8], the

Eclat algorithm is improved in the optimization and pruning of candidate sets, which reduces the computational complexity. The disadvantage is that it is not put into practical application. About the application of students' evaluation of teaching, In [2], when preprocessing information, the attributes of interest are combined to improve the Apriori algorithm and applied to the evaluation. Although it is simple and easy to understand, the conclusion is not scientific and comprehensive. In [5], in order to realize mining rules from multiple data sets, the idea of meta-association rules is proposed, and the explicit meta-association rules mining algorithm and fuzzy-element association rules mining algorithm are compared and summarized, But there is no dedication to evaluation. In [10], the Langbach coefficient, time series based detection method and data analysis tool SQL Server are combined to verify the validity of the evaluation data, and an effective method for solving the evaluation status is proposed, but did not elaborate on improving data mining.

This paper improves the inadequacies of the Eclat algorithm. At the same time, it is applied to the student evaluation system to find out the problems existing in the evaluation of the status quo and to find a solution.

1 Student evaluation status

1.1 Teaching methods

At present, the main ways [5] for students to evaluate teaching are

^aAuthor: 412602026@qq.com

Fund project: Project supported by Liaoning Provincial Natural Science Foundation of China (20180550130).

questionnaires (paper questionnaires and online surveys), discussion methods, and indicator system methods. Relatively speaking, online evaluation has more advantages [1]: the operation process is simple, the data collection is accurate and fast, which is conducive to analysis and summary, green environmental protection and so on.

2 Main factors affecting students' evaluation of teaching [3-7]

2.1 Index system is imperfect

Evaluation indicators play an important role in the entire indicator system, and also serve as an important criterion for testing the level of teachers' teaching. However, because many colleges and universities have different educational purposes and levels, the use of a unified standard format indicator system can not achieve the pre-set goals. Moreover, the indicators are too general and single, not personalized enough, ignoring the differences between different courses in teaching.

2.2 Teachers and students have a shallow awareness of teaching

In terms of students: Students have poor subjective consciousness, do not actively participate in activities, and do not objectively and rationally evaluate.

In terms of Teachers: The teacher did not treat the teaching correctly, and felt that as long as the class was taught well, the evaluation was just a form.

2.3 School management offside and misplacement

The organization of the evaluation needs to be improved. The general evaluation time is set to be the mid-term to the end of the period or after the end of the semester, so that the evaluation of education and teaching cannot be timely feedback and resolved. The rules of evaluation are generally formulated by the school's teaching management department (such as the Academic Affairs Office), ignoring the effective communication between teachers and students.

2.4 Improper operation of the evaluation

The effective implementation of student evaluation involves the common interests of both teachers and students. It is important to establish and implement a reasonable procedural mechanism to ensure that teachers and students can actively participate in the evaluation. If there is no other matching supporting system, this will even have a negative impact on teachers' improvement of teaching level.

3 Data mining association rules

In recent years, Data mining [2] (DM) has aroused great concern from all walks of life. The main reason is that the technology can be converted into useful information and knowledge using a large amount of data analysis and widely used.

Association rules [3] are a method applied to data mining. The "Beer and Diaper" event at Wal-Mart is the most classic case of association rules.

Database association rule mining mainly includes two processes [24]:

(1) Data preprocessing (including discretization, addition and deletion of missing values, etc.);

(2) Find frequent itemsets to generate corresponding strong association rules (meeting min-sup and min-conf).

The specific concepts of support and confidence are as follows: Support is the percentage of items in a database that contain a transaction, that is,

$$\text{support}(A \rightarrow B) = P(A \cup B) \quad (1)$$

$P(A \cup B)$ indicates the probability that A and B transactions are purchased at the same time (the probability of simultaneous occurrence).

Confidence refers to determining how often A and B transactions occur simultaneously in a set of items containing A transactions.:

$$\begin{aligned} \text{confidence}(A \rightarrow B) &= P(B|A) \\ &= \text{support}(A \cup B) / \text{support}(A) \\ &= \text{support_count}(A \cup B) / \text{support_count}(A) \end{aligned} \quad (2)$$

The association algorithm is used to analyze the rules that satisfy a certain strength. The higher the support and confidence, the stronger the rule.

3.1 Apriori algorithm [1] and FP-Growth algorithm

Apriori ranks among the top ten data mining algorithms, using an iterative method called layer-by-layer search (hierarchical search). The algorithm principle is divided into connection and pruning, which reduces the amount of manual calculation and is easy to code. But the disadvantage is that generating a large number of candidate sets results in too much computational effort; the database needs to be scanned repeatedly.

FP-Growth (Frequent-Pattern Growth) algorithm is based on Apriori improved frequent pattern growth algorithm. Compared to Apriori, it only traverses the database twice, which improves search efficiency.

3.2 Eclat algorithm for association rules

The Apriori algorithm and the FP-Growth algorithm introduced above all mine frequent patterns from the transaction set of the TID item set format (itemsets). The data formats of the above two algorithms are horizontal data formats. There are two ways to represent the database format: horizontal and vertical. And the data format that is often vertical in computing is more efficient than the horizontal data format. The Eclat algorithm is a vertical data representation based on depth-first algorithm.

The Eclat algorithm uses an inverted table to increase the speed of frequent item set generation. However, the disadvantage is that the deletion of the candidate set is not performed, resulting in a large number of candidate sets, which affects the efficiency of the algorithm.

3.3 Improved Eclat algorithm for association rules

The traditional Eclat algorithm does not realize the a priori nature of the association algorithm. That is, all non-empty subsets of any frequent itemsets must also be frequent, whereas the parent classes of any infrequent subsets are also infrequent sets. In this paper, the way to improve this algorithm is mainly to reduce the operation time by reducing the number of candidate sets generated by frequent sets of self-join intersections.

The improved Eclat algorithm in this paper is defined as the Eclat' algorithm, which is improved by breadth-first (BFS, similar to hierarchical traversal). The breadth-first algorithm is a graphical search algorithm, starting from the root node and accessing other nodes layer by layer according to the hierarchical structure. First, access all the neighbors from the root node to k, and mark the access, Then continue to visit other nodes with distance (k+1) from these neighbors and so on, until the target is found, the algorithm terminates. Slightly prioritizes the use of queues relative to depth-first implementations with loops, which is slightly better in reducing operating time and improving operational efficiency. Language description of the Eclat' principle:

Input: Vertical format database Tidset, minimum support is min-sup.

Output: All frequent itemsets L.

(1) Scan the database for the first time and get item set L_1 ,
 $L=L \cup L_1$.

(2) For all $x_i \in L_1$ do

(3) For $x_j \in L_1, j>i$ do

(4) Get x_i first neighborw=FirstAdj(L, x_i),

(5) Find all items related to x_i for self-joining Candidate set

$R_w=L_i \cup L_j$.

(6) Tidset R_w =Tidset $L_i \cap L_j$,

(7) Then find x_i next neighbor point relative to w,

W =NextAdj(L, x_i ,w),

(8) If Tidset $R_w \geq$ min-sup, do

(9) $R_{w_i} \in L_i$

(10) $L=L \cup R_w, T_i=T_i \cup R_w$

(11) Find all items related to W_i to self-join to generate candidate sets $R_w=R_{w_i} \cup R_{w_j}$.

(12) Until the end of the algorithm.

L_1 is the item set obtained by scanning the database for the first time, x_i / x_j is the set of items contained in L_1 , and R_w is the set of adjacent items of an item set.

3.4 Comparative experimental design and results analysis

This experiment compares the Eclat algorithm with the Eclat' algorithm using the data set collected by the school evaluation system, to verify the validity of the Eclat' algorithm. Due to the needs of the experiment, the data set is divided into two categories:

dense data sets, and loose data sets. Among them, the intensive data set contains the 2013-2017 teacher evaluation information, and the sparse data set contains the 2016-2017 teacher evaluation information. The experimental platform is PC (Intel i3, CPU 2.30GHz, memory 4GB), and the operating system is Win7 Ultimate. In the course of experimental operation, in order to avoid interference of experimental results with non-procedural factors, In the course of experimental operation, in order to avoid the interference of non-procedural factors on the experimental results, the results show a certain degree of deviation, this experiment takes the average value of several experimental operations to get the algorithm running time. Because the density of the set data set is different, the corresponding support threshold is selected for experimentation, which is convenient for comparing the experimental operation. Figure 1 shows a comparative experiment using a dense data set.

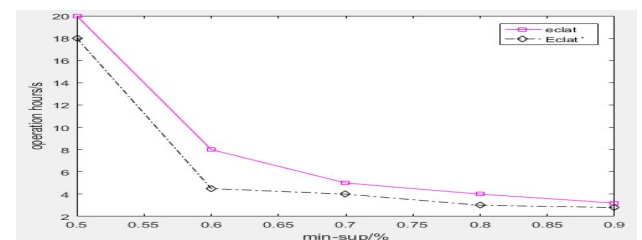


Figure 1. Concentrated data set experiment comparison chart.

Figure 1 shows the experimental operation with a support threshold of 50% to 90%.The results show that when the support degree is large, the running time of the two algorithms tends to be close; when the support degree is small, the running time difference between the two algorithms is large, if the min_sup is 0.6, the improved algorithm is reduced by 10.6% compared with the original algorithm; and when the min-sup is 0.8, the improved algorithm is reduced by 2.4% compared with the original algorithm.

Figure 2 is a comparative experiment using a loose data set.

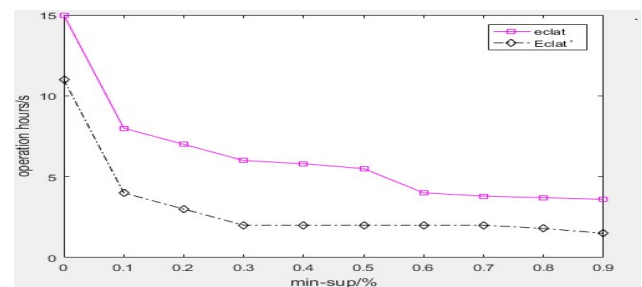


Figure 2. Comparison of loose data sets.

Figure 2 shows the experimental operation with a support threshold of 0 to 90%. The results show that the Eclat' algorithm runs shorter and runs faster. When the min-sup is 0.5, the improved algorithm is reduced by at least 12.5% compared to the original algorithm.

From the experimental results in the above figure, the operating efficiency of the Eclat' algorithm is slightly better than the eclat algorithm. Especially when the data set is slightly loose and the support setting is low, the Eclat' algorithm is more effective.

4 Application of Eclat' algorithm in evaluation system

This paper uses PyCharm to build Eclat' algorithm model based on Python language and applies it to students' evaluation of data mining. This time using Python 3.6.0.

4.1 Collect data for preprocessing

This paper integrates the teacher's comprehensive information table and evaluation transcripts in the school teaching management system into DM. First, it sorts out nearly 3,000 records of data preprocessing. As shown in Table 1.

Table 1. Information consolidation table.

Number	Age	Education	Job title	Party member	Teaching age	Research project	Grade
0001	34	Doctor	Lecturer	YES	3	Sch	Inter
0002	26	master	lecturer	YES	1	No	Passing
0003	29	Bachelor	Assistant	NO	2	No	Failing
0004	47	Doctor	AP	NO	9	Pro	Good
0005	46	Doctor	Professor	YES	12	Pro	Excellent
0006	49	Doctor	professor	YES	13	Nat	Excellent
0007	35	master	Lecturer	NO	4	Pro	Passing
0008	39	Master	Assistant	YES	5	Sch	Inter
0009	40	Doctor	AP	YES	10	Mun	Good
0010	43	Doctor	professor	NO	7	Pro	Inter
.....							

4.2 Statute and convert data [5-6]

Discretize sample data based on MYSQL statement conversion. The age stage is set to: <20-30>G1, <30-40>G2, <40-50>G3, <50-60>G4. The academic classification is set to: Bachelor A1, Master A2, Doctor A3. The job title classification is set as: Assistant B1, lecturer B2, Associate Professor(AP) B3, Professor B4. Party members are classified as: Yes C1, No C2. The teaching years are set to: <1-5> D1, <6-10> D2, <10-15> D3. Scientific research project classification: No E1, Sch E2, Mun E3, Pro E4,

Nat E5. The classification is set as follows: failing F1, passing F2, inter F3, good F4 and excellent F5.

4.3 Association rule application analysis [2]

Using the improved Eclat' algorithm to analyze the sample data of the collected student evaluation, set min-sup to 10% and min-conf to 75%. Some of the results are shown in Table 2 below:

Table 2. Results analysis table.

Num	Age	Edu	Title	Party	Age	Study	Grade	Sup%	Conf%
0002	G2				D2		F2	12	75
0011		A1		C2	D1		F1	20	89
0058		A2			D1	E4	F2	45	96
0111	G2			C1		E3	F3	38	78
0234		A3	B4				F5	67	85
0067		A3	B3		D3		F5	15	79
.....									

The partial rules resulting from the analysis of the results are as follows:

Rule 1 indicates that the teacher who is under the age of <30-40> and the age of <6-10> is a member of the party, the score of passing is 12%, and the confidence is 75%.

Rule 2 indicates that undergraduate degree, municipal research projects, and non-party teachers, the degree of support for failing grades is 20%, and the confidence is 89%.

Rule 3 indicates that teachers with a master's degree and a teaching age of <1-5> have provincial-level scientific research projects, and the degree of support for passing the test is 45%, and the confidence level is 96%.

Rule 4 indicates that teachers who are under the age of <40-50>, who are party members and have a teaching duration of <10-15>, have a good support rating of 38% and a confidence level of 78%.

Rule 5 indicates that teachers with doctoral degrees, professors and national-level research projects have an excellent score of

67% and a confidence level of 85%.

Rule 6 indicates that teachers with doctoral degrees, associate professors, and teaching ages of <10-15> have a good support rating of 15% and a confidence level of 79%.

4.4 Result analysis

Through the above association rules, the teacher's age, education, professional title, teaching age, party members, and research project level all have an impact on the evaluation results. For example, the doctoral degree is higher than the master's degree and the following teachers, the score is high, and the confidence is high.. The professor's assessment is higher than that of the associate professor and below. The longer the age, the higher the confidence. The higher the level of scientific research projects, the stronger the practical ability and the more innovative the knowledge taught in the teaching. Teachers aged between <50-

60> have higher scores and higher confidence than teachers of the following age groups.

The information available from the reasoning rules provides a reference value for the academic management department in the curriculum arrangement, so that the teachers' comprehensive resources are fully utilized. At the same time, the school should adjust the structure of the teacher team and improve the title selection system. Encourage participation in practical research projects to improve teachers' own level and quality of school teaching. In academic research, key teachers drive young teachers to make continuous progress, so as to achieve the goal of improving teaching level.

5 Conclusion

The association rule is applied to the student evaluation teaching system. By improving the Eclat algorithm, the complexity of sorting and analyzing teaching data is reduced, and the efficiency of the improved algorithm is improved (fast). However, it is necessary to take into account the preparatory work in the early stage, that is, the initial working time required to apply the breadth priority. So the Eclat' algorithm still has some shortcomings. At the same time, we will consider using Weka to implement the Eclat' algorithm and further explore its application. This is the main topic of this paper.

In order to facilitate the integration and analysis of evaluation information, this paper proposes the Eclat' algorithm. For the effectiveness of information collection, this paper has the following measures to improve the student evaluation system:

(1) May strengthen the student to evaluate the teaching propaganda work. For the evaluation system indicators, students are mobilized before the evaluation and special training to clearly define the purpose of evaluation.

(2) According to the classification design index system of different disciplines, the evaluation and evaluation form is scientifically and reasonably improved to make it more flexible and humanized. At the same time to organize experts, students, teachers and other people to discuss. The overall setting should pay attention to teaching attitude, teaching content, teaching methods, teaching methods and teaching effects, and need to be divided according to the emphasis.

(3) The related functional departments of the organization and management evaluation shall fully implement the evaluation rules, deal with the evaluation results in a fair and equitable manner, and formulate rigorous solutions to the problems.

(4) Focus on outstanding achievements at home and abroad, and learn from each other's strengths.

References

1. X.Y. Cui, J.Y. Sun, Research and Application of Improved Eclat Algorithm, *Computer Engineering and Design*, **39**, 1059-1063 (2018).

2. G.J. Ren, Y.N. Bai, Application of Association Rules Mining in Teaching Evaluation in Colleges and Universities, *Computer & Digital Engineering*, **42**, 1526-1529 (2014).

3. X.J. Feng, X. Pan, Parallel Eclat algorithm based on Spark, *Computer Application Research*, **36**, 1-7 (2018).

4. C. Zhang, L.J. Yan, Research and application of improved algorithm of Eclat based on MapReduce, *Journal of Beijing Jiaotong University*, **40**, 1-6 (2016).

5. X.Y. Liu, J.Wang, A method for mining fuzzy association rules based on multi-database, *Computer Applications And Software*, **35**, 42-48 (2018).

6. X.L. Ma, K.D. Zhong, L.C. Liu, Analysis of the effectiveness of students' evaluation of teaching from the perspective of big data mining, *Learning Resource and Technology*, 78-84 (2014).

7. Y.F. Zhang, Z.Y. Xiong, X.F. Yan, J.M. Chen, Analysis and Improvement of Eclat Algorithm, *Computer Engineering*, **47**, 28-30 (2013).

8. H.R. Bai, Constructing a "student-centered" evaluation index system for academic evaluation," *Frontier perspective*, **19**, 4-9 (2018).

9. J.J. Xu, An Improved Association Rules Algorithm, *Computer Knowledge and Technology*, **14**, 257-260 (2018).

10. X.Y. Shi, W.L. Xu, The Reform of College Students' Evaluation of Teaching under the Background of Big Data Era, *HeiLongJiang Higher Education Research*, **256**, 33-35 (2015).

11. X. Yin, Y.M. Xie, Q. Yin, Problems and Reflections on College Students' Evaluation of Teaching --- Taking 22 "985 Universities" as an Example, *Hot Spots*, **4**, 44-48 (2013).

12. P.E. Feng, Y. Liu, Q.Y. Qiu, L.X. Li. Strategies to improve the efficiency of Eclat algorithm, *Journal of Zhejiang University (Engineering Science)*, **47**, 223-230 (2013).

13. K. Kala, Dr. E. Ramaraj, A Novel Approach for Risk Evaluation of Multidimensional Risk Prediction Clustering Algorithm, *International Journal on Computer Science and Engineering (IJCSSE)*, **5**, 894-903 (2013).

14. B. Huang, Research on the Influencing Factors of Students' Evaluation of Teaching --- An Empirical Analysis Based on Sorting Logit / Probit Regression, *Higher Education Research in Finance and Economic*, **17**, 1-8(2014).