Clustering Analysis of Sports Performance based on Ant Colony Algorithm

Wang Jian1，Hong Zhi-hua2，Zhou Zhi-yong3

(The School of Humanities and Social Sciences, Jingdezhen Ceramic Institute, Jingdezhen, 333403, China)
wangjian2014_cn@126.com

***Abstract*—Cluster analysis is one of the modes of data mining, which classifies the sample data to different types according to similarity rules. It has also been used in education management field. This paper investigates the principle of k-means clustering algorithm. Because it is easy to converge into local minimum and is also sensitive to noise, isolated point data have a great impact on the average value, an improved clustering algorithm based on ant colony optimization is proposed. The improved algorithm is used in student sports performance management system. It can be concluded that clustering results obtained by the improved algorithm based on ant colony optimization is more scientific, fair and reasonable.**

Keywords-Siphonic Bedpan;Flow Channel Structure; Numerical Simulation

I. INTRODUCTION

The evaluation of the student is a vital factor in the process of teaching, which consists of many elements, among which the evaluation of grades is the most visual[1,2]. However, on account of the one-sidedness of grades, they can hardly reflect the learning state of a student.Thus, the essay makes use of an advanced method to deal with grades of the student, in which case, the student's learning state within a certain period can be reflected effectively[3].

Data mining, which based on specific algorithms, is a technology for researchers to analyze large amounts of data to find potential knowledge, and widely used in the major fields, makes itself be valuable[4,5]. Cluster analysis is one of the modes of data mining, which classifies the sample data to different types according to similarity rules, so that in the same type data have high resemblance, different type data become high differential[6-9]. By clustering,it's easy to describe the characteristic of different types.

Swarm intelligence is the interdisciplinary of biology, computer technology and other fields[10-14]. It means in the nature, due to the interaction and cooperation, the single individual with no intelligent or less intelligent to finalize the complex intelligence mission. In the 1990s, the Italian scholar M. Dorigo, V. Maniezzo, A. Colorni were enlightened by the behavior of ants which could find the shortest path between their home and food source to propose the most important algorithm of swarm intelligence ant colony algorithm. They use it successfully to solve many combinatorial optimization problems such as traveling salesman TSP, graph coloring problems and so on.

Cluster analysis based on the ant colony algorithm is a hot research in recent years,since it don't need the pre-conditions to cluster, which is the requirement of traditionalclustering algorithm. This paper achieves two improvements for ant colony clustering algorithm which proposed by Lumer and Faieta. One is to record the data object that visited by the ant so that can ensure all data will be visited by ants and prevent the same data picking up and putting down repeatedly; the second is to establish a "historical position of global memory, which helps ants to put down the data object faster with target as well as the emergence of local optimal solution. Through the experimental comparative analysis prove that the improved algorithm has better performance and clustering result than the older one.

Finally, the improved algorithm is applied to mining the evaluation of student performance, including the comprehensive evaluation and single subject level evaluation. By comprehensive evaluation it's easy to find out the difference of students in different type, which used to support teaching, and it is valuable for teachers and students to improve their work. And for the other one evaluation, the clustering can relieve the insufficient of traditional grades which lack of flexibility to change with the change of external influences. Finally, according to the compare results of traditional grade and the dynamic grade confirm that dynamic grade is reasonable and practicable. Cluster analysis of the two models,indicating that data mining used in the teaching and learning environment is meaningful.

The paper is organized as follows. In the next section, clustering algorithm based on K-means is given. In Section 3, the improved clustering algorithm based on ant colony optimization is proposed. In Section 4, in order to test the performance of the algorithm, this paper adopts student sports performance in some university as experiment object. Finally, we conclude our paper in section 5.

II. CLUSTERING ANALYSIS

K-means is one of the simplest unsupervised learning algorithms, which takes a simple and easy way to classify a given data set through a certain number of clusters. For given data set, k-means algorithm requires the user to enter the user expected classification number k and it choose k number of points as initial clustering center according to parameter k randomly. Usually Euclidean distance is used as the measure to evaluate similarity and the objective function to evaluate the quality of the division is defined as

978-1-4799-4261-9/14 $31.00 © 2014 IEEE
DOI 10.1109/ISDEA.2014.71

$$J = \sum_{i=1}^{k} \sum_{j=1}^{c_i} d_{ij}(x_j, z_i)^2 .$$

z_i is clustering center of c_i, x_j is data point of cluster c_i. $d_{ij}(x_j, z_i)$ represents the distance between x_j and z_i. Objective function J is the sum of distance between each data point and the cluster's center. The smaller J means better clustering effect. K-means algorithm is to find good clustering solution through continuous optimization value of J. When J takes minimum value, clustering center will no longer change and clustering stops.

The process of K-means algorithm is as follows. Input is data set $S = \{x_1, x_2, \ldots, x_n\}$ and the number of clustering is k. Output is k number of clustering.

Step1. Choose k number of initial clustering center $z_1, z_2, \ldots, z_k$ from data set $S = \{x_1, x_2, \ldots, x_n\}$ randomly.

Step2. Cluster data set according to initial clustering center $z_1, z_2, \ldots, z_k$ and obtain k number of class $\{C_1, C_2, \ldots C_k\}$. The determination of C_i is as follows.

For any data point $x_i \in S$, if $d(x_i, z_j) < d(x_i, z_m)$, x_i belongs to C_i and $m = 1, 2, \ldots, k, m \neq j$.

Step3. According to (2) adjust center of cluster $z_1^*, z_2^*, \ldots, z_k^*$. n_i is the number of sample included by cluster C_i.

$$z_i^* = \frac{1}{n_i} \sum_{x_i \in C_i} x_i .$$

Step4. Determine whether iteration stops. When objective function J converges, the algorithm stops and output optimal clustering result $\{C_1, C_2, \ldots C_k\}$. Otherwise $z_i = z_i^*$, return to step 2.

Step5. Output clustering result and algorithm stops.

The following will be showed by an example of the working process of the k-means algorithm.

Given data set is $S = \{2, 7, 25, 21, 13, 27, 30, 20, 16, 9\}$. Choose three initial center $z_1 = 2, z_2 = 16, z_3 = 30$ randomly. Calculate the distance between sample point and three sample center. Because $|7-2| < |7-16| < |7-30|$, so 7 is divided into class C_1 that z_1 belongs to. 9, 13, 18, 22 are divided into class C_2 that z_2 belongs to. 24, 27 are divided into class C_3 that z_3 belongs to. Recalculate clustering center of each class according to classified data objects.

$$z_1' = \frac{2+7+9}{3} = 6,\ z_2' = \frac{9+13+16+18+22}{5} = 15.6,$$

$$z_3' = \frac{24+27+30}{3} = 27 .$$

Then data set is divides again using the new clustering center and the result are $C_1 = \{2, 7, 9\}$, $C_2 = \{13, 16, 18\}$, $C_3 = \{22, 24, 27, 30\}$ and recalculate clustering center.

$$z_1' = \frac{2+7+9}{3} = 6, z_2' = \frac{13+16+18}{3} = 15.67,$$

$$z_3' = \frac{22+24+27+30}{4} = 25.75 .$$

After the third iteration, results are the same with results of the second iteration. The optimal three clustering are $C_1 = \{2, 7, 9\}$, $C_2 = \{13, 16, 18\}$, $C_3 = \{22, 24, 27, 30\}$.

III. THE IMPROVED CLUSTERING SCHEME BASED ON ANT COLONY ALGORITHM

Initialize each ant in the ant colony, the number of ant in the ant colony is N. The maximum iteration times of ant colony algorithm is M. The length of local area is s. All the clustering objects are mapped into a given plane. That is each data object is assigned to a point of two dimension plane. The location of coordinate is (x, y). Choose a data object in the plane randomly. Calculate average similarity of data objects. When ants do not load, calculate probability P_1 of ants picking up data objects. If P_1 is bigger than some random probability, and this data object is not picked up by other ants. Then this ant picks up this data object and labels itself as having load. Then turn to other places in the given two dimension plane. If P_1 does not meet the condition, the ant will refuse to pick up this data object and selects other data objects in the area randomly. When the ant has data object load, it is necessary to calculate probability of putting down data object, which is P_d. If P_d is bigger than some random probability, the ant will put down this data object and labels itself as unloading. Then the ant will choose another new object. If the probability of putting down data object does not meet the condition, the ant will move to another new position. If an object is in a state of isolation, or the neighborhood object number is less than a constant, the object is isolated point, which needs to be marked. Otherwise a clustering sequence number is assigned to the object, and recursively is assigned the serial number to the neighborhood object as a tag. If at time t, the ant finds a data object o_i at position r, the average similarity is calculated by

$$f(o_i)=\max\left\{0,\frac{1}{s^2}\sum_{j\in n(r)}\left[1-\frac{d(o_i,o_j)}{a(1-(v-1)/v_{\max})}\right]\right\}$$

a is similarity parameter, v is moving speed of ant, $v_{\max}$ is the maximum moving speed of ant. $n(r)$ represents square local area near r, the side length of which is s. $d(o_i,o_j)$ represents distance of object o_i and o_j in the attribute space.

$$d(o_i,o_j)=1-sim(o_i,o_j)$$

$$sim(o_i,o_j)=\frac{\sum_{k=1}^{m}(o_{ik}\cdot o_{jk})}{\sqrt{\sum_{k=1}^{m}(o_{ik})^2\cdot\sum_{k=1}^{m}(o_{jk})^2}}$$

$$P_1=1-sigmoid(f(o_i))$$

$$P_d=sigmoid(f(o_i))$$

$$sigmoid(f(o_i))=\frac{1-e^{-cx}}{1+e^{-cx}}.$$

Record the data object that visited by the ant so that can ensure all data will be visited by ants and prevent the same data picking up and putting down repeatedly. The second improvement is to establish a historical position of global memory, which helps ants to put down the data object faster with target as well as the emergence of local optimal solution.

IV. THE VOF ANALYSIS ON SIPHONIC BEDPAN'S FLUSHING PROCESS

In order to test the application effect of proposed clustering algorithm in student sports performance management system, we choose student sports performance in some university as experiment object, a part of data sample is shown in Table 1. Number represents serial number. PS represents usual performance. FP represents final performance. TP represents total performance. According to k-means algorithm, the performance is divided into five levels (excellent, good, medium, pass, fail),as shown in table 2 and table 3. The clustering results of ant colony clustering algorithm are shown in table 4 and table 5.

Table 1. Sports performance

Number	PS	FP	TP
1	26	48	74
2	26	35	61
3	26	38	64
4	26	51	77
5	26	34	60
6	24	41	65
7	28	32	60
8	25	42	67

Table 2. Clustering results of K-means clustering algorithm

Classification	Number of member	Grade range
1	21	79-90
2	39	70-78
3	17	65-69
4	24	56-64
5	7	43-53

Table 3. Clustering results of K-means clustering algorithm

The minimum error inside class	The maximum error inside class	The average error inside class
0.24	7.24	2.42
0.15	4.15	2.21
0.18	2.18	1.01
0.08	5.92	1.68
0.14	6.86	2.78

Table 4. Clustering results of ant colony clustering algorithm

Classification	Number of member	Grade range
1	12	82-90
2	26	75-81
3	27	68-74
4	35	60-67
5	8	43-56

Table 5. Clustering results of ant colony clustering algorithm

The minimum error inside class	The maximum error inside class	The average error inside class
0.26	5.26	2.00
0.36	3.32	1.81
0.26	3.26	1.68
0.46	3.55	1.98
0.62	6.39	2.97

It can be concluded that clustering algorithm based on ant

colony is more reasonable than k-means algorithm, which can be used in actual sports performance management system.

V. CONCLUSIONS

Data mining, which based on specific algorithms, is a technology for researchers to analyze large amounts of data to find potential knowledge, and widely used in the major fields, makes itself be valuable. Cluster analysis is one of the modes of data mining, which classifies the sample data to different types according to similarity rules, so that in the same type data have high resemblance, different type data become high differential. By clustering, it's easy to describe the characteristic of different types. In order to overcome shortcomings of K-means algorithm, a new clustering algorithm based on ant colony optimization is proposed. The experiment results show that the improved algorithm based on ant colony optimization is more scientific and fair.

REFERENCES

[1] Hong Fan, "Research on the Physical Education Teaching Effectiveness of the Higher Colleges and Universities with Intuitionistic Fuzzy Information", journal of Convergence Information Technology, Vol. 7, No. 15, pp. 463-469, 2012.

[2] Li Xinwu, "A New Volume Segmentation Algorithm for Medical Image based on Improved K_means", International Journal of Digital Content Technology and its Applications, Vol. 5, No. 6, pp. 1-7, 2011.

[3] Dost Muhammad Khan, Nawaz Mohamudally, "An Agent Oriented Approach for Implementation of the Range Method of Initial Centroids in K-Means Clustering Data Mining Algorithm", IJIPM, Vol. 1, No. 1, pp. 104-113, 2010.

[4] Meng-Dar Shieh, Tsung-Hsing Wang, Chih-Chieh Yang, "Affective Response Dimension Selection for Product Design: A Comparison of Cluster Analysis and Procrustes Analysis", International Journal of Digital Content Technology and its Applications, Vol. 5, No. 1, pp. 305- 318, 2011.

[5] Zhengxin Huang, Yongquan Zhou, "Using Glowworm Swarm Optimization Algorithm for Clustering Analysis", journal of Convergence Information Technology, Vol. 6, No. 2, pp. 78-85, 2011.

[6] Gong Zhe, Li Dan, An Baoyu, Ou Yangxi, Cui Wei, Niu Xinxin, Xin Yang, "An Analysis of Ant Colony Clustering Methods: Models, Algorithms and Applications", International Journal of Advancements in Computing Technology, Vol. 3, No. 11, pp. 112-121, 2011.

[7] Xin-Guo Li, "Computer-Web-Based Multimedia College English Teaching Assessment Model with Uncertain Linguistic Information", Advances in Information Sciences and Service Sciences, Vol. 4, No. 13, pp. 178-183, 2012.

[8] Y.K. Wu, Z.H. Tang, "Collaborative Filtering System based on Classification and Extended K-Means Algorithm", Advances in Information Sciences and Service Sciences, Vol. 3, No. 7, pp. 187-194, 2011.

[9] Yu Yuecheng, Wang Jiandong, Zheng Guansheng, Gu Bin, "Semi-supervised Distributed Clustering with Mahalanobis Distance Metric Learning", International Journal of Digital Content Technology and its Applications, Vol. 4, No. 9, pp. 132-140, 2010.

[10] Gong Zhe, Li Dan, An Baoyu, Ou Yangxi, Cui Wei, Niu Xinxin, Xin Yang, "An Analysis of Ant Colony Clustering Methods: Models, Algorithms and Applications", International Journal of Advancements in Computing Technology, Vol. 3, No. 11, pp. 112-121, 2011

[11] Zhengxin Huang, Yongquan Zhou, "Using Glowworm Swarm Optimization Algorithm for Clustering Analysis", Journal of Convergence Information Technology, Vol. 6, No. 2, pp. 78-85, 2011.

[12] Yong Ye, Shaowen Li, "The Model of Support Vector Machine Ensemble based on Artificial Fish Swarm Algorithm", International Journal of Digital Content Technology and its Applications, Vol. 6, No. 16, pp. 356-363, 2012.

[13] WANG Ting, WANG Heng, LI Yong, WANG Ping, "Reliability Analysis for Wireless Sensor of Networked Synchronization Control System by PSO- Least Squares Support Vector Machine", International Journal of Digital Content Technology and its Applications, Vol. 6, No. 21, pp. 492-498, 2012.

[14] Qing Chen, Yuxiang Shao, BaoZhong Liu, "BP Neural Network Based on Improved Artificial Fish Swarm Algorithm and Its Application", International Journal of Advancements in Computing Technology, Vol. 4, No. 19, pp. 507-514, 2012.