Collaborative Learning Group Formation with Density Clustering

Mengxiong Zhou[1], Yanming Ye[*1,2,3], Yizhi Ren[3] and Yueshen Xu[2]

[1]*School of Information Engineering, Hangzhou Dianzi University, Hangzhou
310018,P.R.China*
[2]*School of Computer Science and Technology,Zhejiang University, Hangzhou
310018, P.R.China*
[3]*School of Computer Science and Technology, Hangzhou Dianzi University,
Hangzhou 310018, P.R.China*
[1]*zhoumx@hdu.edu.cn,* [2]*yeym@hdu.edu.cn,* [3]*renyz@hdu.edu.cn,*
[4]*xyshzjucs@zju.edu.cn*

Abstract

The formation of the collaborative learning group is a challenging problem, especially the principle construction and multi-objective optimization. In this paper, we propose an improved density clustering algorithm, and further implement the group formation based on the proposed algorithm. The proposed algorithm first constructs the distance matrix for learners, and clusters learners into different groups. After the random selection of the initial cluster centers, our algorithm achieves superior clustering result by iteratively adjusting the clustering centers. We conduct sufficient experiments and the experimental results demonstrate that the proposed algorithm well satisfies our aim, that is, inter-group heterogeneity maximization and intra-group homogeneity maximization.

Keywords: collaborative learning, group formation, density clustering, inter-group heterogeneity, intra-group homogeneity

1. Introduction

Group formation is one of the key problems in computer supported collaborative learning (CSCL for short). The key task is to effectively cluster learners into different groups based on their features. Such a clustering is the first step of computer supported collaborative learning, and also the necessary step to improve the learning effectiveness. There exist two traditional methods of group formation, including random selection and enumeration. The random selection method ignores the feature attribute of learners, so it is hard to achieve the satisfactory optimization result. The enumeration method has high time complexity. Under the case that there are a large number of learners or attribute features, the enumeration method usually gives the result in a long time. In recent years, some researchers proposed several group formation methods. In [6], the authors proposed an extended particle swarm optimization algorithm, which was based on learners' understanding level and interest. In [5], the authors adopted an enhanced genetic approach, which selected two features, i.e., the concept quantity of a certain class and the predicted scores. In the method proposed in [7], the authors focused on estimating the basic features and personalized features of a learner, and adopted an extended ant colony optimization algorithm for clustering. The mentioned papers only studied a few or very limited learner features. Based on the existing works, the authors in [8] built an improved genetic algorithm, which could leverage a large number of features. However, this algorithm ignores the difference among different features, so it fails to satisfy the multi-objective group formation. To solve these problems, the paper [9] proposed a collaborative learning group formation algorithm extended from the bacterial foraging algorithm, and the paper [10] proposed a collaborative learning group formation algorithm based on improved ant

---

colony optimization. However, the two algorithms can only be used in homogeneous group formation, but cannot be used in heterogeneous group formation, the case that is studied in our paper.

Generally speaking, the homogenous group is effective for learners achieving their plan. In contrast, in most cases, the heterogeneous group is favorable for enhancing complementation among learners. This paper enhances the density clustering algorithm that is presented in paper [13]. In this paper, we cluster learners into different groups using the enhanced algorithm, and further select learners from clusters to form the final groups, which is based on the grouping quality. The proposed algorithm has the capability of involving multiple features. The importance of different features is measured by the predefined weights, so our algorithm is applicable for multiple optimization problems. Also, our algorithm is capable of taking sparse features as input, which gives the chance to improve the importance of personalized features in collaborative learning, and improve the efficiency.

2. Problem Description and Definition

2.1 Problem Description

In collaborative learning, we have an aim to improve the competence for a certain learner, and also improve the balanced development for all learners. To achieve this aim, we need to keep the following principle, i.e., inter-group heterogeneity maximization and intra- group homogeneity maximization.

For inter-group heterogeneity, the difference among learners in a group should be large so that they can fully leverage the complementary and individual advantage. For intra-group homogeneity, it requires that the difference among group members in different groups should be similar. According to these two requirements, first, it is necessary to determine the number of learners in a group. Second, after clustering, every learner in every group should have a counterpart that has similar features in other groups. In our proposed group formation algorithm, we first partition learners by clustering, and then select suitable learners to form groups. Our algorithm can satisfy the mentioned two requirements.

2.2 Definition

Definition 1: The group space L is a quad tuple $L = (S, C, V, F)$, where S is the set of all learners, C is the set of learners' features, V is the value domain of feature attributes and $F : S \times C \rightarrow V$ is the feature function.

Since different feature attributes have different value range, it is necessary to map the values into a unified transformed space according to a standard interval, to make different feature attributes comparable.

Definition 2: Assuming that in the group space, the number of learners is N(N=|S|), and the size of the feature attribute set is T(T=|C|), then the feature vector of the ith learner is defined as $V_i = [f(S(i), C(1))\ f(S(i), C(2)) ... f(S(i), C(T))]'$ where $f(S(i), C(j))$ represents the value of the jth feature attribute of the ith learner.

Definition 3: The covariance matrix of the feature vectors $\{V_i \mid i \in [1, N]\}$ is

$$M = \begin{bmatrix} \text{cov}(V_1, V_1) & \text{cov}(V_1, V_2) & ... & \text{cov}(V_1, V_N) \\ \text{cov}(V_2, V_1) & \text{cov}(V_2, V_2) & ... & \text{cov}(V_2, V_N) \\ ... & ... & ... & ... \\ \text{cov}(V_N, V_1) & \text{cov}(V_N, V_2) & ... & \text{cov}(V_N, V_N) \end{bmatrix}$$

where $\mathrm{cov}\left(V_i, V_j\right) = \dfrac{\sum\limits_{k=1}^{N}(V_{k,i} - \overline{V}_{C(i)})(V_{k,j} - \overline{V}_{c(j)})}{N-1}$ and $V_{k,i} = f(S(k),\ C(i))$ and $\overline{V}_{c(i)}$ is the average value of feature $C(i)$.

Definition 4: The learner distance is defined as the difference between the feature attributes of two learners. The distance is computed by Mahalanobis distance with adjusting factor, as shown below.

$$d(x,y) = \sqrt{(V_x - V_y)^{'} * M^{-1} * \Gamma * (V_x - V_y)}$$

$$\Gamma = \begin{bmatrix} \lambda_1 & 0 & ... & 0 \\ 0 & \lambda_2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & \lambda_T \end{bmatrix}$$

where x and y are the identification labels of two learners in the group space of learner set S, and λ_i is the weight of the ith feature attribute Ci in difference computation. So λ_i can be regarded as the adjusting factor of Ci ($0 < \lambda_i \leq 1$).

Definition 5: The learner distance matrix (notated as MD) is defined as the matrix that records the distance between each pair of learners in the group space, and represented as follows.

$$MD = \begin{bmatrix} d(1,1) & d(1,2) & ... & d(1,N) \\ d(2,1) & d(2,2) & ... & d(2,N) \\ ... & ... & ... & ... \\ d(\mathrm{N},1) & d(N,2) & ... & d(N,N) \end{bmatrix}$$

Since the distance function d(x,y) satisfies the symmetry property, the distance matrix can be simplified into a lower triangular matrix, which can reduce the storage and computation complexity.

Definition 6: The density follows the Gaussian distribution with the general formula as following:

$$f_D\left(x,k\right) = \frac{e^{(-(x-(\overline{x})_k)^2/2(\theta k)^2)}}{\theta k \sqrt{2\pi}}$$

Where the value of constant θ is give optimum results with value $\theta = 3.5$ for all cases (that is determined empirically by [13]) and k is the group size that is a predefined certain value in the case.

Definition 7: The group g is the independent group space L_g, and the learner distance matrix MD_g is represented as the adjacent matrix of the connected graph G. Then the grouping quality of group g is defined as the weighted average of the minimum spanning tree of G. The group quality is used to measure the quality of grouping, which is notated as Q(g).

There exist a large number of personalized sparse features after the collection of feature attributes of learners. We need to use a quite large feature matrix to store the sparse features, which lower the computation efficiency and increase the difficulty for the distance computation of learners. In most cases, the contributions of the personalized features to the grouping cannot be treated equivalently with the basic features. So in a certain application, we just need to know whether a personalized feature exists or not, and do not need to use the specific value. In this paper, we map the personalized features into the binary value range, i.e., {0, 1}.

Definition 8: Assuming that the number of groups in the group space is R, and the quality of each group g is Q(g) (R$\geq$g$\geq$1), then the variance of group quality is notated as *DG(L,R)*.

3. The Proposed Algorithm

Our proposed algorithm includes three parts, i.e., pre- processing, clustering and group formation, which are shown in Figure 1. The pre-processing component produces the learner distance matrix, which is the input of the clustering component. After clustering, we can get the learner clustering group, which is used to construct the group formation. Finally, we get the learner group set.
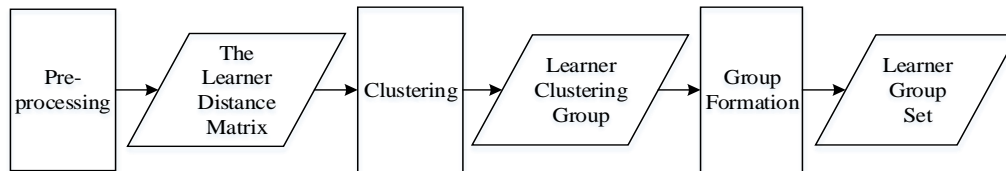

Figure 1. The Framework of our Proposed Algorithm

3.1 Pre-processing

The task of the pre-processing component is to generate the learner distance matrix, which is the input of the subsequent components, i.e., clustering and group formation. The flow chart of pre-processing is given in Figure 2.
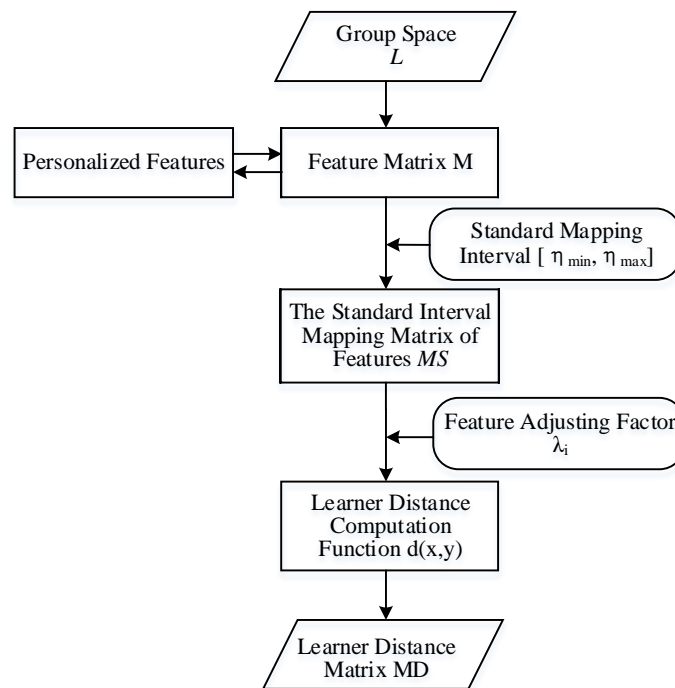

Figure 2. The Flow Chart of Pre-Processing

3.2 Clustering Analysis

The proposed clustering algorithm is conducted in the clustering component. The algorithm takes the learner distance matrix that is generated from the pre-processing component (see Section 3.1) as the input. Also, this component needs the cluster size cn and group size k ($cn=N/k$) as input. Here the value of k is equal to the k that is used in the computation of k nearest neighbor density. In this component, the cumulative density distribution of the learners is computed firstly by following steps:

1). Every learner forms a cell and a cell only contains one learner.

2). Find the k-nearest neighbor cells of the cell I.

3). For every neighbor cell J, calculate distance d between centroid of cells J and I.

4). Calculate the density value t for J using this distance d.

5). Increment the density value at J by t.

6). Repeat sets 1-5 for all learners.

Then, the cell with largest density value is selected and whose learner can be recognized as the center of one of the result clusters.

The clusters in the cluster group are ranked according to the order of their generation, and labeled as *C1, C2···Ck*. After the new cluster is formed, the algorithm will delete the rows and columns of the learners that are contained in the new clusters from the learner distance matrix MD, to form the new MD. If the size of MD is larger than *cn*, then the algorithm starts the next iteration, from the step of learner density computation. Otherwise, the corresponding learners will be as the final cluster group, and the algorithm ends.

3.3 Group Formation

After we get the cluster group using the proposed clustering algorithm (see Section 3.2), in some cases, the cluster center density of the last several clusters are likely to be abnormal. Such abnormal cases can be used to improve the group density.

In this component, the proposed algorithm takes the learner distance matrix (see Section 3.1) and the cluster group set $\{Ci: 1 \leqslant i \leqslant k\}$ (see Section 3.2) as input. With a clustering adjusting parameter ξ, the algorithm computes the density (DC_i) and the mean (DM) of each cluster group center. The set $\{Ci\}$ is divided into two sub-sets according to whether DCi is larger than $DM \times \xi$. The two sub-sets are the basic cluster set $\{Cl\}$ ($DC_l \geqslant DM \times \xi$ and $|Cl|=cn$) and adjusting cluster set $\{Cg\}$ ($DC_g < DM \times \xi$ or $|Cg| < cn$). The principles of group formation are as follows.

1. Each learner can be only assigned to one group.

2. Each cluster contains at least *k* learners.

3. The toppest $|\{Cl\}|$ members in group g_i consist of those learners in which one learner is selected from one cluster in the basic cluster set $\{Cl\}$.

4. The remaining members of group g_i consist of those learners in which one learner is selected from one cluster in the adjusting cluster set $\{Cg\}$.

5. The group quality in the final groups satisfies that the variance is the minimum or is lower than the pre-defined threshold.

4. Experiment and Evaluation

We evaluate the proposed algorithm on a discipline competition in a school. We select 200 students from 8 disciplines as the learner set in the data source. Also, we collect the feature attributes and the corresponding scores, including 9 feature attributes and 7 personalized feature attributes reported by students. In detail, the 9 feature attributes include interest, motivation, comprehension, experience, communication skill, foreign language skill, computer skill, invocation capability and management level. All of 7 personalized feature attributes are combined into one unified personalized attribute with binarization. So there are totally 10 feature attributes as the source data of our experiments. The standard interval $[\eta_{min}, \eta_{max}]$ is set to [1, 10]. The *k* is set to *k*=6, meaning that each group contains 6 learners.

For the common discipline competitions, we set all adjusting parameters to 1 ($\lambda_i = 1$). In contrast, for the international discipline competitions, since the foreign language proficiency of a learner is required to reach to some extent, we set $\lambda_i = 0.6$ ($i \neq 6$) and $\lambda_6 = 1$. Note that, the setting of adjusting parameters is closely related to the feature selection in the group space, and also is related to the grouping

purpose. In the real-world case, the exact values of adjusting parameters are usually hard to be precise, and hard to be versatile in all disciplines or in all sectors. Thus, in most cases, an adjusting parameter is first set to a preliminary value according to the experiences in a certain area. Then, such a preliminary value continues to be improved in the real applications to achieve the optimal value or be set into an optimal value interval.

We give a detailed example of the value setting of λ_i, i.e., $\lambda_i=1$ and $\lambda_i=0.6\,(i \neq 6)$. The distribution of multi-dimensional scaling (**MDS** for short) is shown in Figures 3(1) and 3(2), which is generated from the relative location computed from the learner relative distance. We can observe that in the two cases, the distributions of the relative location of learners are clearly different. The location space in Figure 5.1 is smaller to some extent. One reason is that the adjusting parameters change the relative distance of learners. In our case, such a change makes the relative distance of two learners smaller.
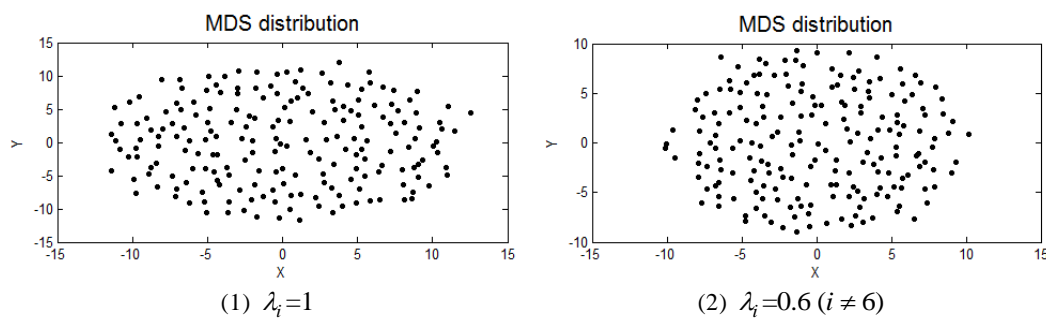

(1) $\lambda_i=1$ (2) $\lambda_i=0.6\,(i \neq 6)$

Figure 3. The Location Distribution Graph of Learners

In the case that the group adjusting parameter $\xi=0.95$, the clustering results are given in Tables 1 and 2. We can observe that for different settings of adjusting parameters, the cluster group centers and group center density are different. In the cases of $\lambda_i=1$ and $\lambda_i=0.6\,(i \neq 6)$, the cluster center density DM is 0.2010 and 0.2559, respectively. Correspondingly, the adjusting cluster set is {C6,C7} and {C5,C7} respectively.

Table 1. The Clustering Result ($\lambda_i=1$)

Cluster Num.	Cluster Center Num.	Cluster Size	Group Center Density DC_i	$DC_i \geq DM \times \xi$ and $|C_i|=cn$
C1	163	33	0.2217	Yes
C2	197	33	0.2037	Yes
C3	117	33	0.1977	Yes
C4	78	33	0.2012	Yes
C5	119	33	0.2110	Yes
C6	107	33	0.1864	No
C7	170	2	0.1853	No

Table 2. The Clustering Result ($\lambda_i = 0.6\ (i \neq 6)$)

Cluster Num.	Cluster Center Num.	Cluster Size	Group Center Density DC_i	$DC_i \geq DM \times \xi$ and $|C_i| = cn$
C1	163	33	0.2802	Yes
C2	108	33	0.2577	Yes
C3	78	33	0.2513	Yes
C4	119	33	0.2688	Yes
C5	107	33	0.2339	No
C6	10	33	0.2574	Yes
C7	117	2	0.2416	No

(1) $\lambda_i = 1$ (2) $\lambda_i = 0.6\ (i \neq 6)$
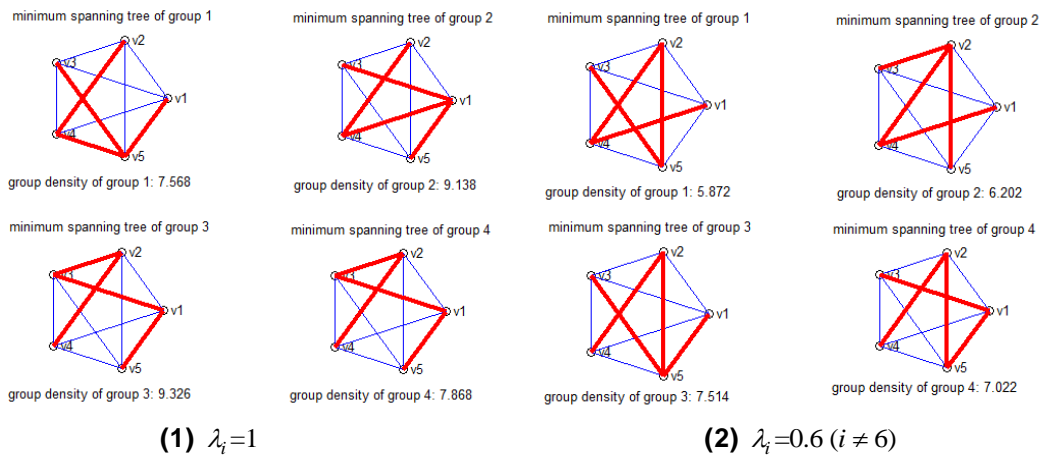
Figure 4. The Minimum Spanning Tree Generated from the Basic Group Set

(1) $\lambda_i = 1$ (2) $\lambda_i = 0.6\ (i \neq 6)$
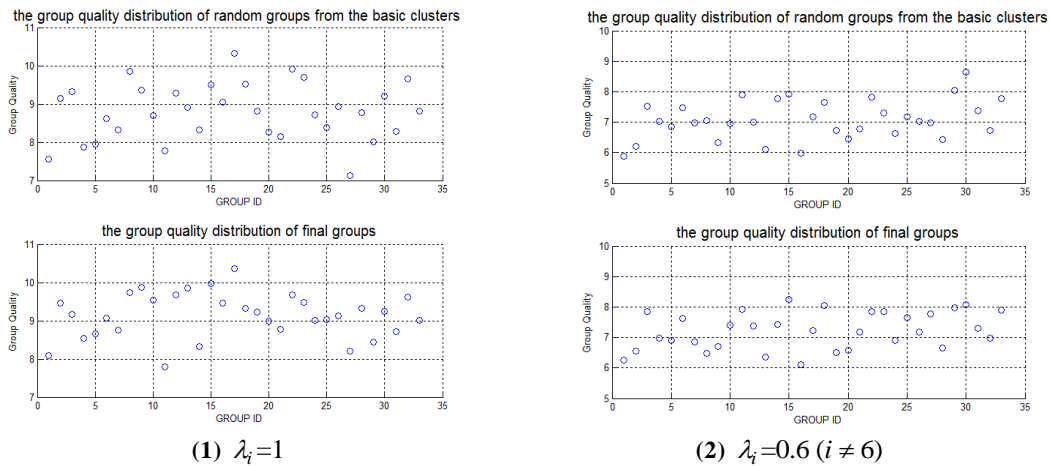
Figure 5. The Group Quality Distribution of the Random Groups in the basic Clusters and Final Groups

We select a learner from each cluster in the basic cluster set, to form the random group of basic clusters ($cn=33$). After that, the learners in each group are regarded as a node in an undirected graph, and the distance between two learners is treated as the weight in the edge. Then, we generate the minimum spanning tree and computer the group quality. The minimum spanning trees are shown in Figure 4, where we give an example of 4 groups selected from each case. The distribution of the group

quality is shown in Figure 5. The variances of the group quality under the two cases are 0.5498 ($\lambda_i = 1$) and 0.4212 ($\lambda_i = 0.6\ (i \neq 6)$).

Next, we select one learner from one cluster in the adjusting cluster set {Cg}, and add the selected learners into the existing groups, so that the final result satisfies that the variance of group quality of the final groups can be reached at the minimum or be smaller than a predefined value. We can use the combinatorial optimization algorithm to get the group result. In this paper, we give an alternative algorithm. In this algorithm, for the groups with low quality, the algorithm adds the learners with a large value of δ_i. For the groups with high quality, the algorithm adds the learners with a small value of δ_i.

5. Conclusion and Future Work

In this paper, we propose an algorithm for heterogeneous group formation based on the group formation principle in collaborative learning. The proposed algorithm first utilizes the extended density-based clustering algorithm to cluster learners. Then, the algorithm divides the cluster group into two sub-sets, i.e., basic cluster set and adjusting cluster set. The basic cluster set randomly select learners to form the basic group. The adjusting cluster set is to adjust the basic group, and form the final heterogeneous group. We conduct the experiment in a real-world dataset. The experimental results show that the proposed algorithm can fully reach the demand, i.e., inter-group heterogeneity maximization and intra- group homogeneity maximization. Also, our algorithm supports the formation of heterogeneous group in collaborative learning.

In the future, we will focus on studying the impact of the setting of the distance adjusting parameter on the multi-objective optimization problem. We also plan to study the impact of the group adjusting parameter on the improvement of clustering performance, and find an effective clustering algorithm in group set building and combinatorial optimization algorithm in group adjusting.

Acknowledgments

References

[1] L. Lipponen, "Exploring foundations for computer- supported collaborative learning", Proceedings of the Conference on Computer Support for Collaborative Learning, vol.65, no.3, (2002), pp. 156-162.

[2] J. Li, X. Wang, K. Li, "The research on Design and Development of Collaborative Group Formation System in CSCL", Journal of South China Normal University (Natureel Science Edition), vol.3, no.3 , (2007), pp. 1-12.

[3] M. Huxham, R. Land, "Assigning Students in Group Work Projects．Can we do Better Than Random?", Innovations in Education and Teaching International, vol.37, no. 1, (2000), pp. 17-22.

[4] Y. Lou, P. C. Abrami, J. C. Spence, Within-class Grouping, "A Meta-analysis. Review of Educational Research", vol.66, no.4 , (1996), pp. 423-458.

[5] G. J. Hwang, P. Y. Yin, C. W. Hwang, et al, "An Enhanced Genetic Approach to Composing Cooperative Learning Groups for Multiple Grouping Criteria", Educational Technology & Society, vol.11, no.1 , (2008), pp. 148-167.

[6] S. Graf, and R. Bekele, "Forming heterogeneous groups for intelligent collaborative learning systems with ant colony optimization", Proceedings of International Conference on Intelligent Tutoring Systems, Jhongli, Taiwan, June, (2006), pp. 217-226.

[7]   Y. T. Lin，Y. M. Huang，S. C. Cheng "An automatic group composition system for composing collaborative learning groups using enhanced particle swarm optimization", Computers & Education, vol.55, no.4, (2010), pp. 1483-1493.
[8]   J. Moreno, D. A. Ovalle, R. M. Viccari, "A genetic algorithm approach for group formation in collaborative learning considering multiple student characteristics", Computers and Education, vol.58, no.1 , (2012), pp. 560-569.
[9]   Z. Sang, J. He, "Grouping Algorithm for Collaborative Learning Based on Improved Bacterial Foraging", Computer Engineering, vol.40, no.10 , (2014), pp. 137-142.
[10]   H. Hu, J. He, "Research of Composing Cooperative Learning Group based on Enhanced Ant Colony Optimization Algorithm", Computer Engineering and Applications, no.13 , (2014), pp. 137-141.
[11]   B. A. Nijstad, C. K. W. Deu, "Creativity and Group Innovation", Applied Psychology, vol.51, no.3, (2002), pp. 400–406.
[12]   E. Martín, P. Paredes, "Using Learning Styles for Dynamic Group Formation in Adaptive Collaborative Hypermedia Systems", Proceedings of International Workshop on Adaptive Hypermedia and Collaborative Web-based Systems, (2004), pp. 188-198.
[13]   S. Khalid, S. Razzaq,"TOBAE: A Density-based Agglomerative Clustering Algorithm", Journal of Classification, no.32, (2015), pp.1-27.
[14]   C. Guo, "Research on Group Cooperative Learning (Dissertation)", Shandong Normal University, (2003).

Authors

Mengxiong Zhou is working as a lecturer at the Hangzhou Dianzi University, College of information engineering, computer department, Hangzhou,CHINA. His areas of interest include Software engineering, artificial intelligence.
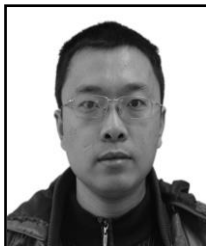
Yanming Ye received Ph.D degree from Zhejiang University and now he is an associate professor and tutor for graduates in Hangzhou Dianzi University. His main research interests include workflow, social computing and artificial intelligence.

Yizhi Ren is an associate professor in the college of Computer, Hangzhou Dianzi University. He received the doctor degree in computer science and technology from the Dalian University of Technology in 2011. His research interests include Network Security, Social Computing and Evolutionary game.
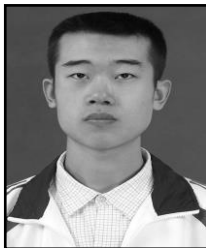
Yueshen Xu is currently a fifth-year Ph.D. student, in the College of Computer Science and Technology of Zhejiang University. He was a visiting Ph.D. student in the Department of Computer Science, University of Illinois at Chicago. He has published several papers in international conferences and journals, including WISE, WAIM, ESWA, EAAI, IJWGS, ACM DEBS, SCC, AMIS, IJSEK, TWSJ, etc. His research interests include Recommender System, Text Mining and Natural Language Processing.