

Student Achievement Prediction Based on Factor Analysis and BP Neural Network

Hongchen Guo^{1, a}, Junsheng Liang^{2, b} and Zhiqiang Li^{2, c}

¹Network Information Technology Centre, Beijing Institute of Technology, Beijing, 100081, China

²School of Software, Beijing Institute of Technology, Beijing 100081, China

^a guohongchen@bit.edu.cn, ^b 2120161101@bit.edu.cn, ^c lizq@bit.edu.cn (Corresponding Author)

Keywords: Student Achievement Prediction; Factor Analysis; BP Neural Network

Abstract. The study proposed in this paper analyses the problems existing in the college course score prediction methodologies. Because of diversity and multiple choices in course selection, the analysis of college students' course score can be quite complex. This paper proposes a student achievement prediction method based on factor analysis (FA) and Back-Propagation (BP) neural network. Our method is based on the improvement of FA algorithm. Firstly, special factors will be added to complement the equation of common factor score. Secondly, the initial equation of common factor score will be improved. Thirdly, a new equation intended to give an estimation of the special factors mentioned in the first point will be proposed. Finally, an improvement on the common factor loading matrix will be made. We use the improved equation of common factor score to calculate the score of each common factor. Then we use these scores as the input vector of the BP neural network. The output of the neural network is brought into the final equation to get the final prediction result. The experimental results show that the prediction accuracy is very high and the prediction model can be used for most of the college courses. The error on the prediction result is reduced by using the prediction model proposed in this paper. Therefore, the model developed in this paper is very effective and has high application value.

Introduction

College course score prediction is different from other score prediction [1] [2] [3]. What make it difficult is first the high numerous of courses proposed in a curriculum. Additionally, the composition of the student program can be highly variable: it is very common for a program to be composed of a fixed number of mandatory courses which has to be completed by courses personally chosen by the student. These courses can even be part of a different curriculum than one he chose (a course in another department for example). The highly variable composition of the course program of a student makes the prediction more difficult. This is why a system as complex as FA and BP neural network have to be used.

BP neural network [4] [5] is a multi-layered feedforward neural network trained by error back propagation algorithm. It is a very popular tool in the research about data compression, function approximation, etc. It is a kind of neural network composed of three or more layers of neurons, including the input layer, the middle layer (the hidden layer) and the output layer. The dimensions of the input and output feature vectors are determined by the problem that has to be solved. The neurons in each layer of the BP neural network are independent. Information can only be transferred to neurons in the next layer. The number of neurons in the BP neural network is related to the performance of the neural network. However, one of the difficulties in the application of BP neural network is how to accurately determine the number of hidden layers and the number of neurons.

FA [6] [7] is an important multivariate statistical analysis method, which transforms many feature variables into few synthetic variables. It can make use of several new comprehensive variables to represent a large number of original variables that may have an inherent relationship. FA can be used to analyze the sample data, find the basic structure of the sample data, and then use a small number of new comprehensive variables to represent the basic structure of the sample data. These new comprehensive variables can reflect most of the information contained in the large

number of original variables. We call these new comprehensive variables common factors. Assuming that there are n samples, and each sample contains p variables in the problem we discussed. Each sample can be represented as $X = (x_1, x_2, \dots, x_p)$, and then we can get the formula shown as equation (1).

$$X_i = a_{i1}F_1 + a_{i2}F_2 + \dots + a_{im}F_m + \varepsilon_i \quad (1)$$

This equation can be shown in its matrix form in equation (2).

$$X = AF + \varepsilon \quad (2)$$

Where a_{ij} is factor loading of the matrix A in the i_{th} row and the j_{th} column, express the correlation coefficient between the i_{th} variable and the j_{th} factor, it reflects the i_{th} variable in the importance of the j_{th} factor. ε_i is a part in the i_{th} variable that cannot be explained by common factors, we call it special factor. In this way, we can use a small number of common factors to represent a large number of variables, to be able to reduce the dimension.

The method proposed in this paper cleverly combines the advantages of FA and BP neural network, and do the necessary improvement on the traditional FA method. The improved FA method was used to analyze the college students' course score and combined with the BP neural network algorithm to build a model for the prediction of college students' course score. The experimental results show that the prediction model can accurately predict the future course score of the students, and has a high application value.

Method

Overall Design of Prediction Model

There are many different kind of courses in university. However, all the courses are not independent from each other. In fact, some connections could exist between several courses. This means that some courses which appear to be different could finally be linked, so that the common factors which affect the course score are objective existence. As long as we find the common factors which affect the course score, we can use these common factors to objectively describe the course score of each student.

FA can be used to find out the comprehensive factors which can correctly reflect the main information contained in the original variables, so as to simplify the complex relationship among the variables. Then, the common factor score obtained from scores in all courses can reflect the students' normal state of learning ability and also permit us to effectively eliminate the negative impact on the prediction model (caused by student's abnormal play in one exam).

We extract the common factors that affect the course score by the FA of students' course score data samples. Then, the factor loading matrix and the special factors of each course are also obtained. In order to make a better use of FA, this paper makes an improvement on the traditional FA method. We can calculate the scores of each student in each common factor according to the improved factor score formula. The improved factor score formula is shown as the equation (3).

$$F_i = \frac{\sum_{j=1}^n f(X_j) * (X_j - \varepsilon_j) * a_{ji}^* * 100 / (100 - \varepsilon_j)}{\sum_{j=1}^n f(X_j) * a_{ji}^*} \quad (3)$$

Where F_i is the score of the i_{th} common factor. X_j is the score of the j_{th} course. ε_j is the special factor of the j_{th} course. a_{ji}^* is factor loading of the improved factor loading matrix in the j_{th} row and the i_{th} column. And in order to ensure that only the student has achievements in the course, the course will affect the common factor score, we use the formula as the following equation (4).

$$f(x) = \begin{cases} = 1(x > 0) \\ = 0(x = 0) \end{cases} \quad (4)$$

Then we use the scores of common factor as the input vector of BP neural network and obtain an optimal BP neural network model in the end by constantly adjusting the topology of BP neural network, the number of hidden layers and the number of neurons in each layer.

The input and output data of the prediction model are normalized after data processing and the value of the data is in the [0 - 1]. So the transfer function of BP neural network all use tansig function which can be able to better respond to the output. While the output of BP neural network also needs to carry out anti-normalization.

After that we select the appropriate learning algorithm of BP neural network and build BP neural network model and input training samples for training. Because BP neural network has strong nonlinear mapping ability and self-learning ability, we can finally get a BP neural network model which can accurately predict the results.

There are many factors that determine the score of a student's course, and the common factors can only determine a part of it, even a big part. Therefore, the output of the BP neural network should be the partial score affected by the common factors, otherwise the prediction accuracy of the trained BP neural network will be reduced. Besides the common factors, other factors that determine the score of a student's course include marking criteria which specific to each different teacher, student's psychological state and so forth. The relationship between these factors and the course score is uncertain, so we unify these uncertain factors to the special factor. And we will estimate the special factor's value of each course.

The final output of the prediction model proposed in this paper is the course score which jointly influenced by common factors and that course's special factor. That is, the final output is the result which is calculated by the following final equation (centesimal system).

$$\text{output} = \text{result} * \frac{(100 - \varepsilon)}{100} + \varepsilon \quad (5)$$

Where the result is the output of the BP neural network, ε is the special factor of specific course and it also can be called the partial score which can't be explained by common factors.

The prediction model of this paper is shown as Figure 1.

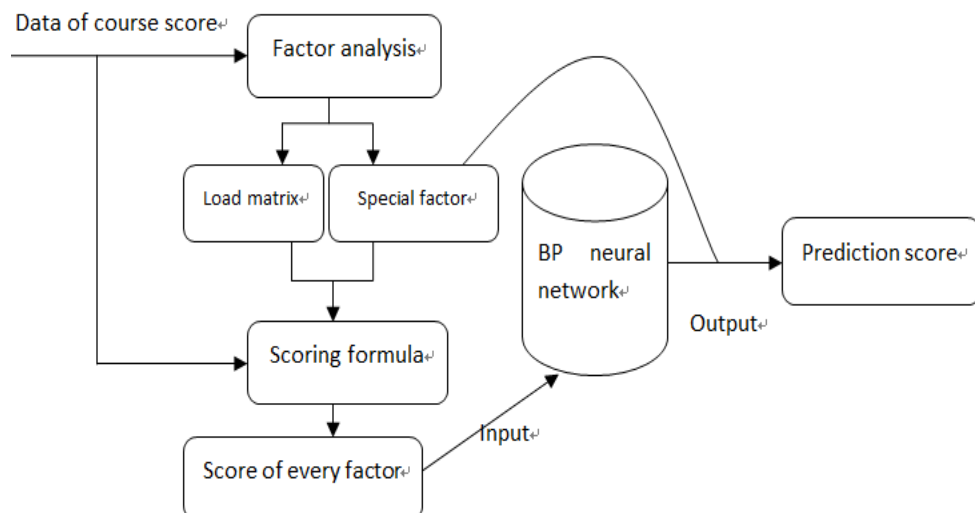


Fig.1. Prediction model

The Improvement of FA

There are several key steps in the process of extracting the common factors and special factors of the course score by using FA. The first step is to select the variables you need to analyze. The second step is to carry on the correlation analysis. The third step is to extract the common factors and use the appropriate method to calculate the loading matrix of the common factors. In this paper, we use the principal axis factoring method. The fourth step is to calculate the special factor value.

The traditional FA method neglects the impact of special factors on the score of common factors, but we can't ignore it, because this impact is huge. If we do ignore it, the prediction result will be greatly affected. Therefore, in this paper, we use the following equation to estimate the value of special factors.

$$\varepsilon_i = \left(1 - \sqrt{h_i^2}\right) * 100 \quad (6)$$

Where ε_i is the special factor of the i_{th} course, and h_i^2 is the quadratic sum of the i_{th} row's factor loading in the matrix.

After the standardization of the data used for FA, the variance of X_i is equal to 1, and the sum of the variance of the common and special factors is equal to variance of that variable. We arrive then to the following equation.

$$1 = h_i^2 + \sigma_i^2 \quad (7)$$

Where σ_i^2 is the variance of the special factor of the i_{th} course.

That is, the contribution of common and special factors to the variables is equal to 1. So using h_i as the degree of interpretation of common factors on variables is credible. And h_i^2 can be calculated according to the factor loading matrix.

In order to make the common factors' loading matrix can fully explain the part that affected by common factors, the sum of the factor loadings in each row of the matrix should be equal to 1. That is, the part that is affected by the common factor is only dependent on the common factor, and the part which is affected by the special factor depends only on the special factor.

Therefore, we improved the spun factor loading matrix to make the sum of factor loadings in each row of the matrix equal to 1. This can be done under the following assumptions: When using the traditional method to obtain the spun factor loading matrix, the impact of special factors on each common factor are average. Therefore, we obtain a formula permitting to improve the spun factor loading matrix, expressed in the following equation.

$$a_{ij}^* = a_{ij} / \sum_{k=1}^m a_{ik} \quad (8)$$

Where a_{ij}^* is the factor loading of the improved factor loading matrix in the i_{th} row and the j_{th} column. a_{ij} is the factor loading of the spun factor loading matrix in the i_{th} row and the j_{th} column. And there are m common factors.

Experiment

Experimental Setting

MYSQL, Microsoft Excel and Eclipse were used to do the data preprocessing. The software we used to do FA and build BP neural network are SPSS Statistics20 and MATLAB R2015b.

The data used in the experiment came from the students following the software engineering program from 2001 to 2016. And all the courses' scores are in a range of 0 to 100 points (centesimal system).

We establish a database of student achievement, and do the data preprocessing through MYSQL. In the end, we get the course information table, student course score table and other data tables. Then we used the sample data of 500-1000 students according to the specific experiments.

We use BP algorithm with adaptive learning speed and momentum gradient-falling (traingdx) as the BP neural network learning algorithm in the experiment. We got an optimal topology structure (contain 2 hidden layers, and the number of neurons in each hidden layer is 4) through changing the number of neurons in each layer (First, Second, Third) and comparing the experimental results. The parameters we use to evaluate experimental results are consist of the prediction result's average magnitude of error (AME), the convergence rate and the minimum value of convergence (MIN of CON).

FA Results

We use SPSS (Statistical Product and Service Solutions) to complete the FA of students' course scores. Its calculation results have a high reliability and reputation in the field of statistical analysis.

The first step of FA procedure is to complete the correlation analysis. Through the experiment we get the KMO (Kaiser-Meyer-Olkin) value is 0.944, which shows that the sample data is very suitable for FA. In the second step, we extract 4 common factors that affect the students' course score, and then use the principal axis factor method to calculate the factor loading matrix of these common factors and spin the factor loading matrix to get the spun factor loading matrix. In the third step, we assume that the impact of special factors on each common factor are average and improve the spun factor loading matrix to make the sum of the factor loading in each row of the matrix equal to 1. Finally, the improved factor loading matrix of some courses is shown in Table 1.

Table 1: The improved factor loading matrix

<i>Course</i>	<i>Factor1</i>	<i>Factor2</i>	<i>Factor3</i>	<i>Factor4</i>
<i>Mathematics for technology I</i>	0.161579	0.512286	0.178704	0.147431
<i>Linear algebra</i>	0.216918	0.504188	0.11474	0.164154
<i>C programming language</i>	0.368263	0.4002	0.03493	0.196607
<i>College English I</i>	0.110704	0.092406	0.294602	0.502287
<i>College English VLS I</i>	0.10026	0.09375	0.00651	0.799479
<i>Principle of Database</i>	0.487027	0.13788	0.26464	0.110452

Some interesting observations can be made through the analysis of the results in Table 1. The courses like C language program design and database design principle are highly affected by factor 1. The courses like Mathematics for technology I and Linear algebra are highly affected by factor 2. The courses like College English are highly affected by factor 4. And through the FA of students' score in more courses, we found that this kind of law is universal in college courses. That is why we analyzed the results of FA and named factor 1 as professional ability factor, named factor 2 as logical thinking factor, named factor 3 as class performance factor and named factor 4 as communication ability factor.

Then we use the equation (6) to calculate the special factor of each course. Finally, we got the proportion influenced by common factors and the proportion influenced by the special factor in each course (centesimal system).

The result of some courses is shown in Table 2.

Table 2: Course score distribution

<i>Course</i>	<i>Common factors</i>	<i>Special factor</i>
<i>Mathematics for technology I</i>	79	21
<i>Linear algebra</i>	70	30
<i>C programming language</i>	58	42
<i>College English I</i>	66	34
<i>College English VLS I</i>	62	38
<i>Principle of Database</i>	78	22

The Influence of Special Factor on Prediction Model

The final output of the prediction model proposed in this paper is the result that influenced by 4 common factors' score and the specific course's special factor. The calculation formula has been shown in equation (5). The prediction model without considering the influence of special factors will use the output of the BP neural network as the final output. We compare the effects of these two models through the following experiments.

This experiment uses 900 data samples to train these 2 prediction models, and then compares the prediction results with 100 test samples. All the experiments in this paper, the index used in AME and prediction error are centesimal system.

The result of this experiment is shown in Table 3.

Table 3: Contrast of experimental results

Number	AME	Convergence rate	MIN of CON
1	2.9087	23s	0.0143
2	4.6035	25s	0.0229

The result of group 1 is the result of the prediction model proposed in this paper, and the result of group 2 is the result of the prediction model without considering the influence of special factors. And the comparison between the predicted results and the real results of the 100 test samples is show as Figure 2.

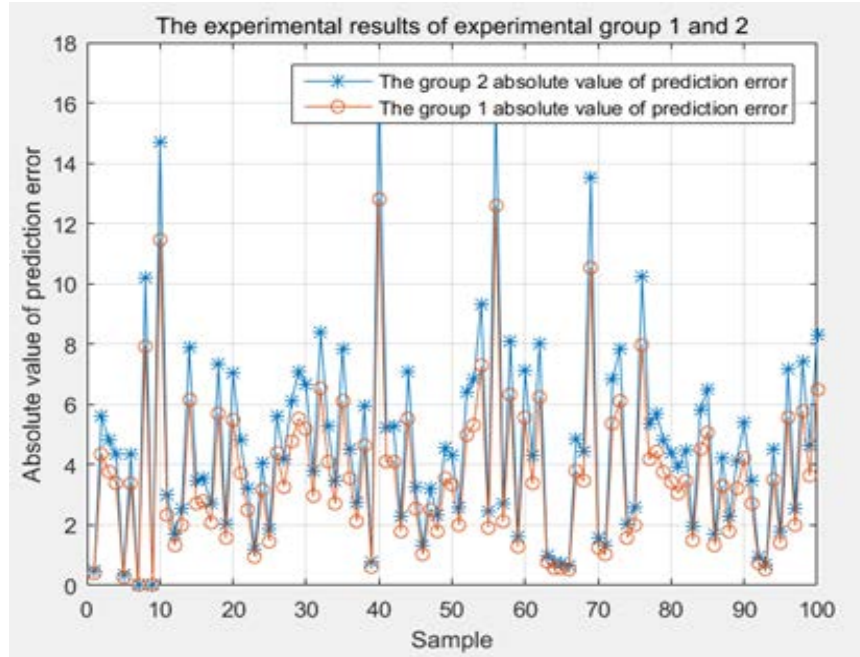


Fig.2. Contrast of prediction error between group 1 and 2

From Figure 2, we can find that, for each sample, the results obtained by considering the special factor's influence can effectively reduce the prediction error. And the average prediction error is reduced from 4.6035 to 2.9087. As it can be seen, the prediction accuracy of the model proposed in this paper has been greatly improved by considering the special factor's influence.

Comparison with Baseline

In this experiment, we will compare our method with the Association Rules (AR) method. The most famous AR mining algorithm is Apriori algorithm [8] [9]. We will use Apriori algorithm to select a number of the most relevant courses and use these course scores set as the input vector of BP neural network and train it.

Compared with AR method, the prediction model proposed in this paper is more adaptable. AR method needs the multidimensional data with a unified data format and complete vector. The prediction model proposed in this paper improved the scoring formula and it can calculate the score of every common factor according to the specific situation of the multidimensional data vector.

AR method can't deal well with certain situations, for example if some students did not take a certain course and do not have a score of a course. Consequently, the experiment had to use the data of scores of 500 software engineering students in 30 different courses with each student having scores in all the 30 courses. 400 samples were selected as training samples, and the remaining 100 samples were used as test samples. The results of the experiment are shown in Table 4.

Table 4: Contrast of experimental results between our method and Apriori method

Number	Method	Input dimension	AME	Convergence rate	MIN of CON
1	Our method	4	4.5721	19s	0.0277
2	Apriori	4	6.2990	20s	0.0322
3	Apriori	6	5.1100	20s	0.0339
4	Apriori	8	6.0143	20s	0.0337

And the 100 test samples prediction result's comparison between experimental group 1 and 2 are shown as Figure 3.

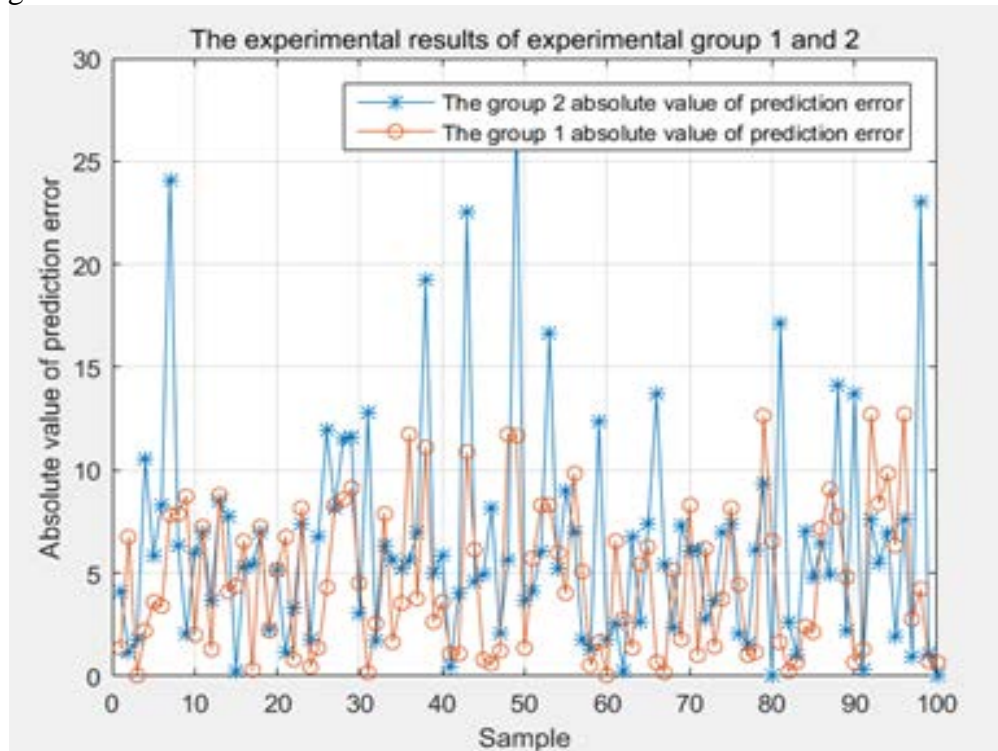


Fig.3. Contrast of prediction error between our method and Apriori method

The experimental results show that the prediction model proposed in this paper has higher prediction accuracy than the prediction model based on AR and BP neural network, and can be applied to a wider range. In Figure 3, there are a lot of test sample's prediction results which are far from the true value in the prediction model based on AR and BP neural network. And most of the test sample's predicted deviation of the prediction model proposed in this paper can be kept within 10 points. This shows that the prediction model based on FA and BP neural network is not only accurate, but also more stable than other models.

Generalization Ability of Prediction Model

In order to prove that the prediction model proposed in this paper can well be used for other various courses. We use the student's score in courses of the first three semesters to calculate the score of every common factor and use these 4 common factor scores and the specific course's special factor to predict the student's score in courses of the fourth semester. There are 22 courses in the first three semesters. Finally, we use 900 training samples to train the final prediction model, and get the different prediction model of several specific courses, and use 100 test samples to test the performance of the model. The result of this experiment is shown in Table 5.

It can be found from Table 5, for each course, the prediction model proposed in this paper has a good prediction power. According to the prediction model of Software Engineering Foundation, the average value of the prediction error is lower than 3 points, which has a high value of application.

Table 5: The experimental results of the specific prediction model for each course

Number	Predicted course	AME	Convergence rate	MIN of CON
1	computer network	5.6377	25s	0.0140
2	Principle of Database	3.3092	25s	0.0156
3	Fundamentals of Digital Electronics	3.3733	23s	0.0093
4	Software Engineering Foundation	2.9990	25s	0.0153

Conclusion

Research on the prediction of student achievement performance has made remarkable achievements, and has been successfully applied to the field of education. This paper analyses the

problems existing in the college student course score prediction and puts forward a new idea about the student achievement prediction through the research and study of data mining, neural network, statistical analysis and other related fields. In this paper, the prediction model based on improved FA and BP neural network are introduced. This model combines the advantage of the traditional prediction model based on statistical analysis method with those of the prediction model based on nonlinear theory.

Overall, the prediction model proposed in this paper achieves a desired goal. The prediction model proposed in this paper can accurately predict the student's course score in the future on the basis of a single attribute of students' scores in their history. The average value of the prediction error reduced about 40% than other prediction model, and it has a high value of application.

References

- [1] Cripps A. Using artificial neural nets to predict academic performance[C]//Proceedings of the 1996 ACM symposium on Applied Computing. ACM, 1996: 33-37.
- [2] Kalles D, Pierrakeas C. Analyzing student performance in distance learning with genetic algorithms and decision trees[J]. Applied Artificial Intelligence, 2006, 20(8): 655-674.
- [3] Kotsiantis S, Pierrakeas C, Pintelas P. PREDICTING STUDENTS' PERFORMANCE IN DISTANCE LEARNING USING MACHINE LEARNING TECHNIQUES[J]. Applied Artificial Intelligence, 2004, 18(5): 411-426.
- [4] Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators[J]. Neural networks, 1989, 2(5): 359-366.
- [5] Yi J, Wang Q, Zhao D, et al. BP neural network prediction-based variable-period sampling approach for networked control systems[J]. Applied Mathematics and Computation, 2007, 185(2): 976-988.
- [6] Spearman C. "General Intelligence," objectively determined and measured[J]. The American Journal of Psychology, 1904, 15(2): 201-292.
- [7] Wang C, WANG J. Comprehensive Evaluation on the Marks of Statistical Practices Graduates Based on Factor Analysis and Cluster Analysis [J][J]. Journal of Gansu Lianhe University (Natural Science Edition), 2011, 1: 005.
- [8] Borgelt C, Kruse R. Induction of association rules: Apriori implementation[C]//Compstat. Physica-Verlag HD, 2002: 395-400.
- [9] Agrawal R, Srikant R. Fast algorithms for mining association rules[C]//Proc. 20th int. conf. very large data bases, VLDB. 1994, 1215: 487-499.