

# An Improved Genetic Algorithm Approach for Optimal Learner Group Formation in Collaborative Learning Contexts

Ya-Qian ZHENG, Jia-Zhi DU, Hai-Bo YU, Wei-Gang LU & Chun-Rong LI\*

*Department of Educational Technology, Ocean University of China, China*

\*lchh@ouc.edu.cn

**Abstract:** Collaborative learning has been widely used in educational contexts. Considering that group formation is one of the key processes in collaborative learning, the aim of this paper is to propose a method to obtain inter-homogeneous and intra-heterogeneous groups. In this method, the group formation problem is translated into a combinatorial optimization problem, and an improved genetic algorithm approach is also proposed to cope with this problem. To evaluate the proposed method, we carry out computational experiments based on eight datasets with different levels of complexity. The results show that the proposed approach is effective and stable for composing inter-homogeneous and intra-heterogeneous groups.

**Keywords:** Collaborative learning, group formation, combinatorial optimization problem, genetic algorithm

## 1. Introduction

Collaborative learning (CL) is an instructional strategy which allows students to work together in small groups toward a common learning goal (Dillenbourg, 1999). To apply this strategy, the initial task is assigning students into groups, which is a crucial process. Previous research has indicated that well-structured CL groups allow for a good interaction among members, which is fundamental to obtain appropriate learning results (Sadeghi and Kardan, 2015). Building inter-homogeneous and intra-heterogeneous groups based on student characteristics has been considered as an effective approach for the formation of well-structured groups (Dascalu et al., 2014; Moreno, Ovalle and Vicari, 2012). However, when the number of students and characteristics under consideration is very large, it is almost impossible for an instructor to organize optimal learning groups that meet multiple criteria. Therefore, we translate the group formation problem into a combinatorial optimization problem. In order to cope with this problem, an improved genetic algorithm is presented.

## 2. Problem Description and Mathematical Formulation

In this section, the group formation problem is formally described and mathematically formulated. Assume that there is a set of  $n$  students in a class  $S$ ,  $S = \{s_1, \dots, s_i, \dots, s_n\}$ , which are required to be divided into  $g$  groups  $G$ ,  $G = \{G_1, \dots, G_k, \dots, G_g\}$ . Each group  $G_k$  is made up of a  $z_k$  number of students. Let  $C = \{c_1, \dots, c_u, \dots, c_{p+q}\}$  be the set of all required attributes for the grouping of students, where  $p$  and  $q$  are the number of student characteristics for achieving inter-homogeneous and intra-heterogeneous groups. Let  $\underline{v}_{iu}$  denote the value of attribute  $c_u$  associated with student  $s_i$ , normalized between 0 and 1. Let  $\bar{v}_u$  denote the mean value of attribute  $c_u$  associated with all the participated students, and let  $\bar{v}_{k,u}$  represent the mean value of attribute  $c_u$  associated with students in group  $G_k$ .

For inter-homogenous groups (all the groups have similar compositions), the sum of the squared differences ( $F_{homo}$ ) to measure such homogeneity with regard to the  $p$  characteristics between each group and the whole sample is calculated as:

$$F_{homo} = \frac{\sum_{k=1}^g \sum_{u=1}^p w_u (\bar{v}_u - \bar{v}_{k,u})^2}{\sum_{u=1}^p w_u} \quad (1)$$

where  $w_u$  is the importance degree of attribute  $c_u$ .

For intra-heterogeneous groups (each group with members who are as different among themselves as possible),  $F_{hete}$  indicates the heterogeneity within groups is computed as:

$$F_{hete} = \sum_{k=1}^g \sum_{i=1}^{z_k-1} \sum_{j=i+1}^{z_k} \left( 1 - \frac{\sum_{u=1}^n w_u |z_{iu} - z_{ju}|}{\sum_{u=1}^n w_u} \right) \quad (2)$$

In order to achieve the optimal groups, the mathematical model is defined as follows:

$$Min F = \frac{w_{homo} \times F_{homo} + w_{hete} \times F_{hete}}{w_{homo} + w_{hete}} \quad (3)$$

Subject to

$$s_i \in G_k, i=1,2, \dots, \quad (4)$$

$$n; k=1,2, \dots, g$$

$$G_x \cap G_y = \emptyset, x=1,2,\dots, g; y= \quad (5)$$

$$1,2, \dots, g; x \neq y \quad (6)$$

$$|z_x - z_y| \leq 1, x=1,2, \dots, g; y=1,2, \dots, g; x \neq y$$

where  $w_{homo}$  and  $w_{hete}$  are the importance degrees of inter-homogenous and intra-heterogeneous grouping respectively. The objective function (3) is to achieve the optimal groups. Constraint (4) and constraint (5) ensure that each student can be assigned exactly into one group. Constraint (6) guarantees that the size difference of any two groups is not more than one.

### 3. An Improved Genetic Algorithm for the Group Formation Problem

To solve the group formation problem, an improved genetic algorithm is proposed. Genetic algorithms (GAs) have been successfully applied to solve a variety of combinatorial optimization problems. However, previous studies reported that using the standard genetic algorithm (SGA) easily led to unsatisfactory searching behaviors (e.g. premature convergence) when faced with large-scale and complex real-world problems (Wang and Tang, 2011). In order to overcome the shortcomings of the SGA, we present an improved genetic algorithm (IGA) to solve the group formation problem.

The IGA starts with a set of initial feasible solutions which are represented by adopting an integer permutation encoding scheme. Thus, each solution is encoded as a list with a length equal to  $n$  (i.e. the list is a permutation of  $n$  students). We generate the initial population by a random method which ensures the diversity of the population and improves the convergence speed and the quality of final solutions. The performance of each solution is evaluated by a fitness function that corresponds to the objective function of the optimization problem. In general, the chromosomes with greater fitness value are more likely to be selected to survive and replicate. For the objective function is defined as a minimization problem and the value of  $f$  is always positive (Section 2), the fitness function can be calculated as follows:

$$Fitness = 1/F \quad (7)$$

In each generation, some individuals are selected by applying the stochastic universal sampling selection, and then GA employs the partially mapped crossover and the swap mutation to generate new offspring chromosomes. Besides the traditional crossover and mutation operators, the simple inversion mutation operator is also performed, which is used to increase the diversity of the population. After the genetic operation, an additional elitist reinsertion strategy is used to prevent the loss of good information and fill the generation gap. The elitist strategy guarantees that the best chromosomes always survive intact from one generation to the next.

### 4. Computational Experiments

To evaluate the effectiveness and efficiency of the proposed algorithm, eight datasets with different levels of complexity are generated randomly for the computational experiments. We evaluate the performance of the IGA approach by comparing it with another two competing algorithms, the exhaustive method (EM) and the random method (RM). Table 1 show the results obtained by the EM, the RM and the IGA for the 8 datasets (all the three methods are implemented in MATLAB). Columns Obj. and T report, respectively, the average value of the objective function during ten runs and the average computational time in seconds.



The results show that the EM only obtains optimal solutions for the first four datasets, and the executing time increases significantly as the number of students grows, so it is practically infeasible for solving the group formation problem. Unlike the EM, the RM and the IGA can find solutions for all 8 datasets within an acceptable time. Although the computation time of the RM is the shortest, the value of objective function obtained by the RM is the highest. For the IGA, it also can obtain the optimal solutions for the first four datasets and the quality of the solutions found by the IGA is always better than the RM. Therefore, the IGA is a more effective method which can achieve a satisfactory solution for a reasonable computational time by comparison with the EM and the RM.

**Table 1: Results obtained by the exhaustive method, the random method and the proposed method.**

Dataset	Exhaustive Method		Random Method		Proposed Method	
	Obj.	T(s)	Obj.	T(s)	Obj.	T(s)
1	4.4107	0.0458	4.4635	<0.0001	4.4107	0.1894
2	6.1213	0.2928	6.6010	<0.0001	6.1213	0.5115
3	7.5917	60.5364	8.4560	<0.0001	7.5917	1.5994
4	8.6020	29479.7019	10.3985	<0.0001	8.6020	2.6686
5	N/A	N/A	17.1713	0.0001	14.9955	7.1883
6	N/A	N/A	40.7732	0.0002	34.4190	42.3692
7	N/A	N/A	82.2104	0.0004	64.8140	165.5741
8	N/A	N/A	123.3962	0.0006	101.1219	727.5106

N/A: not available.

We analyze the robustness of these two stochastic methods (i.e. the RM and the IGA) by calculating the standard deviation during ten runs on each dataset to measure the dispersion of the value of objective function. The variation of the standard deviation corresponding to the IGA fluctuates between 0.0000 and 0.0548, while that corresponding to the RM fluctuates between 0.0458 and 3.6469. Thus, we consider that the IGA is more stable to solve the group formation problem with different sizes.

## 5. Conclusion

In this paper, we propose a method based on an improved genetic algorithm to solve the group formation problem. To evaluate the performance of the proposed method, a series of computational experiments are conducted. The simulation results indicate that the proposed method is an effective and stable to method to compose inter-homogeneous and intra-heterogeneous groups.

## Acknowledgements

This work has been supported by the MOE (Ministry of Education in China) Project of Humanities and Social Sciences (No. 16YJC880057) and the Fundamental Research Funds for the Central Universities (No. 201564041).

## References

- Dascalu, M. I., Bodea, C. N., Lytras, M., De Pablos, P. O., & Burlacu, A. (2014). Improving e-learning communities through optimal composition of multidisciplinary learning groups. *Computers in Human Behavior*, 30, 362-371.
- Dillenbourg, P. (1999). What do you mean by collaborative learning. *Collaborative-learning: Cognitive and computational approaches*, 1, 1-15.
- Moreno, J., Ovalle, D. A., & Vicari, R. M. (2012). A genetic algorithm approach for group formation in collaborative learning considering multiple student characteristics. *Computers & Education*, 58(1), 560-569.
- Sadeghi, H., & Kardan, A. A. (2015). A novel justice-based linear model for optimal learner group formation in computer-supported collaborative learning environments. *Computers in Human Behavior*, 48, 436-447.
- Wang, L., & Tang, D. B. (2011). An improved adaptive genetic algorithm based on hormone modulation mechanism for job-shop scheduling problem. *Expert Systems with Applications*, 38(6), 7243-7250.