

## A Multi-weight Adaptive Analysis Method of Students' Learning Behavior

Enqiang Shen<sup>1, a</sup>, Ran Wei<sup>1, b</sup>, Jin Ding<sup>1, c</sup>, Quanyin Zhu<sup>1, d\*</sup>

<sup>1</sup>Faculty of Computer and Software Engineering Huaiyin Institute of Technology, Huaian, 223005, China

<sup>a</sup>email:15950376029@163.com, <sup>b</sup>email:weiran@hyit.edu.cn, <sup>c</sup>email:hyicdj@163.com, <sup>d</sup>email:hyitzqy@126.com

**Keywords:** Data mining; K-means algorithm; Behavior analysis

**Abstract.** With the development of Internet information technology, all aspects of society are in development to digital, information-oriented. Student management system is widely used in Colleges and Universities. Student management system contains a lot of valuable data, and has not been fully excavated. This article is aimed at the university student attendance system. An algorithm of students' learning behavior analysis is designed. In the experiment, the number of the students is 423. Students are divided into four categories. The numbers of students in the four categories were: 7, 267, 112 and 24 respectively. The average scores of all the students in each category were: 107, 113, 117 and 123 respectively. This algorithm can effectively analyze students' learning behavior and provide effective support for students' management.

### Introduction

Nowadays, student management system is widely used in Colleges and Universities. In the era of Big Data, a large number of potential value in the systems have not been explored. Data mining is a technology that searching for hidden information from large number of data. Data mining technology [1] can dig out the potential value of Student management system.

The students' behavior can be analyzed by using the data of the students' system. Students' behavior can be divided into many kinds, including learning behavior [2], consumer behavior and so on. Understanding the behavior of the students is good for improving the quality of teaching and strengthening the management of students.

Clustering algorithm [3] is a key technology of behavior analysis [4]. Clustering algorithm can be used to gather students into different categories, each category of students have similar behavior habits. Commonly used clustering algorithms are: partitioning clustering [5], hierarchical clustering [6], neural network clustering [7], etc. K-means [8] clustering is a kind of partition clustering algorithm, it has the advantages of fast and simple. K-means has higher efficiency and better scalability for large data sets.

In this paper, a method of analysis of students' learning behavior is proposed. This method can divide the students into four classes, each of which has different learning behaviors.

### Data Preprocessing

The experimental data of this paper comes from the student attendance system of the school. Among them, including student morning exercises attendance data, class attendance data, evening class attendance data and late bedtime attendance data. Select students' morning exercises normal attendance rate, morning exercises non normal attendance rate, class attendance rate, classroom non normal attendance rate, evening class attendance normal attendance rate, evening class non normal attendance rate, late bedtime attendance rate to construct the feature vector of students, the feature vectors of each student has seven dimensions.

### Design Cluster Analysis

First of all, set up the attendance data set Records according to the Equation(1),  $Stu_i$  has seven

dimensions,  $Stu_i = \{s_1, s_2, \dots, s_7\}$ ,  $s_1, s_2, \dots, s_7$  stand for different attendance rates which have been mentioned in Data preprocessing. Simultaneously, define the clustering data set according to the Equation(2),  $U_1, U_2, \dots, U_K$  is the different cluster group and  $K$  is the number of cluster centers. Define the evaluation data set according to the Equation (3),  $(wei_z, v_z)$  represents weight subset  $wei_z$  and its corresponding evaluation value  $v_z$ .

$$Records = \{Stu_1, Stu_2, \dots, Stu_M\} \quad (1)$$

$$Urecord = \{U_1, U_2, \dots, U_K\} \quad (2)$$

$$VALUE = \{(wei_1, v_1), (wei_2, v_2), \dots, (wei_L, v_L)\} \quad (3)$$

Then, set up the weight data set according the Equation(4),  $wei_j$  is the subset of  $W$ ,  $wei_j = \{wa_j, wb_j, wc_j, wd_j, we_j, wf_j, wg_j\}$ .

$$W = \{wei_1, wei_2, \dots, wei_L\} \quad (4)$$

Using a cycle traversal the weight set  $W$ , then, using each subset of  $W$  to add weight to  $Records$  and get the weighted attendance dataset  $Wrecords$  according Equation(5), then, using hierarchical clustering to determine the number of cluster centers  $K$ . After that, using  $K$ -means clustering algorithm for clustering  $Wrecords$  and get the clustering Results data set  $Urecord$ . Evaluating the  $Urecord$  according to the Equation(6) and get the evaluation value  $v_z$ . The function of  $GetCenter(Stu_i)$  is to achieve the clustering center by  $Stu_i$ . Then, add  $(wei_z, v_z)$  to the data set  $VALUE$ .

$$Wrecords = \{Stu_1 * wei_j, Stu_2 * wei_j, \dots, Stu_M * wei_j\} \quad (5)$$

$$v_z = \sum_{i=1}^M (Stu_i - GetCenter(Stu_i))^2 \quad (6)$$

Define a function  $Min(VALUE)$  to get the minimum value  $v_{min}$  from the  $VALUE$  data set. Then get the corresponding weight  $wei_{fin}$  by  $v_{min}$ . Using  $wei_{fin}$  to add weights to  $Records$  data set and get the final weighted data set  $FWrecord$  according to the Equation(7). After that, using  $K$ -means clustering algorithm for clustering  $FWrecord$  and then get data set  $FUrecord$  according to the Equation(8),  $F_1, F_2, \dots, F_K$  represent different cluster groups.

$$FWrecord = \{FStu_1, FStu_2, \dots, FStu_M\} \quad (7)$$

$$FUrecord = \{F_1, F_2, \dots, F_K\} \quad (8)$$

Define student achievement data set  $FGrade$  according to the Equation(9),  $g_1, g_2, \dots, g_M$  respectively represent different student's achievement. Following, calculating the average achievement of each cluster group and get the data set  $FGrade$ , The total number of cluster group is  $K$ .

$$FGrade = \{ag_1, ag_2, \dots, ag_K\} \quad (9)$$

Finally, using PCA reduction algorithm to reduce the  $FUrecord$  to two-dimensional data, and display in the two-dimensional plane, then, analyzing the relationship between students' learning behavior and students' achievement based on  $FGrade$  data set and  $FUrecord$  data set.

## Test Results

Fig.1 is the result of experiment that don't add any weight to the  $Records$  data set, we can see that students are divided into four categories but the four categories of students have not been clearly divided, four categories of students have some areas of overlap together, clustering effect is not very good.

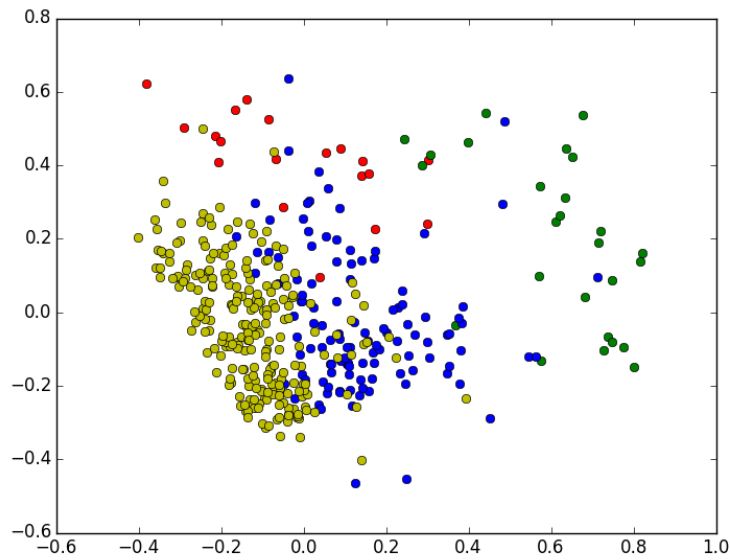


Fig.1. Clustering result1 (No weight added)

Table 1 Average Achievement of each Category (No weight added)

Category	Category1	Category2	Category3	Category4
Color	Blue	Yellow	Green	Red
Average Achievement	113	114	118	123

Table 1 shows the average achievement of each category without adding any weight, we can see from table 1, the gap between the scores of each category is not big enough, especially, category 1 and category 2 differ only 1 point.

Fig.2 is the result of the experiment that add a proper weight to the attendance data set Records, compared with figure 1, the effect is obviously improved, we can see from the figure 2, students are clearly divided into 4 categories.

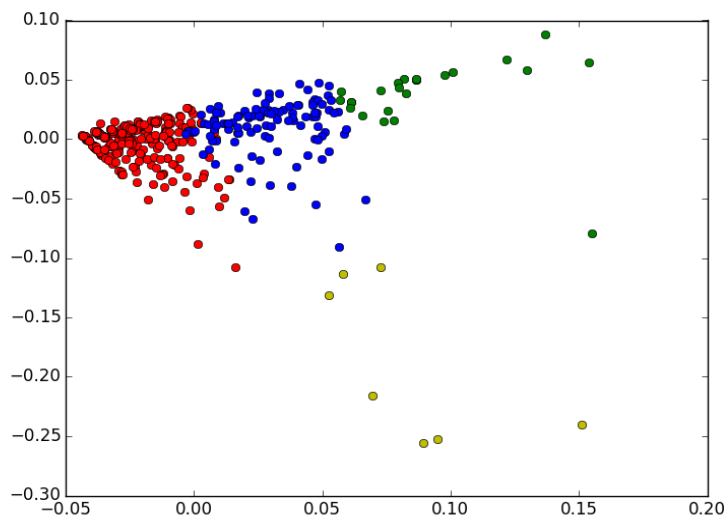


Fig.2. Clustering result2 (proper weight added)

Table 2 Average Achievement of each Category (proper weight added)

Category	Category1	Category2	Category3	Category4
Color	Yellow	Red	Blue	Green
Average Achievement	107	113	117	123

Table 2 show us the result of experiment with proper weight added. Each category of students has a different average achievement. Compared with table 1, the result is significantly improved.

## **Conclusion**

This paper puts forward a method of analysis of students' learning behavior. The experiment data comes from the attendance system which is using in our institution. First, the data set is established by data cleaning, normalization and so on. Then, the program use to calculate the best weight according to the principle of minimum square error sum. After that, using the best weight to add weight to attendance data and using K-mean algorithm to get the clustering result. Finally, analyses students' learning behavior with students' achievement.

## **Acknowledgement**

This work is supported by the Key Research Plan of Jiangsu Province, China (BE2015127), the University Science Research Project of Jiangsu Province (15KJB520004), the Science and Technology Projects of Huaian (HAG2015060, HAG2014028), the Project of National Undergraduates Innovation under the Grant No.201511049013Z and Scientific Foundation Project of Huaiyin Institute of Technology (HGC1412).

## **Corresponding Author**

Quanyin Zhu, Faculty of Computer and Software Engineering, Huaiyin Institute of Technology, Huaian, 223005, China.

## **References**

- [1] Ren Yanna, The design of algorithm for data mining system used for Web Service, , IEEE 3rd International Conference on Communication Software and Networks [C], 2011, 480 - 483.
- [2] Tsung-Ho Liang, Kun-Te Wang and Yueh-Min Hung, An exploration study on student online learning behavior patterns, ITME IEEE International Symposium on IT in Medicine and Education [C], 2008, 854 – 859.
- [3] Fuge, Mark, B. Peters, and A. Agogino, Machine Learning Algorithms for Recommending Design Methods [J], Journal of Mechanical Design, 2014, 470 - 495.
- [4] He, Gang, Y. Deng, and X. Wu, Analysis of web-surfing behavior of enterprise network users, International Conference on Computer Science and Network Technology [C], 2012, 916 - 920.
- [5] G. Singh, J. Kaur and Y. Mulge, Performance evaluation of enhanced hierarchical and partitioning based clustering algorithm (EPBCA) in data mining, International Conference on Applied and Theoretical Computing and Communication Technology [C], 2015, 805 - 810.
- [6] Bouguettaya, Athman, et al, Efficient agglomerative hierarchical clustering, Expert Systems with Applications [J], 2015, 2785 - 2797.
- [7] Chen, Lifang, X. Lu, and Z. Du, RBF neural network modeling based on PCA clustering analysis, IEEE International Conference on Granular Computing [C], 2014,35 - 38.
- [8] Wu, Guohua, et al, An Improved K-means Algorithm for Document Clustering, International Conference on Computer Science and Mechanical Automation [C], 2015, 65 - 69.