

## Design of student information system based on association algorithm and data mining technology

CaiYan, ChenHua

Jiangxi Environmental Engineering Vocational College, Ganzhou 341000, China

**Keywords:** association algorithm; data mining; information system; GUI interface; confidence

**Abstract:** With the advent of the information age, students in the day-to-day management of information will inevitably encounter some data processing problems. When the amount of data processing is large, it is difficult to meet the needs of calculation if the manual method is adopted. Accordingly, this paper proposes a data mining algorithm based on Association and clustering, and uses it in the student information analysis system, which effectively improves the efficiency of data mining. In order to improve the ease of operation of the system, the GUI operation interface is designed by using Matlab software, and the data association analysis of student achievement can be achieved by simple button operation. Finally, the feasibility of the system is verified, and tested the feasibility of data analysis in different confidence conditions, the test results indicate that the clustering analysis using the analysis system of student performance can be achieved successfully. The data were correlated and results, is a kind of efficient student information analysis system.

### Introduction

In recent years, big data applications have been integrated into people's daily life and work, has become a very popular word, which marks the data mining has been integrated into all aspects of our life, with the development of computer and cloud platform technology, we have entered the era of information explosion. In the era of information explosion today, thousands of times every day of our contact information will be even more than the ancient people a day contact information, data information every day around the world are in exponential growth, in order to efficiently use the data information, data mining algorithms emerge. But it requires the aid of interface design how to use these data algorithms and how to get the ordinary staff to do simple data mining algorithms. Using a simple interface, even people who do not understand the algorithm can easily achieve massive data mining.

### Association and clustering data mining algorithms

Correlation algorithm is based on data mining to analyze the relevance of the data itself, in the actual data analysis, the first is the simple statistics of all items containing a set of elements of the frequency to determine the largest project in one dimension. Then the distribution is calculated to get the support of the project set, in which the data can be listed on the record table. This time, with a simple data record table as an example, as shown in Table 1, it is divided into four data records.

Table 1 data sheet

TID	<i>Items</i>
T1	21,23,24
T2	22,23,25
T3	21,22,23,25
T4	22,25

As shown in Table 1, in the correlation algorithm, we need to create each data record candidate

support, and by setting the confidence associated comparison data records to get the data correlation, the data can be classified, the use of the clustering algorithm, the procedure is as follows.

(1) Determine the clustering center, you can choose  $k$  centers, specifically for  $Z_1^1, Z_2^1, \wedge Z_k^1$

(2) The data  $\{X\}$  will need to be classified with certain rules, assign  $Z_j^{(1)}$  to the cluster center.

(3) Calculate the new vector values of each cluster center  $Z_j^{(k+1)}$ ,  $j=1,2,\dots,K$ , the expression is as follows

$$Z_j^{(k+1)} = \frac{1}{N_j} \sum_{X \in S_j^{(K)}} X \quad (1)$$

$N_j$  is the number of samples contained in the  $j$ -th clustering domain  $S_j$ , after finding the mean vector, it is used as the clustering center to calculate the mean, and the function rules are as follows

$$J = \sum_{j=1}^K \sum_{X \in S_j^{(K)}} \|X - Z_j^{(k+1)}\|^2 \quad (2)$$

The data of each cluster is calculated by means of the mean value.

(4) If  $Z_j^{(k+1)} \neq Z_j^{(k)}$ ,  $j=1,2,\dots,K$ , then, second steps are returned. The algorithm is used for cluster analysis and iterative calculation is performed. If equal, the algorithm is convergent and the calculation is over.

### **Student information system interface design**

In order to realize the function of the algorithm in the first section, the Matlab data simulation software is adopted, and the graphical design method is adopted to realize the functional interface of the student information system. There are two ways in which Matlab graphical interface design is enabled, one is directly select the GUIDE button in the Matlab menu, as shown in Figure 2, is a window in the Matlab programming input GUIDE command, and then will pop up the window shown in Figure 2.

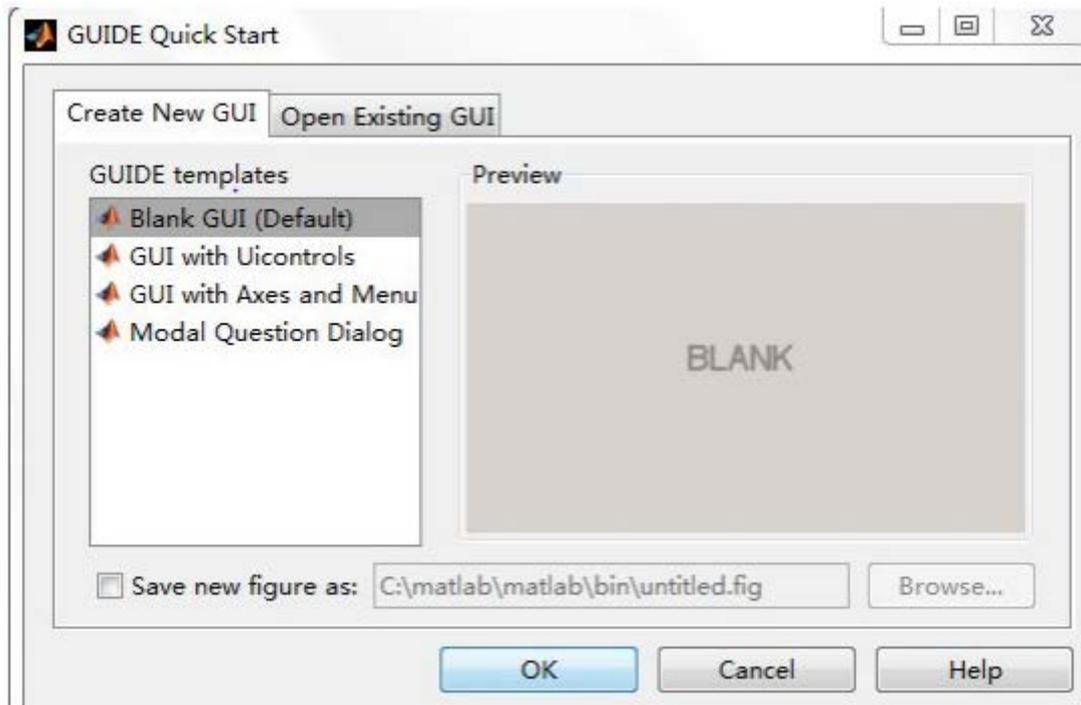


Fig.1 GUI interface design

Select the first item in the pop-up window, Blank GUI options, click “OK” to enter the graphical user interface layout editing interface to design the required system interface, the design of the interface has three main function buttons, the specific function of each button can function as follows.

- (1) Input confidence, the function button can input different confidence values according to the user's needs. According to the confidence value, the correlation analysis of achievement is achieved.
- (2) The performance related clustering analysis, the function button can achieve the results of cluster analysis and data statistics, resulting in the relevance of the results of analysis.
- (3) Output results, the function button is to display the analysis results in the form of image data, so as to improve the visualization of data analysis.

### Mining and analysis of student information data

This analysis divides the student's information into several groups according to several categories, the relevance of information can be discovered mainly through association rules.

$X\{1\}=\{\text{'scholarship' 'student origin' }\};$

$X\{2\}=\{\text{'scholarship' 'students origin' 'comprehensive quality' 'the proportion of men and women' }\};$

$X\{3\}=\{\text{'scholarship' }\};$

$X\{4\}=\{\text{'scholarship' 'comprehensive quality' 'students origin' }\};$

$X\{5\}=\{\text{'scholarship' 'comprehensive quality' 'proportion of men and women' }\};$

$X\{6\}=\{\text{'scholarship' 'students origin' 'experimental ability' 'the proportion of men and women' }\};$

When the confidence level is  $k = 0.5000$ , the output correlation results are:

Output association results

'scholarship' 'student origin'

'scholarships' 'the proportion of men and women'

'scholarship' 'comprehensive quality'

'the proportion of men and women' 'comprehensive quality'

When the confidence level is  $k = 0.6000$ , the output correlation results are:

Output association results

'scholarship' 'student origin'

'scholarships' 'male and female ratio'

'scholarship' 'comprehensive quality'

From the above results, we can see that the correlation increases with the increase of confidence. Finally, the clustering analysis of academic achievement, this cluster uses the first section of the clustering algorithm, the  $k$  is set to 3, that is divided into three categories, then Figure 2 can be obtained.

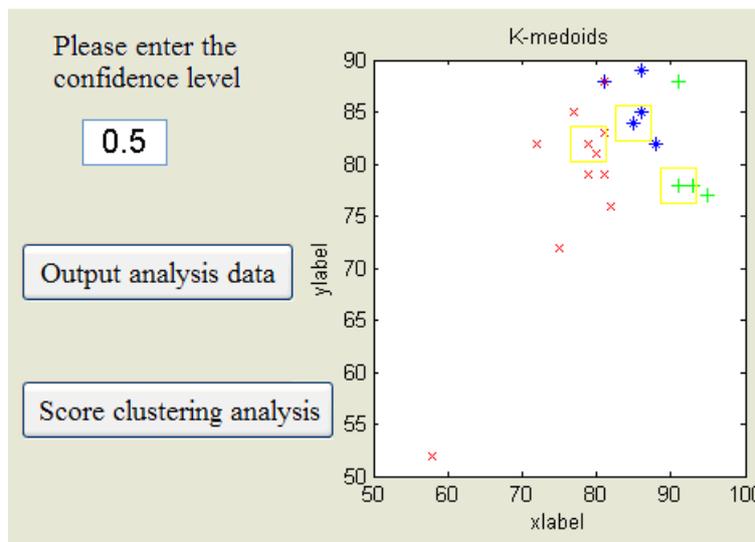


Fig.2 The visualization result of student achievement clustering analysis

This simulation mainly carries on the cluster analysis to 800 students' achievement of 20 classes, from the analysis results, we can see that the results are concentrated in more than 80 points, and some have failed, but very few, so as to verify the system data analysis and visual display function.

**Conclusion**

In order to achieve student information data analysis and processing, an algorithm for mining association and clustering data based on student achievement has been presented in this paper, and the design is easy to operate by using the Matlab software of the GUI interface, which effectively improves the efficiency of data analysis and visualization. In order to verify the reliability of the system, taking the relevance score system as an example, the relationship between student scholarships and students origin, comprehensive quality, experimental ability and the proportion of men and women has been analyzed, through the use of the achievement data mining, clustering algorithm has a visualization of the results of the cluster analysis. The test results indicate that the analysis system can effectively get the relevance of the scholarship, and can display the clustering classification and visualization, is a very efficient student information mining system.

**Reference**

[1] He Yao, Wang Wenqing, XueFei. Research on massive data mining based on cloud computing [J]. computer technology and development, 2013, 23 (2): 69-72.  
 [2] Ding Yan, Yang Qingping, Qian Yuming. Platform architecture and its key technology research of [J]. ZTE cloud based data mining technology, 2013,19 (1): 53-60.

- [3] Li Kai, Chang Zheng. Design and implementation of parallel data mining system based on cloud computing [J]. *microcomputer information*, 2011, 27 (6): 121-123.
- [4] Sun Guang Lu, Qi Haoliang. *Journal of Tsinghua University*[J]. online scheduling of spam filtering based on logistic regression 2013, 53 (5): 734-740.
- [5] Liu Botao. Research on data mining algorithms based on rough sets [J]. *Western China Science and technology*, 2011, 10 (14): 11-12.
- [6] Hu Jiance, Wu Guoping. Improved genetic BP neural network data mining algorithms and applications [J]. *microcomputer and applications*, 2011, 2: 25-30 .
- [7] ChuBin, Wu Chen, Yang Xibei. Algorithm [J]. *computer technology and development of RBF neural network and data mining based on rough set*, 2013,23 (7): 87-91.
- [8] Sun Jigui, Liu Jie, Zhao Lianyu. Study on clustering algorithms [J]. *Journal of software*, 2008, 19 (1): 48-61.
- [9] Ceng Donghai, m, Hong, Liu Lifeng. A clustering algorithm based on mesh density and space partitioning tree [J]. *systems engineering theory and practice*, 2008, 28 (7): 125-133.
- [10] Li Guangqiang, Deng Min, Liu Qiliang, et al. A spatial clustering method adapted to local density change [J]. *Journal of Surveying and mapping*, 2009, 38 (3): 255-263.
- [11] Liu Ming, Wang Xiaolong, Liu Yuanchao. A fast clustering algorithm for large-scale high-dimensional data [J]. *Journal of automation*, 2009, 35 (7): 859-866.
- [12] Xu Li, Ding Shifei. Study on the granularity of [J]. *clustering algorithm in computer science*, 2011, 38 (8): 25-29.
- [13] Zhou Tao, Zhang Yanning, Yuan Hejin, et al. Rough kernel k-means clustering algorithm [J]. *Journal of system simulation*, 2008, 20 (4): 921-925.