Research on Grouping Strategy of Virtual Learning Community Based on the Fuzzy C-means Clustering

Yan Cheng

Post doctoral mobile station of computer science and technology, Tongji University
College of Computer and Information Engineering, Jiangxi Normal University;
Nan Chang, China; Shanghai, China
chyan88888@jxnu.edu.cn

Jianhua Xie

College of Computer and Information Engineering, Jiangxi Normal University
Nan Chang, China
971383331@qq.com

Weisheng Xu

College of Electronics and Information Engineering, Tongji University
Shanghai, China

Abstract—**Virtual learning community, as a new model of network education, is widely used in the process of learning. But this kind of way is not gradually satisfying people's needs for more intelligent and personalized. Teaching strategy in virtual learning community is an important research direction of intelligent reasoning and it is the key to realize intelligent which has great role in promoting for it. In this paper, the fuzzy c-means clustering algorithm is applied to analysis of the characteristics of learning and grouped from the perspective of teaching strategy and reasoning in virtual learning community. And the model of teaching strategy reasoning mechanism based on learning characteristics is designed which use the cosine similarity and grouping clustering center vector and teaching part of the matching degree of policy rule conditions by setting the threshold of matching processing. Formulate specific feasible learning tasks and teaching strategies will be made for the learners. Through the simulation experiments it is found that the teaching strategy reasoning algorithm is feasible and it has achieved good results.**

Keywords—virtual learning community; the fuzzy c-means clustering algorithm; learning characteristic; teaching strategy reasoning mechanism.

I. INTRODUCTION

Teaching strategy is the mutual concern topic in pedagogy, psychology and education technology and many other areas. Since 1964, foreign scholars Taba and Hilda formally put forward the teaching strategy (teaching strategy) concept [1].And it has a wide range of influence which becomes an important content of teaching and research. At the same time, Serena found that use charts, including experience oral and restructuring words have meanings, and in the process of teaching strategies can help students improve their reading comprehension [2]. Starting in the 70s, some psychologists put forward the systematic teaching strategy and it begins to circulate, such as Eggen (1979) and Gagne's system to guide the strategy (1985), Tennyson's goal, the teaching of strategy (1990), Moore (1992), Freiberg's general strategy system teaching strategy (1996) and so on. These theories are widely abroad and also translated by domestic scholars.

The scholars Dajun zhang and Lin Yu think teaching strategy is an integral part of the teaching design and also the specific teaching situation of complete teaching goal .The students' cognition develop a plan of teaching program and teaching implementation of the measures.[3] Ruizhen Shao ,the editor of the book of education psychology, thought the teaching strategy is the relative system behavior that a teacher achieve certain teaching goals in the teaching process.[4] While Xinwu Chen thought teaching strategy is a teacher, according to specific situation and the special teaching method, designed the teaching methods which has a certain regularity and in many cases can be universal. YiPing Wei thought teaching strategy is the field of education technology research, one of the important contents of the teaching design[5]. The view of teaching strategy is essentially to the teaching goal into teaching practice as the core of the link. In order to effectively implement objective, and for the form, process and methods of teaching activities in organic integration with the reasonable design of a series of behavior is a result of the comprehensive plan.

Teaching strategies, as the construction blueprint of a teaching process, can be used as a transition Bridges from teaching to the teaching content. In virtual learning community, there are all sorts of study resource and the learner's study background, study behavior also vary from person to person. In this paper, it designed the teaching strategies and reasoning algorithm with the characteristics of learning as the research object aiming at the problem of insufficient intelligent teaching in virtual learning community.

II. The Characteristics Classification of Learning Based on The Fuzzy C-Means Clustering

Fuzzy C-means algorithm (FCM) is a kind of common fuzzy classification method which Bezdek had used in pattern recognition [7]. According to the similarity of the learning characteristics, we grouped the students by fuzzy c-means algorithm and calculated the center characteristic vector of each group. That's the main work this paper does. After the learners have been divided into groups, the domain experts make some corresponding teaching strategies for targeted teaching so as to realize the community learners' overall improvement.

A. The Description of Learner Clustering Analysis Problem

After classifying learners by learning characteristics, each type of learners have a clustering center, respectively form a vector of clustering center. Each type center vector of the learners' learning characteristics defined as follows.

Definition 2: The learners will be classified by their learning characteristics through cluster algorithm. And the cluster center vector of each group is defined as $Z_k = (Z_{k1}, Z_{k2}, ..., Z_{km})$. Z_k represents learners' overall learning characteristics of group k, Z_{km} represents the value of the mth learning vector in the group k. The value of k is equal to the number of category C.

Clustering analysis was carried out on the learners, namely, the learners are divided into category c. The learner's each classification corresponds to a c*n membership degree matrix: $U = \{u_{ij}\}$, u_{ij} represents for the membership degree between learners with cluster subset S_j. $V = (v_1, v_2, ..., v_c)$ Represents the clustering center set for all learners.

B. Cluster Analysis of the Objective Function Learners

The fuzzy c-means algorithm [8] based on clustering analysis was carried out on the learners' learning characteristics. The objective function of clustering algorithm is as below.

$$\text{Min} J_m = (U, \ V) = \sum_{i=1}^{c} \sum_{j=1}^{n} (u_{ij})^m d_{ij}^2 \tag{1}$$

To make formula (1) minimum, namely, the sum of square of the distance between each learner sample to the center of the cluster. In the formula 1, $d_{ij} = \|x_j - v_i\|$ represents the distance between the jth learner S_j and the ith clustering center v_i, m represents the weighted index, usually value to 2 [9]. The weighted index controls the distribution of the membership degree and the degree of fuzzy clustering.

C. Learners Membership Degree and the Calculation of the Clustering Center

The calculation formula about the learners for each category of membership degree is shown as equation 2. d_{ij} represents the distance between the jth learner and i^{th}

clustering center, $\sum_{k=1}^{c} d_{kj}$ represents the sum of distance between the jth learner and c^{th} clustering center. Actually m value of 2.

$$U_{ij} = \left[\sum_{k=1}^{c} \left(d_{ij} \middle/ d_{kj} \right)^{2/(m-1)} \right]^{-1} \tag{2}$$

The cluster center of the learners is calculated by the formula 3.

$$V_i = \sum_{j=1}^{n} (u_{ij})^m x_j / \sum_{j=1}^{n} (u_{ij})^m \tag{3}$$

D. The Steps of Fuzzy C-means Algorithm

The fuzzy C clustering algorithm steps are as follows: The first step: Determine the number of learner classification c, $2 \le c \le n$, n represents the number of learners. Given the initial cluster centers $V(0)$ randomly, set the termination value ε of the iteration process. Set the initial numerical iteration times p = 0.

The second step: According to the type (2) calculating the value of the membership degree u gets $U^{(p)}$.

The third step According to the formula (3), a new clustering center matrix $V^{(p+1)}$

The fourth step: Using the formula (1) to calculate J_m, if $\left| J_m^{(p)} - J_m^{(p-1)} \right| > \varepsilon$, then let p=p+1 and turn to the second step, otherwise stop to calculate the fuzzy membership matrix U and cluster center matrix V.

III. Teaching Strategy Reasoning Mechanism Based on Learning Characteristics

Virtual Learning Community is not only a learner learning network platform, but also provide teaching services to a variety of learning backgrounds learners organization. In this paper, in order to provide a reasonable teaching strategy for learners, we analyzed the learners' learning characteristics and discuss how to make a division for the study group. Taking the students of ideological political education classes of grade 2014 in Jiangxi Normal University as the research object. Therefore, we designed a model of teaching strategy based on learning characteristic vector. (see Fig.1) This model includes three modules, which includes data preprocessing module, learners feature analysis module and teaching strategies recommended module. Following the first two modules, we will make a specific analysis.

A. Data Preprocessing

Through the virtual learning community platform to record the learner's learning dynamics, data preprocessing module after data cleaning, data integration, data conversion, data protocol and other operations to clutter, irregular, and no regular data to transform the data into useful data.
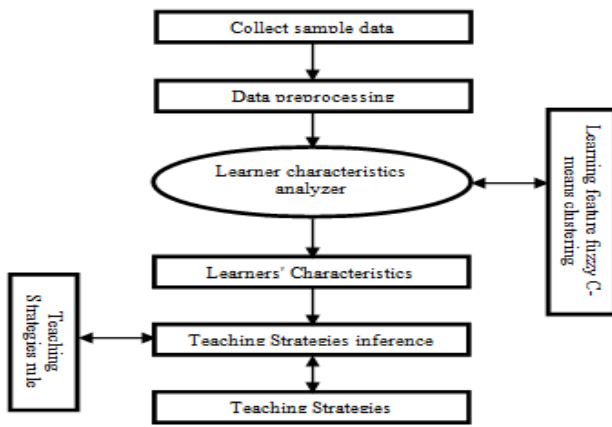
Fig. 1. Feature-based learning teaching strategies Inference System Model

B. Learning Feature Analysis Module

Both traditional education and Internet education, the levels of knowledge and independent learning are the typical indicators of learners' characteristics. In addition, according to the characteristics of the network education, learners' collaborative learning enthusiasm will serve as another important study characteristic. Therefore this paper selects the knowledge level of learners, independent learning, cooperative learning enthusiasm of the three aspects to analysis, and combined with the actual teaching situation of virtual learning community, respectively, the knowledge level, autonomous learning and collaborative learning initiative analysis of its influence factors and their weights.

Among them, the learner's learning feature vector is defined as follows.

Definition 1: Suppose there are n learners, each the dimensions of the learners' learning characteristics is m, then the learner's learning characteristic vector is $S_i = (S_{i1}, S_{i2}, ..., S_{im})$. S_i represents the learning feature vector of the ith learner, S_{im} indicates that the ith learning characteristic values of the mth learners. Where i is not more than n.

In this paper, we do not discuss the weight evaluation, and the weight value determined by assuming [6]. The learner's knowledge level depends on the scores they get at the usual time and final. And we assume each weight is assumed as Q1=0.3,Q2=0.7; The independent learning consists of three parts ,the practice times in the homework module , the learning times in the course module and the test times in examination module . And each of their weight is assumed as Q3=0.3,Q4=0.4,Q5=0.3; Collaborative learning enthusiasm consists of the learning times in the resource sharing module and the times of uploading data ,and each of their weight is assumed as Q6=0.3,Q7=0.7. In this paper, the learning characteristic vector is $S_i = (S_{i1}, S_{i2}, S_{i3})$, S_{i1} represents the knowledge levels, S_{i2} represents the independent learning , S_{i3} represents the cooperative learning enthusiasm . The specific learning characteristics are shown in TABLE I:

TABLE I. CHARACTERISTICS OF LEARNERS IMPACT ANALYSIS

Learning feature	Usually test scores (P1)	Final exam scores (P2)	Job data module number of exercises (T1)	Learning course modules learning number (T2)	Quiz module test times (T3)	Resource sharing module learning number (X1)	Data Upload times (X2)
Level of knowledge(Si1)	Weight Q1	Weight Q2					
Self-learning(Si2)			Weight Q3	Weight Q4	Weight Q5		
Collaborative learning initiative(Si3)						Weight Q6	Weight Q7

TABLE I shows the weight of each learning characteristic. Based on the autonomous learning website of Jiangxi Normal University computer course .And this learning website is study community based on Moodle(Modular Object-Oriented Dynamic Learning Environment) which has recorded the action of the students in this website .Through the Moodle we can get the data we need in the TABLE I .

Here we take the students of ideological political education classes of grade 2014 in Jiangxi Normal University as the research object. And there are 60 students in this class. We will consider one course named fundamentals of computer as an example that 5 invalid records and 55 valid records. According to the setting of the weight to each influence factor,

the learners learning characteristics had been calculated and shown as blow. (See TABLE II)

TABLE II.   THE LEARNER'S LEARNING CHARACTERISTICS

Learning feature / Student ID	Knowledge level (unit: fraction)	Collaborative learning enthusiasm (unit: times)	Autonomous Learning (unit: times)
14*****001	83.91	1.7	42.1
14*****002	40.00	2.2	12.8
14*****003	83.33	2.9	7.7
14*****004	95.65	1.0	23.1
14*****005	88.51	9.9	67.9
14*****006	90.90	1.9	8.2
14*****007	53.89	1.4	31.1
14*****008	44.44	7.2	14.7
14*****009	59.88	4.0	9.1
14*****010	66.29	0.7	2.8
......	......	......	......

IV. THE EXPERIMENTAL PROCEDURE OF THE FUZZY C-MEANS CLUSTERING GROUPING

1) Initialize clustering center learning characteristics. The initial clustering center was selected by center by the system functions selected randomly. Then the appropriate category c can be obtained by the above process.

2) The actual number of iterations. The experiment of the maximum number of iterations is set to 150 times. Actually when the learning characteristics of the recent change of weighted error sum of squares is less than the set termination conditions of 0.00001, the iteration is stopped when the iterate number is 51.

3) Iterative 51 times, the final clustering center of learning characteristics and the membership are determined.

4) Learners are divided into 6 classes according to their membership.

Each learner have a membership degree to the 6 clustering centers, the learner will be divided into one of maximum value membership category.

V. GROUPING STRATEGY OF VIRTUAL LEARNING COMMUNITY BASED ON THE FUZZY C-MEANS CLUSTERING

After the termination of algorithm, learners are divided into 6 categories according to learning characteristic by the fuzzy c-means algorithm.

According to the learner's level of knowledge, self-learning and collaborative learning initiative, the 55 students in the class were divided into six groups by fuzzy c-means clustering algorithm. And it gives the corresponding teaching strategies through the teaching rules reasoning mechanism model. The teaching strategy rule base is given by the expert system, and teaching strategy of the rule base is designed according to the ideas of the method of production rules.

The clustering center of grouping a learning characteristic vector Z_1= (72.8604, 3.8411, 81.1484).For cluster grouping learners, the match degree is calculated by clustering center vector Z_1 and six teaching strategies through cosine similarity calculation characteristic. The results are as follow which are calculated in MATLAB through cosine similarity function:

$$P\left(Z_1, T_1\right) = \sim\left(Z_1, T_1\right) = \frac{Z_1 \bullet T_1}{\|Z_1\|\|T_1\|} = 0.8918;$$

$$P\left(Z_1, T_2\right) = \sim\left(Z_1, T_2\right) = \frac{Z_1 \bullet T_2}{\|Z_1\|\|T_2\|} = 0.7286;$$

$$P\left(Z_1, T_3\right) = \sim\left(Z_1, T_3\right) = \frac{Z_1 \bullet T_3}{\|Z_1\|\|T_3\|} = 0.9731;$$

$$P\left(Z_1, T_4\right) = \sim\left(Z_1, T_4\right) = \frac{Z_1 \bullet T_4}{\|Z_1\|\|T_4\|} = 0.5197;$$

$$P\left(Z_1, T_5\right) = \sim\left(Z_1, T_5\right) = \frac{Z_1 \bullet T_5}{\|Z_1\|\|T_5\|} = 0.9473;$$

$$P\left(Z_1, T_6\right) = \sim\left(Z_1, T_6\right) = \frac{Z_1 \bullet T_6}{\|Z_1\|\|T_6\|} = 0.8846;$$

The calculation result shows that the matching degree is $P(Z_1,T_2),P(Z_1,T_4)$, less than the threshold value f = 0.85, to be eliminated directly. $P(Z_1,T_1)$, $P(Z_1,T_3)$, $P(Z_1,T_5)$, $P(Z_1,T_6)$, more than the threshold value f = 0.85, which are saved to the target teaching strategy in the rule base. Then according to the matching degree of big priority rules to choose the rules of matching degree of the largest, by the calculating results, the teaching strategy C_3 is to be chosen because $P(Z_1,T_3)$ is the largest. That is to say, for the first group of learners, the most suitable teaching strategies for the teaching strategy is C_3 because the matching degree of the highest. And according to the teaching strategy C_3 to perform some conclusions of the corresponding learning activities to improve the teaching effect.

So the most appropriate cluster grouping teaching strategies of the learner is C_3, its strategy is: focuses on the organization form of network study group for learners to participate in the study discussion and sharing of teaching resources, and adopt virtual reality, video, slide form to enrich the content of teaching resources, etc.

VI. CONCLUSION

In this thesis, we analyze the problems of network education in the new era. By analyzing the characteristics of the learners, the learning feature vector is set up. The feature vector of the learners is the research object, and the feature vector of the clustering center is studied. The fuzzy C means algorithm is used to divide the learners. The specific teaching strategies are provided to the learners through the teaching strategy reasoning mechanism model, and also the formulate specific feasible learning tasks and goals. Through the simulation experiment results, we found that the teaching strategy has achieved good results; the learner's knowledge level, collaborative learning enthusiasm and autonomous learning have been improved to a certain extent. And it can also be established by the virtual learning community to establish a discussion area for each group. In short, we hope that the study of the grouping strategy based on the learning

characteristics can provide the theoretical and practical support for the further development of the network education.

REFERENCES

[1] Taba & Hilda & Elzey & Freeman. F. Teaching strategies and thought processes[J]. Teachers College Record. 1964, 65(6): 524- 534.

[2] Niensted Serena. Meaninglessness for beginning readers. Reading Teacher[J]. 1969, 23(2): 112- 115.

[3] Dajun zhang, Yu Lin. Try to talk about the teaching strategy of the reference for the basic connotation and [J]. Journal of course. Teaching materials. Teaching, 1996,09:6-8

[4] Ruizhen Shao. Learner-centredness as language education [M]. Shanghai: Shanghai education press, 1997:80. (In Chinese)

[5] YiPing Wei. Three-dimensional research perspective of teaching strategies [J]. Journal of teacher education research, 2006,01:29-33 + 54. (In Chinese)

[6] Xiaoxia Sun,Xiaoxia Liu,Qianru Xie.The implementation of the fuzzy C-means clustering algorithm [J].Computer Applications and Software2008,03:48-50. (In Chinese)

[7] Xinbo Gao, Jie Li,Weixin Xie.Optimal choice of weighting exponent in a fuzz C-means clustering algorithm [J]. Pattern Recognition and Aitificial Intelligence,2000,01:7-11. (In Chinese)

[8] Han Xue, Xie Lili, Keep righteousness. Motor fault diagnosis expert system based on fuzzy reasoning study [J]. Computer measurement and control, 2010, 01:8-10. (In Chinese)

[9] Yanxing Niu, Songhua Shen, Shiliang Dong. The aircraft power supply system based on fuzzy reasoning fault diagnosis expert system [J]. Journal of electronic measurement technology, 2007, 12:59-62. (In Chinese)

[10] Gaiyun Gong, Yongcai Mao,Xinbo Gao. Based on the fuzzy c - average clustering of microarray gene expression data analysis [J]. Journal of xi 'an university of electronic science and technology, 2004, 11:291-295. (In Chinese)