Student Model and Clustering Research on Personalized E-learning

Sheng Cao[1], Songdeng Niu[1], Guanghao Xiong[2], Xiaolin Qin[3], Pengfei Liu[1]

[1] School of Computer Science and Engineering, University of Electronic Science and Technology of China, China
[2] School of Aeronautics and Astronautics, University of Electronic Science and Technology of China, China
[3] Chengdu Institute of Computer Applications, CAS, China
caosheng@uestc.edu.cn, {842549301, 714461296}@qq.com, qinxl2001@126.com, liupengfei@uestc.edu.cn

Abstract

As the internet and data mining technologies are developing rapidly, how to provide various students with high-quality education services has become the hotspot in the internet environment. In order to promote the characteristics of online education and enhance the quality of personalized e-learning, in this paper, we propose a novel algorithm named MK-means by exploiting the cluster-wise weighing co-association matrix mechanism and improving the K-means algorithm based on the mean shift theory. The experimental results on the UCI's Iris and Wine test sets demonstrate its effectiveness and superiority, finding that the total F-measure of MK-means achieves better performance than the Hierarchical Clustering, FCM, K-means, SOM, and X-means algorithms. Finally, the new algorithm combined with the student model explains the clustering results in detail from perspectives of cognitive model and knowledge map respectively and can extend to support the personalized e-learning in a wide range.

Keywords: E-learning, Mean shift, K-means, CWWCA, Student model.

1 Introduction

E-learning is a method of content dissemination via applying information and internet technologies. Online learning can expand students' scope of knowledge, improve their personalities, and develop the interests of students [1]. Meanwhile, it also enhances the learning quality of all aspects you can imagine. E-learning has become a primary way of further education and lifelong learning [2]. The ideal online learning environments should offer close cooperation between teachers and students. Thus, knowledge construction of students, personalized learning guidance, learning ability, characteristics of the full development, and optimal learning path selection of intelligent online education are being concerned by more and more educational technology researchers and internet products. The capital goal of intelligent e-learning is to implement the targeted study for each online student so as to guide them under their aptitudes. Namely, the future of e-learning will become more and more personalized and social.

Research on myriad existing online e-learning platforms has found that personalized features have been changed continuously due to different learning abilities and hobbies from different students [3]. But the guidance of learning strategies is still imperfect and has great development potential. Actually, online e-learning generates extensive data that can be fully analyzed and utilized. As data mining and cluster ensemble technologies have developed rapidly, it's possible to mine the knowledge from these educational data and analyze the characteristics of different online students to guide their development of individual character. At present, in the field of online education, applying data mining and clustering methods to analyze the behavioral data of online students has been widely studied [2, 4].

It can be assumed that for all scenarios, no clustering method is better than all other methods. For instance, linkage-based clusters may not be suitable for clusters with a normal distribution. By adopting different clustering methods, different partitions of a given dataset can be obtained. Therefore, it is quite challenging to choose the best clustering algorithm for a certain application and dataset. Besides, since the data generated by e-learning reflect the characteristics of students in different dimensions, the data cannot be processed equally when using clustering algorithms. Existing methods fail to fully describe the student online learning process due to insufficient feature dimensions of students' behavior data. This is also the reason why we consider using weighted integrated clustering methods to cluster online learning data. The cluster-wise weighing co-association (CWWCA) matrix evaluates the independence of the clusters while considering the weighting mechanism and defines the

*Corresponding Author: Sheng Cao; E-mail: caosheng@uestc.edu.cn
DOI: 10.53106/160792642021072204020

diversity at the level of clustering [5]. In the clustering ensemble process, it considers the quality and diversity of clusters to improve consistency partition.

In this paper, we propose a new ensemble clustering technology of data mining and extend it to the students' personalized online learning. We cluster various features of learning data from students' learning behaviors and then divide students into groups with similar characteristics. Subsequently, we make suggestions on implementing learning contents and strategies according to different characteristics of each group. The main contributions of this paper are summarized as follows:

(1) An improved K-means algorithm named MK-means is proposed by allocating the co-association weighing of student feature data and mean shift theory to converge the student feature data to high-density sets based on the cluster-wise weighing co-association (CWWCA) matrix.

(2) Through the comparative experiments with hierarchical clustering, SOM, FCM, and original K-means algorithm on the internationally used UCI data sets of Iris and Wine, our new clustering algorithm is superior in the quality of clustering results.

(3) Our clustering algorithm is applied to analyzing the personalized student model and provides favorable evidence for guiding students to the next learning direction and path.

The second section of this paper introduces the research of ensemble clustering algorithms in recent years. The third section represents the designed student model, including the knowledge model, the cognitive model, and the emotion model. The fourth section provides the prior knowledge and accurate definition of our MK-means algorithm and describes the corresponding algorithm flow. The fifth section shows the effectiveness and superiority of MK-means through comparative experiments. In the sixth section, the MK-means algorithm is applied to the students' feature data, and the result is compared with X-means and K-means. The last section ends with the conclusion and future work.

2 Related Work

The purpose of ensemble clustering is to obtain a better and more robust consistent clustering result by combining multiple base clusters. In the past decade, many ensemble clustering methods have been developed and mainly consist of three categories: methods based on the pair-wise cooccurrence, graph partition, and median partition.

In pair-wise cooccurrence methods, the co-association (CA) matrix is usually constructed regarding the number of times that two objects in multiple base clusters appear in the same cluster. Using the CA matrix as a similarity matrix, the traditional clustering technology can be employed to build the final clustering results. For instance, Fred and Jain proposed the CA matrix and evidence aggregation clustering (EAC) approach [6]. Wang et al. improved the EAC approach focusing on the size of clusters and proposed a probability accumulation method [7]. Alizadeh presented another modified EAC (EEAC) method based on evidence aggregation clustering by utilizing the consensus function of the co-association matrix to combine the selected clusters [8]. In Mojarad and Nejatian's work, a clustering similarity matrix was obtained from fuzzy clusters, and then a hierarchical clustering algorithm was applied to dividing the fuzzy cluster on the cluster similarity matrix [9]. Nazari designed the concept of cluster-level weighting co-association matrix to replace the traditional co-association matrix [10]. In their papers, two consensus functions are introduced and exploited to generate consensus partition.

In the methods based on graph partition, the problem of ensemble clustering is solved by constructing a graph model reflecting the ensemble information. The graphs are partitioned into a certain number of segments to get the consensus clustering. Hamid Parvin proposed the consensus function of Local Weighted Graph Partition (LWGP), using cluster diversity and combining the local weighting strategy [11]. Dong Huang described two ensemble clustering methods based on a factor graph (ECFG) and sparse graph representation [12-13]. The Hierarchical Clustering Ensemble Selection (HCES) method was presented by Akbari et al. and is a complete solution of pair-wise diversity measure and cluster-based similarity partition algorithm [14].

The method based on median partition regards the ensemble clustering problem as an optimization problem, whose purpose is to find the median segmentation (or clustering) by maximizing the similarity between the cluster and multiple base clusters. The median partition is an NP-hard problem [15]. For large data sets, it is not feasible to find the globally optimal solution in the large space of all possible clusters. Cristofor et al. proposed a genetic algorithm to obtain an approximate solution, in which clustering could be regarded as chromosome [16]. More partition methods based on median segmentation are discussed in the following work [17-18].

These algorithms mentioned above attempt to solve the problem of ensemble clustering in various ways. However, the common limitation of most existing methods is that they usually treat all clusters and all base clusters equally and may be affected by low-quality clusters or base clusters. In order to reduce this limitation, some weighted integrated clustering methods have been proposed recently by Alguliyev and Parvin et al. [19-20]. Besides, Rashidi et al. exploited the concept of cluster uncertainty for consensus clustering based on the research of Parvin and utilized the cluster independence based on information theory to calculate the contribution weight from the entropy between the data label points of the clustering [5]. However, this uncertainty measurement is highly sensitive to the cluster size. It punishes clusters with larger sizes and makes it unstable when facing large clusters, which is not suitable for analyzing student models and enhancing online learning quality.

Contrary to Rashidi's work, referring to the concept of clustering weighted association matrix and combined with the K-means method (a dynamic clustering algorithm based on partitioning) [21], a mean algorithm based on shift K-means is proposed in this paper, which does not need to calculate the entropy between the data labels of the clusters. The sensitivity of this method is far lower than that of clustering uncertainty for consensus clustering. Also, with the increase of iteration times, the sensitivity to noise and the lack of easy access to local extremum points of K-means are addressed. The improved algorithm makes the selection of initial clustering centers focus on high-density areas, increases clustering stability, and reduces the probability of local optimization.

3 Student Model Research

The student model is to infer the degree of knowledge mastery, the ability to learn knowledge, and the personality learning characteristics of students through analyzing the students' basic information, knowledge background, action behaviors, and learning data. Therefore, developing a suitable and adaptable student model is vital to realize personalized e-learning. Compared with the traditional student model [1-2], we introduce the knowledge model (i.e., knowledge map), cognitive model, and emotion model in this section to improve the accuracy and efficiency of online learning.

3.1 Knowledge Map

Generally speaking, books are the carriers of knowledge, where a book has many chapters and each chapter contains several subsections. Inside each subsection, there are many knowledge points, which are called knowledge units [22-23]. The knowledge connections between different units and the order required to learn are called knowledge maps, which identify the positions of knowledge and reveal the relationships between different knowledge units.

According to the sequence of knowledge units, the "road" is called knowledge path or knowledge navigation. At the same time, each student has different knowledge levels and different choices of learning order, which forms a unique path of knowledge. For example, the data structure course contains several major chapters such as the linear table, stack, queue, broad table, array, tree, string, graph, sorting, search, and so on. We can divide this course into multiple knowledge units and draw the knowledge map according to the relationships between the knowledge units, such as the knowledge map shown in Figure 1 [23].

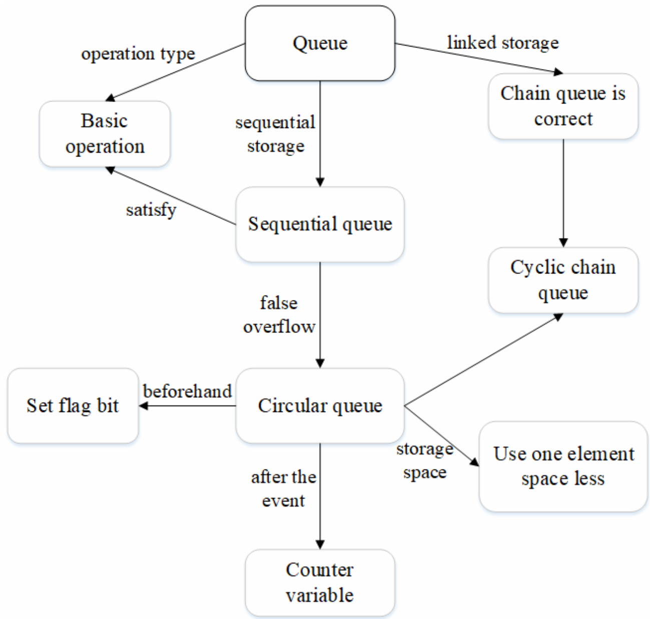


Figure 1. Knowledge map

The knowledge map contains the expert knowledge basis, which forms the best-recommended knowledge navigation for the students. When a student is learning online via his/her knowledge map, knowledge units can record the very knowledge that has been studied according to the paths of the knowledge map. Thus, a knowledge map can not only point out the students' learning path and learning progress but also reflect the students' mastery degree of all the knowledge points.

3.2 Cognitive Model

Learning is a highly complex process. Based on the theory of Bloom [24], the dilemma of mastering knowledge units is not only associated with students' learning ability but also has great relevance with their cognitive ability.

The cognitive ability is divided into six grades according to the complexity of intellectual activities, including memorizing, understanding, application, analysis, synthesis, and evaluation, respectively. The cognitive ability is represented by a vector named a_i, which indicates the cognitive ability of item $i, i = (1, 2, 3, 4, 5, 6)$. Gagne believes that the ability to test the topic is progressive [24]. Therefore, when a topic reaches a certain level, the ability value less than or equal to the current cognitive ability level is set to 1. When students give the right answers to the questions, the cognitive level is set to 1, whereas the wrong cognitive ability level is set to -1; otherwise, it is set to 0, as expressed in formula (1).

$$a_i = \frac{r_i(1)}{r_i(1) + r_i(-1)} \tag{1}$$

where $r_i(1)$ is the number of correct questions, and $r_i(-1)$ represents the wrong ones. Based on each student's matrices, six cognitive abilities are calculated for each student. Then according to different cognitive abilities, we recommend the corresponding ability of the topic, which can also be made into complementary learning groups by different value abilities and improve the learning enthusiasm and efficacy of students.

3.3 Emotion Model

A student's emotion model can be analyzed to find out the emotional changes when he/she is learning the related materials online, such as watching a video, reading a document, or in other learning processes. The research on emotion models from different aspects has achieved good results. For example, external devices are applied to detect emotional changes based on facial expressions [25], biosensor [26], etc. As the emotion model involves human emotion factors, this paper mainly aims at the objective descriptions and research on the knowledge units, so we consider the emotional model as the supplement of our proposed method.

3.4 The Working Process of the Student Model

The working process of the student model is shown in Figure 2. Entering an e-learning system for the first time, the initial student model can be analyzed from basic information, and then the inference engine distributes learning content and other learning paths to students according to the knowledge map. The learning content can be various media, such as PPT, video, and documents. While a student is in the process of learning, the length of the media needs an appropriate allocation according to the current personal status since everyone's attention span is different. When the student begins to do exercises after learning each knowledge unit, the knowledge model and the cognitive model work together to recommend suitable exercises from the student's difficulty ratio to improve his/her cognitive ability and strengthen the knowledge points. In the meantime, the emotion model adjusts the student's emotional changes according to his/her emotional self-feedback and accuracy of exercises to make him/her in an efficient learning state.

The student forms his/her personal knowledge base, wrong question database, and personality database when he/she studies and does exercises, and these data are fed back to the student model to dynamically update the model and the basic information base.

The student personality database records the student's learning behavior during online learning and can analyze the characteristics of student's knowledge structure and learning style. According to these characteristics, adaptive learning content, learning methods, and teaching strategies can be assigned to different students.

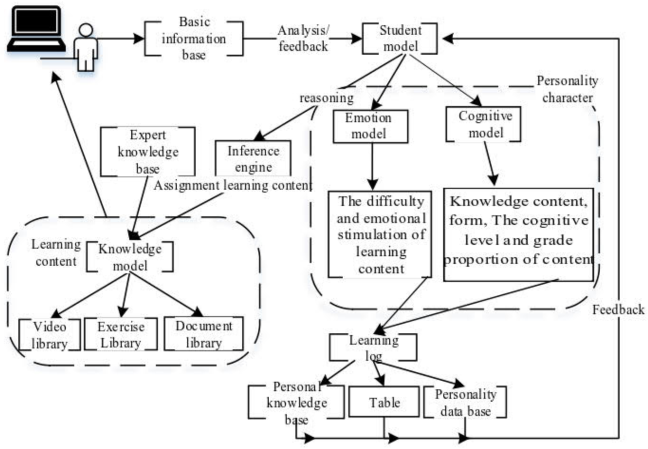


Figure 2. Working diagram of student model

4 MK-means Algorithm

4.1 Cluster-wise Weighing Co-association Matrix

Fred and Jain first proposed a co-association (CA) matrix, reflecting the number of times that two data objects are grouped into the same cluster in multiple basic clustering in the ensemble [4].

Definition 1. Given an ensemble Π, the co-association (CA) matrix is computed as

$$A = \{a_{ij}\}_{N \times N} \tag{2}$$

with

$$a_{ij} = \frac{1}{M} \cdot \sum_{m=1}^{M} \delta_{ij}^{m} \tag{3}$$

$$\delta_{ij}^{m} = \begin{cases} 1, & if\ Cls^{m}(o_i) = Cls^{m}(o_j), \\ 0, & otherwise. \end{cases} \tag{4}$$

where $Cls^{m}(o_i)$ denotes the cluster in $\pi^{m} \in \Pi$ that object o_i belongs to.

The CA matrix is a classical and widely used tool to deal with ensemble clustering problems [9, 12]. Despite its remarkable success, one limitation of the CA matrix is that it treats all clusters and all base clusters equally and lacks the ability to evaluate and weigh the reliability of set members. Parvin et al. constructed a weighed association matrix (WCA) by weighting the basic clustering with the NCAI index [10]. But this matrix only considers the reliability of the basic clustering and ignores the diversity of clustering in the same basic clustering. Rashidi and Nejatian refined the CA matrix through measurement based on clustering independency or uncertainty and proposed the concept of cluster-wise weighing co-association (CWWCA) matrix for representing a clustering ensemble [5], which can effectively capture the implicit relationship among the objects with higher clustering accuracy.

Definition 2. Given an ensemble Π, the cluster-wise weighing co-association matrix is computed based on formula (5).

$$CWWCA_{ij} = \frac{1}{B} \cdot \sum_{k=1}^{B} W_{ij}^{k} \quad \textbf{(5)}$$

where W_{ij}^{k} is defined based on formula (6)

$$W_{ij}^{k} = \begin{cases} Dep\left(\tau_{\pi_i^k}^{k}, \Pi\right), & \pi_i^k = \pi_i^k \\ 0, & otherwise \end{cases} \quad \textbf{(6)}$$

where $\tau_{\pi_i^k}^{k}$ is the π_i^k th cluster of the kth partition, and $Dep\left(\tau_{\pi_i^k}^{k}, \Pi\right)$ can be either $ED\left(\tau_{\pi_i^k}^{k}, \Pi\right)$ or $ND\left(\tau_{\pi_i^k}^{k}, \Pi\right)$.

Through the above weighing mechanism considering the cluster dependability, a data point that usually falls into a dependable cluster is highly likely to be placed into a correct cluster. By the cluster-wise weighing co-association matrix, the dependability of clusters is taken into consideration and is suitable to be applied in personalized e-learning training.

4.2 K-means algorithm

K-means algorithm, a dynamic clustering algorithm based on partitioning, is proposed by MacQueen [21], whose idea is to calculate with low time complexity and effective support of big data. Therefore, K-means has been used widely in many fields.

Define data set D with n raw data samples $(x_1, x_2, \ldots, x_n)$, where each $x_i (i = 1, 2, 3, \ldots, n)$ represents a vector of d dimensions, and they are divided into $k(k < n)$ clusters. Every cluster contains at least one sample, and each sample only belongs to one cluster.

The process of K-means is as follows:

(a) Initializing clustering centers: taking k data samples from the data set D as the clustering centers.

(b) Calculating the similarity between the remaining data samples and the center of the k clusters and classifying them into the clusters with the highest similarity.

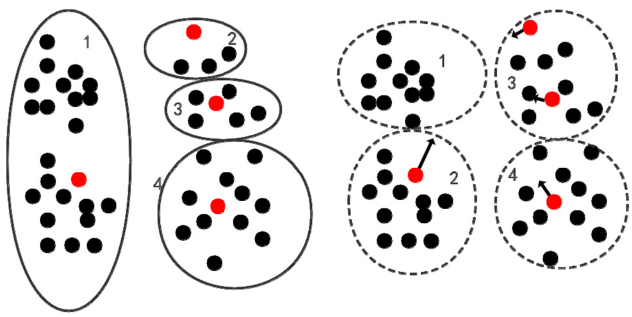
(c) In the new clusters, recalculating the center of k clusters by taking all data samples' average values in the clusters.

(d) Reclassifying the remaining data samples according to the new centers.

Then, the loop runs from (b) to (d) for t iterations until the nearest two clustering centers are no longer changed or repeated. K-means' time complexity is $O(tkn)$, where n is the number of sample data, k is the number of clustering centers, and t is the number of iterations for completing the clustering process. It is also observed that approximately $t \propto n$, hence the effective time complexity becomes $O(n^2)$ [27].

The advantages of the K-means algorithm lie in that it is easy to implement, and it is not necessary to determine the distance matrix for large data sets, which has the features of scalability, high efficiency, low time complexity, and near linear. The main shortcomings include that k must be presented if the number of clusters has been generated.

K-means algorithm has a great influence on setting the k values of the clustering results. The literatures had searched for optimal k values [28-29]. However, the choice of the initial cluster centers is sensitive and easy to fall into local minimum [30-31]. As shown in Figure 3(a) and Figure 3(b), the initial four centers are selected on the left, but it can be clearly found that the data set is divided into four categories on the right. Obviously, randomly selecting initial centers has potential risks.



(a) Four initial centers selected randomly

(b) Final possible clustering results

Figure 3. K-means new sensitivity diagram for initialization

The criterion function used in the K-means algorithm is the criterion function of the square sum of error:

$$J_D = \sum_{i=1}^{k} \sum_{p \in D_i} \| p - \mu_i \|^2 \quad \textbf{(7)}$$

where μ_i is the average value of all data objects in the dataset D_i, and p is the data object in D_i. By analyzing the errors of the square sum, it is found that if the data sets are distributed densely and the boundaries between the sets are distinct, the errors of the sum are valid; otherwise, a large set may be divided into several subsets. Secondly, the extreme point of the objective function should correspond to the optimal points of clustering results. Nevertheless, there is an extreme point of the objective function and multiple local extrema, and the search direction is along with the reduced direction of the criterion function J_D, so different initial values may lead to the local minimum.

If the discrete points are selected as the initial clustering center, the number of cluster iterations will increase and deviate from the dense region. Suppose the discrete points are selected after one iteration. In that case, the clustering centers will be recalculated, which will lead to the clustering centers deviating from the data-intensive regions and reduce the quality of clustering.

4.3 Mean Shift Theory

In 1975, Fukunaga et al. proposed the concept of the mean shift in an article about the gradient of probability density function [32], which referred to the mean vector of offset. With the development of the mean shift theory, the related mean shift algorithm has emerged. As shown in Figure 4, first of all, we select point *A* as the initial point and then calculate the mean offset of *A* to obtain the point *B*. Next, we move the point to *B* as a new starting point, continue to calculate the offset mean, and keep moving to meet the preset end condition. In 1995, Cheng made a generalization of the mean shift theory [33-34], which introduced the kernel function that the offset vector's contribution is related to the sample points and offset distance.

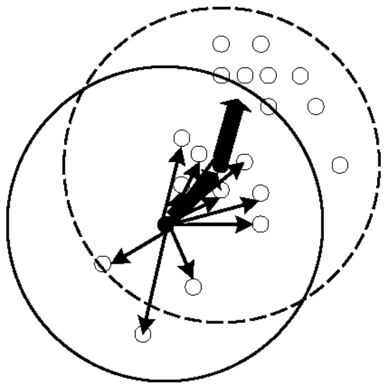


Figure 4. Mean shift sketch map

Given the *D*-dimensional space of *n* data sample points as x_i, where $i = 1, \ldots, n$, the mean shift vector of any point in the space can be represented as:

$$M_h(x) = \frac{1}{k} \sum_{x_i \in S_h} (x_i - x) \quad \textbf{(8)}$$

Among them, S_h is a region of the high-dimensional ball with a radius of *h* (*D* may be greater than 2) to meet the points set formula described in formula (9). *k* indicates the number of points that fall into the S_h area with the *n* samples.

$$S_{\boldsymbol{h}}(\boldsymbol{x}) = \{y : (\boldsymbol{y} - \boldsymbol{x})^T (\boldsymbol{y} - \boldsymbol{x}) < \boldsymbol{h}^2\} \quad \textbf{(9)}$$

Simply speaking, as shown in Figure 4, the solid point is arbitrary in *n* as the center and *h* as the radius to make a high-dimensional sphere, then the mean value of the *k* vectors is calculated. There are *k* points falling into the ball and generate *k* vectors with the center of the ball. Besides, the vector endpoint of a high-dimensional sphere is used as the new starting. Repeating the above steps, then the mean shift vector converges to a relatively large density region.

4.4 MK-means Algorithm Flow

From the above analysis, we can deduce that randomly selecting the initial cluster centers will significantly impact the clustering quality and cause unstable clustering. The sensitivity is associated with noise and easy access to local extreme points. Besides, the dependence between clusters has also been ignored. So, aiming at these shortcomings, this paper proposes a mean algorithm based on the CWWCA weighting mechanism and mean shift K-means, which makes the selection of the initial cluster centers concentrated in a high-density area. Consequently, we can reduce the number of iterations, increase the stability of clustering, and decrease the probability of falling ratio into local optimization.

In the mean shift algorithm, there is a coefficient *h* needed to be set. This experiment finds out the size of *h* is related to the degree of difference between the attributes of the data sets. When the difference is larger, *h* can be a little bigger.

The improved algorithm steps are as follows:

Step 1. Using the mean shift algorithm to process the data set *D* and get *m* points.

Step 2. Selecting *k* points in *m* points that the sum of distances with other points is maximum and calculating their weights with formula (5) and (6).

Step 3. Finding the points closest to the *k* points in the data set *D* and setting the initial center.

Step 4. Calculating the similarity of each remaining data sample point and *k* cluster centers, and then integrating the sample points into the clusters with formula (7), (8), and (9).

Step 5. In the new cluster, the center and the weights of each cluster from *k* are recalculated by computing all data samples' mean in the clusters.

Step 6. Reclassifying all the elements in *D* as the new centers.

The loop runs from Step 4 to 6 for t' iterations until the nearest two clustering centers are no longer changed. Otherwise, it stops when t' reaches the iterations. The flow chart is illustrated in Figure 5.

The time complexity of the mean shift algorithm is $O(mn\log(n))$ in lower dimensions, where *n* represents sample number and *m* represents point number; In higher dimensions, the complexity tends towards $O(n^2)$. Thus, the complexity of MK-means becomes $O(mn\log(n) + nkt')$, where *k* is the number of clustering centers and t' is the iterations. In higher dimensions, the complexity tends towards $O(n^2)$ as well.

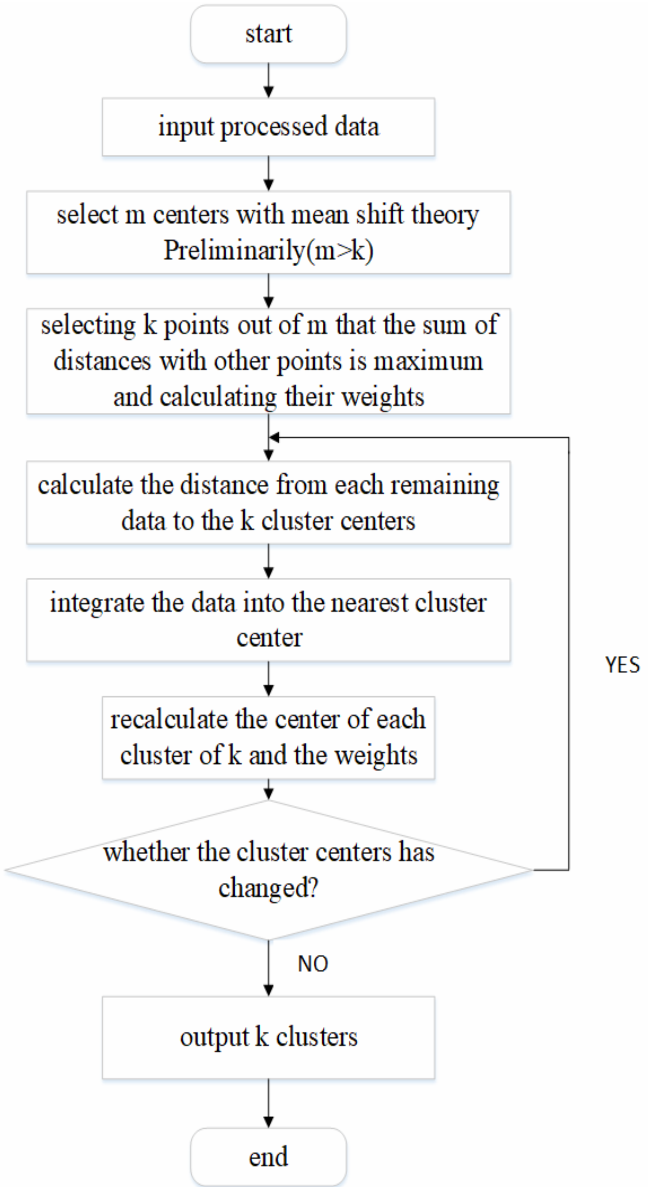


Figure 5. MK-means flow chart

In terms of time complexity, although the complexities of MK-means and K-means are in the same order of magnitude, MK-means significantly reduces the number of iterations to recalculate k centers. In addition, the advantage is obvious: our improved algorithm overcomes the problem that K-means is sensitive to the selection of initial clustering centers and easy to fall into local optimum, which is significant, effective, and applicable to the analysis of student characteristic data.

The m points all in the maximum density are selected by mean shift. Figure 6 is an Iris data setting h = 0.65 with the mean shift algorithm and then obtained seven relatively large points. Figure 7 is a Wine data setting h = 100 with the mean shift algorithm and then obtained eight relatively large points.

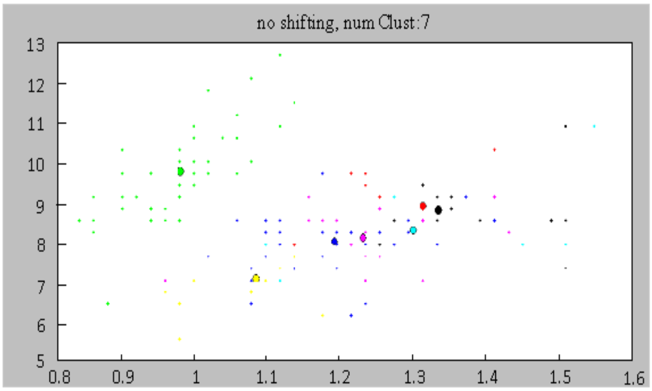


Figure 6. Iris data setting h=0.65

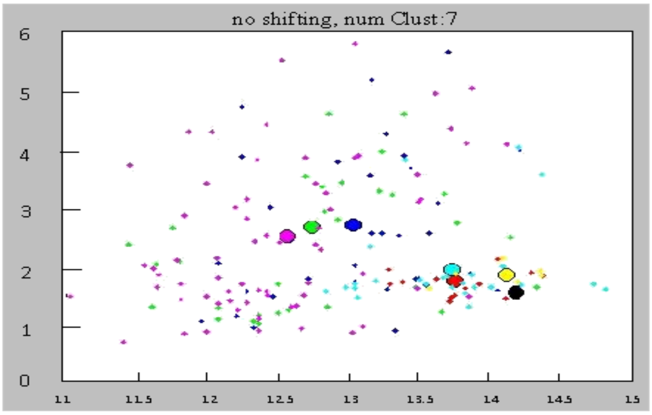


Figure 7. Wine data setting h=100

5 Experimental Results and Analysis of MK-means

5.1 Experimental Tools

The experiment tools were MATLAB, SPSS, and Weka. MATLAB was used to implement comparison experiments for the cluster algorithms, such as hierarchical clustering, K-means clustering, FCM clustering, and SOM clustering [35]. The characteristics of experimental data sets were analyzed by the X-means algorithm and SPSS, and were compared with the characteristics generated by Weka. We selected two widely used data sets of UCI's Wine and Iris sets.

UCI is a specialized database designed to test data mining algorithms and machine learning algorithms. It has clear classifications, so we can use the external criteria to intuitively judge the quality of the clustering algorithm. Iris and Wine are two international data sets that represent two different testing angles. In this paper, experimental results are represented through research and analysis of the two data sets.

5.2 Verify the Effectiveness of the MK-means Algorithm

After filtering by mean shift with CWWCA, the initial clustering center of the K-means algorithm is obtained. In order to make the *m* sample data points meet the condition of $m > k$, the value of *h* should be set. After repeating experiments, the mean value of each attribute value is *r* times in the data set, the value of *h* is close to $r/10$, the mean difference between Iris sets is no more than 10 times, and the value of *h* will be denoted to [0,1]. The mean difference of Wine is more than 1000 times, and the value of *h* is between 100-200.

Subsequently, the algorithm proposed in this paper was compared with several classical algorithms.

Firstly, the Iris data set was used with various hierarchical clustering methods. The similarities between the samples were calculated with the shortest distance and the longest distance, average distance and deviation square sum distance separately to carry out the experimental results, measured by the F-measure value, shown in Table 1.

Table 1. F-measure values of different hierarchical clustering algorithms for Iris data sets

Similarity measure-ment	Shortest distance	Longest distance	Average distance	Deviation distance
Class1 F-measure	100.00%	100.00%	100.00%	100.00%
Class2 F-measure	67.57%	69.23%	87.72%	85.96%
Class3 F-measure	7.69%	80.32%	83.72%	81.90%
All F-measure	58.42%	83.19%	90.48%	89.12%

Next, MK-means, K-means, FCM, and 1-D SOM algorithms were used to practice experiments on the Iris data set, and the experimental results were measured by the F-measure value criterion. The results are listed in Table 2.

Table 2. F-measure values for Iris data are set by MK-means, K-means, FCM, and SOM algorithms

Clustering	MK-means	K-means	FCM	SOM
Class1 F-measure	100.00%	100.00%	100.00%	100.00%
Class2 F-measure	95.31%	82.71%	91.26%	89.10%
Class3 F-measure	95.67%	81.81%	82.22%	91.58%
All F-measure	96.13%	88.17%	91.16%	93.57%

As it can be seen from Table 1 and Table 2, the total F-measure value of the MK-means algorithm is as high as 96.13%, which is up to 8% higher than the hierarchy method, K-means, FCM, and SOM algorithms. It shows that the MK-means algorithm performs well on the Iris dataset.

For the Wine data sets, the various methods of hierarchical clustering were utilized. The similarities of Euclidean distance between the samples, the shortest distance and the longest distance, and average distance and deviation square sum distance were also calculated separately to finish the experiment. The experimental results were measured by the F-measure value, as shown in Table 3.

Table 3. F-measure values of different hierarchical clustering algorithms for Wine data sets

Similarity measure-ment	Shortest distance	Longest distance	Average distance	Deviation distance
Class1 F-measure	15.71%	84.22%	79.62%	86.12%
Class2 F-measure	58.21%	73.16%	0.00%	70.65%
Class3 F-measure	0.00%	41.00%	53.34%	51.04%
All F-measure	24.64%	66.13%	44.32%	69.27%

Next, MK-means, K-means, FCM, and 1-D SOM algorithms were used to execute experiments on the Wine data set, and the experimental results were measured by the F-measure value criterion. The results are shown in Table 4.

Table 4. F-measure of Wine data are set by MK-means, K-means, FCM, and SOM clustering algorithm

Clustering	MK-means	K-means	FCM	SOM
Class1 F-measure	94.03%	87.32%	84.64%	92.11%
Class2 F-measure	93.12%	85.57%	81.35%	82.93%
Class3 F-measure	96.67%	86.52%	79.84%	85.56%
All F-measure	94.91%	84.12%	88.39%	85.68%

As it can be seen from Table 3 and Table 4, the total F-measure value of the MK-means algorithm is as high as 94.91%, up to 10% higher than the hierarchy method, K-means, FCM, and SOM algorithms. It shows that the MK-means algorithm performs well on dispersed data in the Wine datasets.

In conclusion, the proposed MK-means algorithm, using F-measure as the evaluation standard on the two data sets of Iris and Wine, has shown the superior value on clustering applications through comparing with FCM, K-means, SOM, and the centralized hierarchical clustering.

6 Application of MK-means Algorithm in the Personalized Student Model

In this section, we apply the MK-means algorithm to the student feature data set to verify its effectiveness in education data and explain the final clustering results. From the final results, MK-means shows more sufficient evidence to prove its superior clustering value as an improved K-means algorithm.

6.1 Educational Data EPM

Learning analysis (LA) and education data mining (EDM) have received great attention from educational researchers and technical practitioners in the last ten years [36]. Chen et al. studied online short courses via learning behaviors in an early detection prediction method [37] as the application guidance of our paper.

The educational data used in this paper was UCI Educational Process Mining (EPM), which is a learning analytics data set. The EPM data set was collected from Mehrnoosh Vahdat (University of Genoa in Italy) and Luca Oneto (Holland at the Eindhoven University of Technology), and they researched 115 first-grade engineering specialty undergraduate students of the University of Genoa in 2014. Students were required to be in the digital and electronic learning environment. Learning materials were provided through major browsers, and students were required to solve all kinds of related problems with different difficulties in the learning process. The final data set was recorded by the students, including the six experimental classes and the scores in the six experiments, the final exam results and the test paper were also included.

In summary, the EPM data set is a rigorous, authoritative, and learning feature data set of students consistent with the student model. This paper used two EPM data subsets' experimental activity scores as the final exam results. The final test scores and data subsets correspond to the cognitive model. The session data subset of the experimental activity score is consistent with the knowledge map in this paper.

In this section, the internal evaluation criterion and the contour coefficient were used as the evaluation criteria of our proposed algorithm.

6.2 Application of MK-means algorithm

6.2.1 Finding a Learning Group with Different Cognitive Abilities

The purpose is to cluster the students' cognitive ability and obtain different cognitive ability groups to provide personalized data. Three kinds of clustering algorithms, including MK-means, K-means, and X-means, were used to analyze the final data, and the coefficients are used to analyze the clustering results as shown in Figure 8, Figure 9, and Figure 10. The horizontal axis represents the contour coefficients, and the vertical axis indicates the category. Finally, the clustering results are used to analyze the cognitive ability.

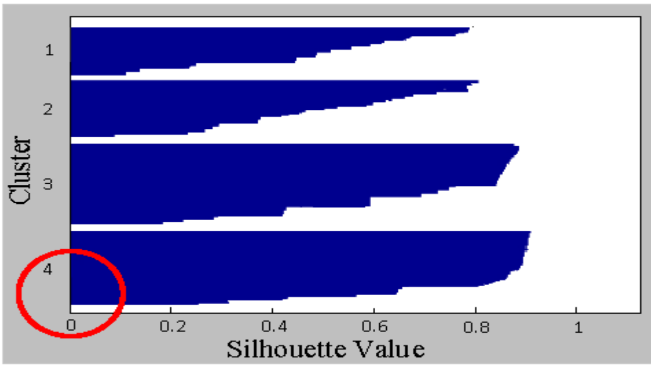


Figure 8. MK-means profile factor

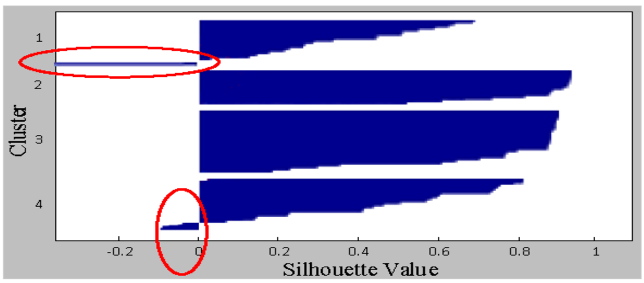


Figure 9. K-means profile factor

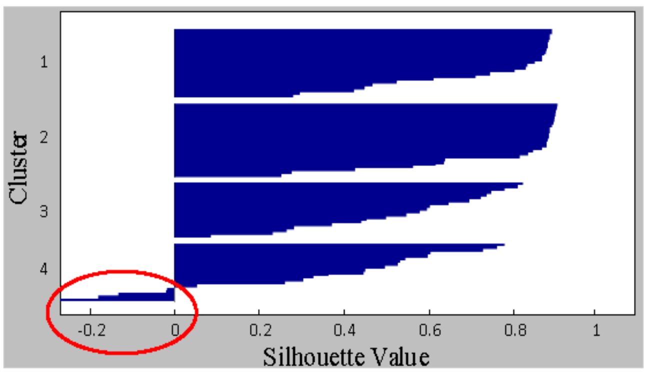


Figure 10. X-means profile factor

In the above clustering algorithms, the X-means algorithm can find the optimal k value automatically. The contour coefficients of the MK-means algorithm are shown in Figure 8. The contour coefficients tend to be 1, and no contour coefficients tend to be -1 of all classes, indicating that all samples are assigned to the optimal category and the MK-means algorithm works well.

Figure 9 is the contour graph of the K-means algorithm. The contour factor of the partial data tending to -1, indicating that the data is not allocated to the best class. So, the K-means algorithm fails in the final data subset of the EPM education.

Figure 10 is the contour coefficient graph of the X-means algorithm, and the contour coefficient of partial data in the graph tends to -1, which means that the X-means algorithm performs poorly in the final subset of EPM educational data.

In the contour coefficient figure of MK-means, K-means, and X-means algorithm, we could find that all categories of contourlet coefficients are close to 1, and there is no category to -1 in the MK-means algorithm. In other algorithms, some numerical and data contour tends to -1, indicating that the data are not assigned to the optimal class. Therefore, the clustering effect of the MK-means algorithm is excellent.

The clustering center obtained by the MK-means algorithm is shown in Table 5, and the clustering results are analyzed by SPSS.

Table 5. Final clustering centers of MK-means

	Memorizing	Understanding	application	Comprehensive	Analyzing	appraising
Class 1	0.49654	0.49654	0.46921	0.42131	0.41287	0.32693
Class 2	0.69791	0.69791	0.68311	0.64941	0.65897	0.66193
Class 3	0.25191	0.25191	0.22316	0.14901	0.14681	0.05302
Class 4	0.88913	0.88913	0.87612	0.91740	0.93191	0.90733

Analysis of the four types' center is shown in Table 5, where the cognitive ability of class 1 is below 0.5 and above 0.30. The cognitive ability of class 2 is between 0.65-0.70, the cognitive ability of class 3 is below 0.25, and the cognitive ability of class 4 is between 0.87-0.93. According to the above four categories, it can be divided into four study groups with different levels to train and improve purposefully.

6.2.2 Finding Different Levels of Knowledge

This experiment's purpose is to grasp the clustering knowledge, which can be obtained from the learning groups with different levels of mastery. On the other hand, it could help students to master the understanding of different plans and give reference to the next learning step. The session data set of the knowledge map conforms to the internal evaluation criteria, so the internal evaluation criterion of the contour coefficient is adopted to evaluate the clustering quality.

We compared the X-means, the K-means, and the improved algorithms of K-means to run the cluster analysis on the session data sets and analyzed the clustering results using the contour coefficient, as shown in Figure 11, Figure 12, and Figure 13. Finally, the clustering results were used to analyze the knowledge grouping.

In the clustering algorithms above, the X-means algorithm can find the optimal k value automatically.

Figure 11, Figure 12, and Figure 13 respectively correspond to the silhouette coefficient of the three algorithms. It can be found that the part of contourlet coefficients of three kinds of algorithm tending to -1.

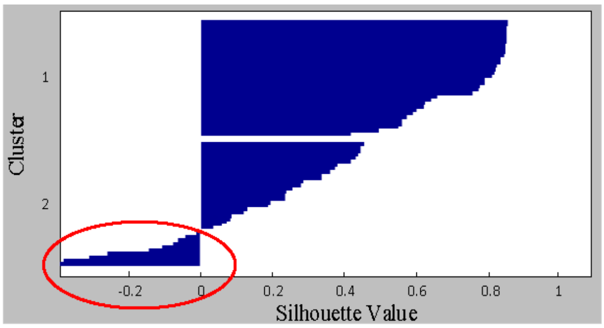


Figure 11. K-means profile factor

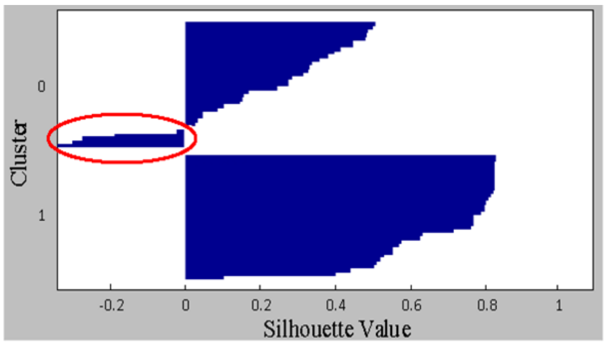


Figure 12. X-means profile factor

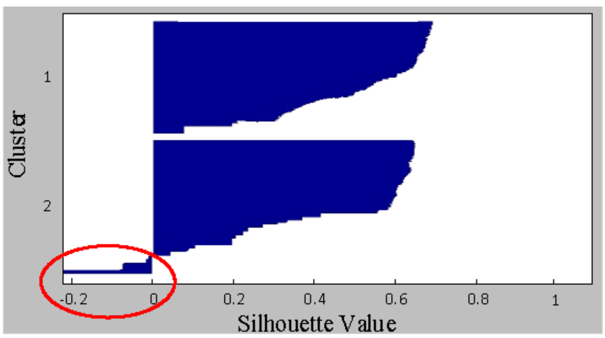


Figure 13. MK-means profile factor

From the figures, it can see in the MK-means algorithm contourlet coefficients, it tends to silhouette value. Therefore, it is proved again that the MK-means algorithm is superior to the other two algorithms.

The session data set represents the knowledge map. The knowledge grouping of the clustering algorithm and the knowledge map is discussed using the best-quality MK-means algorithm.

The MK-means algorithm divides the session into two categories. The clustering centers of the two types are shown in Table 6. Type 1 has 32 students, and type 2 has 38 students. The central point of clustering can explain the student's mastery of knowledge in type 1 is weaker than that of type 2, which can be properly reviewed and consolidated.

Table 6. Clustering centers of MK-means in session

	Session1	Session2	Session3	Session4
type1	2.5	4.5	4	2
type2	3.5	4.5	3.5	2.25

7 Conclusion

In this paper, a novel student personalized e-learning model is proposed. We introduced the K-means algorithm in the model and analyzed a series of problems caused by the random selection of initial centers. In order to optimize the initial centers of clustering, an improved MK-means algorithm, based on cluster-wise weighing co-association matrix mechanism and mean shift theory, was proposed to solve the various problems of K-means. Experimental results of UCI's Iris and Wine datasets demonstrated that the MK-means algorithm outperforms K-means, FCM, and K-means' improved algorithm. From the perspective of the cognitive model and knowledge map, we achieved an excellent explanation of the clustering results by combining the enhanced MK-means algorithm with the student model.

In the future, we'd like to raise more algorithms to apply the emotion model. The new way to make better use of promising techniques to improve the efficiency of online e-learning and students' personalized learning motivation is also under our consideration.

Acknowledgements

This research is supported by the New Generation Artificial Intelligence Science and Technology Major Project of Sichuan Province (2018GZDZX0041), and the Key R & D project of Sichuan Province (2020YFQ 0056, 2020YFG0011, 2020YFG0476, 2021YFG0132).

Reference

[1] S.-C. Wang, B. Cowie, A. Jones, Benefits? or Challenges? University Student Perception of E-Learning, *Journal of Internet Technology*, Vol. 10, No. 5, pp. 505-512, October, 2009.

[2] Q. Alajmi, A. Sadiq, A. Kamaludin, M. A. Al-Sharafi, E-learning Models: The Effectiveness of the Cloud-based E-learning Model over the Traditional E-learning Model, *2017 8th International Conference on Information Technology (ICIT)*, Amman, Jordan, 2017, pp. 12-16.

[3] H. Rodrigues, F. Almeida, V. Figueiredo, S. L. Lopes, Tracking E-learning Through Published Papers: A Systematic Review, *Computers & Education*, Vol. 136, pp. 87-98, July, 2019.

[4] M. Li, S. Cao, Z.-G. Qin, Creation Method and Evolution Evaluation of Concept Knowledge Maps, *Journal of Internet Technology*, Vol. 17, No. 2, pp. 179-189, March, 2016.

[5] F. Rashidi, S. Nejatian, H. Parvin, V. Rezaie, Diversity Based Cluster Weighting in Cluster Ensemble: An Information Theory Approach, *Artificial Intelligence Review*, Vol. 52, No. 2, pp. 1341-1368, August, 2019.

[6] A. L. N. Fred, A. K. Jain, Combining Multiple Clusterings Using Evidence Accumulation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 6, pp. 835-850, June, 2005.

[7] X. Wang, C. Yang, J. Zhou, Clustering Aggregation by Probability Accumulation, *Pattern Recognition*, Vol. 42, No. 5, pp. 668-675, May, 2009.

[8] H. Alizadeh, B. Minaei-Bidgoli, H. Parvin, Cluster Ensemble Selection Based on a New Cluster Stability measure, *Intelligent Data Analysis*, Vol. 18, No. 3, pp. 389-408, April, 2014.

[9] M. Mojarad, S. Nejatian, H. Parvin, M. Mohammadpoor, A Fuzzy Clustering Ensemble Based on Cluster Clustering and Iterative Fusion of base Clusters, *Applied Intelligence*, Vol. 49, No. 7, pp. 2567-2581, July, 2019.

[10] A. Nazari, A. Dehghan, S. Nejatian, V. Rezaie, H. Parvin, A Comprehensive Study of Clustering Ensemble Weighting Based on Cluster Quality and Diversity, *Pattern Analysis and Applications*, Vol. 22, No. 1, pp. 133-145, February, 2019.

[11] H. Parvin, B. Minaei-Bidgoli, A clustering Ensemble Framework Based on Selection of Fuzzy Weighted Clusters in a Locally Adaptive Clustering Algorithm, *Pattern Analysis and Applications*, Vol. 18, No. 1, pp. 87-112, February, 2015.

[12] D. Huang, J.-H. Lai, C.-D. Wang, Ensemble Clustering Using Factor Graph, *Pattern Recognition*, Vol. 50, pp. 131-142, February, 2016.

[13] D. Huang, J.-H. Lai, C.-D. Wang, Robust Ensemble Clustering Using Probability Trajectories, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 28, No. 5, pp. 1312-1326, May, 2016.

[14] E. Akbari, H. M. Dahlan, R. Ibrahim, H. Alizadeh, Hierarchical Cluster Ensemble Selection, *Engineering Applications of Artificial Intelligence*, Vol. 39, pp. 146-156, March, 2015.

[15] R. Yarinezhad, S. N. Hashemi, Exact and approximate algorithms for clustering problem in wireless sensor networks, *IET Communications*, Vol. 14, No. 4, pp. 580-587, March, 2020.

[16] D. Cristofor, D. Simovici, Finding Median Partitions Using Information-Theoretical-Based Genetic Algorithms, *Journal of Universal Computer Science*, Vol. 8, No. 2, pp. 153-172, February, 2002.

[17] L. Franek, X. Jiang, Ensemble Clustering by Means of Clustering Embedding in Vector Spaces, *Pattern Recognition*, Vol. 47, No. 2, pp. 833-842, February, 2014.

[18] E. Weiszfeld, F. Plastria, On the Point for Which the Sum of the Distances to n Given Points is Minimum, *Annals of Operations Research*, Vol. 167, No. 1, pp. 7-41, March, 2009.

[19] R. M. Alguliyev, R. M. Aliguliyev, L. V. Sukhostat, Weighted Consensus Clustering And Its Application to Big Data, *Expert Systems with Applications*, Vol. 150, pp. 113294, July, 2020.

[20] H. Parvin, B. Minaei-Bidgoli, A Clustering Ensemble Framework Based on Elite Selection of Weighted Clusters, *Advances in Data Analysis and Classification*, Vol. 7, No. 2, pp. 181-208, June, 2013.

[21] J. Macqueen, Some Methods for Classification and Analysis of MultiVariate Observations, in: L. M. Le Cam, J. Neyman (Eds.), *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 5.1, University of California Press, 1967, pp. 281-297.

[22] Y.-L. Ji, Y. Zhan, Y. Yang, X. Xu, F.-M. Shen, H.-T. Shen, A Context Knowledge Map Guided Coarse-to-fine Action Recognition, *IEEE Transactions on Image Processing*, Vol. 29, pp. 2742-2752, 2020.

[23] A. Elfaki, A. Aljaedi, Y.-C. Duan, Mapping ERD to Knowledge Graph, *2019 IEEE World Congress on Services (SERVICES)*, Milan, Italy, 2019, pp. 110-114.

[24] R. M. Gagne, *The Conditions of Learning and Theory of Instruction*, Holt, Rinehart and Winston Press of America, 1985.

[25] F. Setiawan, A. G. Prabono, S. A. Khowaja, W. Kim, K. Park, B. N. Yahya, S. L. Lee, J. P. Hong, Fine-grained emotion recognition: fusion of physiological signals and facial expressions on spontaneous emotion corpus, *International Journal of Ad Hoc and Ubiquitous Computing*, Vol. 35, No. 3, pp. 162-178, October, 2020.

[26] Y.-X. Dai, X. Wang, X.-P. Li, P.-B. Zhang, Reputation-driven Multimodal Emotion Recognition in Wearable Biosensor Network, *2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, Pisa, Italy, 2015, pp. 1747-1752.

[27] M. K. Pakhira, A Linear Time-Complexity K-Means Algorithm Using Cluster Shifting, *2014 International Conference on Computational Intelligence and Communication Networks*, Bhopal, India, 2014, pp. 1047-1051.

[28] R.-P. Wang, H. Li, M. Chen, Z.-Y. Dai, M. Zhu, MIC-Kmeans: A Maximum Information Coefficient Based High-Dimensional Clustering Algorithm, in: R. Silhavy (Eds.), *Artificial Intelligence and Algorithms in Intelligent Systems: Proceedings of 7th Computer Science On-line Conference 2018*, Vol. 764, 2019, Springer, Cham, pp. 208-218.

[29] X.-D. Wang, R.-C. Chen, F. Yan, Hendry, K-means Clustering with Feature Selection for Stream Data, *2018 International Symposium on Computer, Consumer and Control (IS3C)*, Taichung, Taiwan, 2018, pp. 453-456.

[30] R. K. Poluru, L. K. Ramasamy, Optimal cluster head selection using modified rider assisted clustering for IoT, *IET Communications*, Vol. 14, No. 13, pp. 2189-2201, August, 2020.

[31] J. Leskovec, A. Rajaraman, J. D. Ullman, *Mining of Massive Datasets, 2nd Edition*, Cambridge University Press, 2014.

[32] K. Fukunaga, L. Hostetler, The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition, *IEEE Transactions on Information Theory*, Vol. 21, No. 1, pp. 32-40, January, 1975.

[33] Y.-Z. Cheng, Mean Shift, Mode Seeking, and Clustering, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 17, No. 8, pp. 790-799, August, 1995.

[34] D. Comaniciu, P. Meer, Mean shift: a robust approach toward feature space analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 5, pp. 603-619, May, 2002.

[35] N. Lethikim, T. Lehoang, T. Vovan, Automatic clustering algorithm for interval data based on overlap distance, *Communications in Statistics-Simulation and Computation*, pp. 1-16, March, 2021, doi: 10.1080/03610918.2021.1900248.

[36] C. Angeli, S. K. Howard, J. Ma, J. Yang, P. A. Kirschner, Data Mining in Educational Technology Classroom Research: Can it Make a Contribution, *Computers & Education*, Vol. 113, pp. 226-242, October, 2017.

[37] W.-Y. Chen, C. G. Brinton, D. Cao, A. Mason-Singh, C. Lu, M. Chiang, Early Detection Prediction of Learning Outcomes in Online Short-Courses via Learning Behaviors, *IEEE Transactions on Learning Technologies*, Vol. 12, No. 1, pp. 44-58, January-March, 2019.

Biographies

Sheng Cao received the Ph.D. degree in Computer Science from the University of Chinese Academy of Sciences, China in 2008. He is currently an Associate Professor in School of Computer Science and Engineering, University of Electronic Science and Technology of China. His current research interests include Intelligent Education and Blockchain.

Songdeng Niu is a graduate student in University of Electronic Science and Technology of China. His research interests are mainly in Clustering, Reinforce-ment Learning and Block-chain.

Guanghao Xiong is a graduate student in University of Electronic Science and Technology of China. His research interests include Clustering and Optical Communication.

Xiaolin Qin received the Ph.D. degree in Computer Science from the University of Chinese Academy of Sciences, China in 2011. He is currently a professor in the University of Chinese Academy of Sciences, China. His current research interests are mainly in Automatic Reasoning, Cluster Intelligence and Intelligent Big Data Analysis.

Pengfei Liu received the Ph.D. degree from University of Electronic Science and Technology of China in 2018. He is currently an assistant researcher with UESTC. His current research interests include Intelligent Education, Multi-Agent System, Swarm Intelligence and Edge Computing.