Applying Group Formation in Practice on a Brazilian Postgraduate Course

Giuseppe Fiorentino[1], Péricles Miranda[1], André Nascimento[1],
Ana Paula Cavalcanti Furtado[1,2] and Rafael Ferreira Mello[1,2]

[1]*Departamento de Computação, Universidade Federal Rural de Pernambuco, Brazil*
[2]*Cesar School, Brazil*

Abstract

Group formation is a critical activity in the collaborative learning process. The literature presents several automatic algorithms to perform this task. However, it fails to compare the results concerning experienced instructors' decisions. In this context, this paper proposes the application of an automatic group formation algorithm in real-world settings. The proposed approach was compared with a manual approach performed by an instructor with ten years of experience on this task. The results proved the potential of the proposed approach as it reached more than 80% of similarity with the groups formed manually by the instructor. The practical implications of the proposed approach are further discussed.

Keywords

Group Formation, Optimization, Collaborative Learning

1. Introduction

There are currently a large number of studies relating to collaborative learning methodology in an educational context [1, 2]. However, with a limited amount in relation to studies referring to the automatic formation of groups [3], even though the literature shows the importance of creating heterogeneous and homogeneous groups, in relation to the characteristics of the students, in the learning process of each student [4, 1, 5**?** , 6]. However, this is a complex problem, as it considers multiple individual characteristics, causing a complexity proportional to the number of students enrolled in a course [7].

The problem of forming groups, due to its complexity, depends on intelligent algorithms so that the process of grouping students with different profiles is effective and results in even more effective learning [8, 9]. Among all the existing approaches, the one based on evolutionary algorithms is the one with the best results [10, 11]. Some authors [12, 13, 14] map this problem into a multi-objective problem with a focus on evaluating different criteria for each group. Others [15, 16], which adopted the adoption of the multi-objective optimization algorithm

LALA'21: IV LATIN AMERICAN CONFERENCE ON LEARNING ANALYTICS - 2021, October 19–21, 2021, Arequipa, Peru
fiorentinogiuseppebcc@gmail.com (G. Fiorentino); pericles.miranda@ufrpe.br (P. Miranda); andre.camara@ufrpe.br (A. Nascimento); anapaula.furtado@ufrpe.br (A. P. C. Furtado); rafael.mello@ufrpe.br (R. F. Mello)
0000-0002-5016-5458 (G. Fiorentino); 0000-0002-5767-7544 (P. Miranda); 0000-0002-9333-3212 (A. Nascimento); 0000-0002-5439-5314 (A. P. C. Furtado); 0000-0003-3548-9670 (R. F. Mello)

called Non-dominated Sorting Genetic Algorithm (NSGA-II) [17] with two and three objective functions, respectively. Others use algorithms that combine objective functions into a [18].

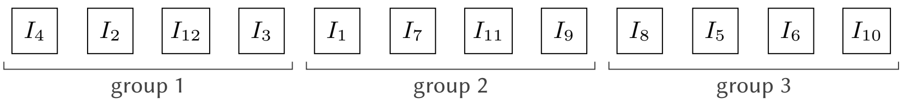
Due to the possibility related to the adoption of different objective functions and the fact that this algorithm was evaluated in real world scenarios, this work focuses on bringing the results of a study that evaluated the effectiveness of the proposed algorithm in [16] called Algorithm Group Formation Multiobjective Optimization (GFMOA). This aimed to test the extent to which the algorithm can replicate the groups created by an expert instructor, using a set of data collected in an on-site master's course in software engineering.

This work is organized as follows: Section 2 introduces the problem formulation. Section 3 presents the experimental methodology used to evaluate the proposed method. Section 4 presents the results achieved. Finally, Section 5 highlights conclusions and future work.

2. Problem Formulation

Based on a set of criteria, the automated group formation process seeks to obtain the best combination of students in each group and in the class as a whole. Given a finite set with k students, $\vec{I} = (I_1, I_2, ..., I_k)$, to be divided into g groups, where $g < k$ and k is divisible by g. Each student is represented by another finite set of n characteristics used to describe the student profile, $I. = (C_1, C_2, ...C_n)$. A solution that is said to be valid for the problem of forming groups is represented as a permutation of $\vec{I}$, where each set of k/g positions represents individuals belonging to a certain group. So we can say that we have two types of way of seeing the students as a whole: a) One of a physical way in which we see the students themselves; b) another in a logical way in which we use a start and an end for each logical subset. The following example, Figure 1, considers a possible candidate solution with $k = 12$ and $g = 3$ and shows the visual representation of each group in the class.

Figure 1: Problem formulation example



As shown in the example the solution has a permutation of 12 students, where I_1 represents $student_1$, I_2 represents $student_2$ and so on. The first 4 elements of the solution correspond to the students allocated in the first class; then, the next 4 elements are in the second group; and, finally, the others are part of the third group, thus representing the logical way to analyze the given set of students.

3. Method

3.1. Data and course context

In order to evaluate the proposed approach we used a dataset containing four offerings (two times in the first semester of 2019, one time in the second semester of 2019, and first semester of 2020) of a professional master-level course in software engineering offered in a face-to-face setting, at a Brazilian private university between 2019 and 2020. As part of the assessment for this program (60% of the credits to conclude the master), the students were divided into groups to develop a practical software engineering project using a problem-based learning methodology. Without completing this project, a student is not able to graduate. In those four offerings, a total of 93 students were divided into 14 groups. The different runs of the course had the numbers present on table 1.

Table 1
The different runs of the course.

Class	Number of Students	Number of groups
2019.1a	30	5
2019.1b	21	3
2019.2	14	2
2020.1	28	4

The data acquired from the students to generate their profiles, which are then employed by the algorithm, is presented in table 2. An online questionnaire was used to gather data throughout the first week of the course. Students were asked for information such as their location of residence, information about their graduating course, professional profile, time availability, and MBTI categorization.These materials were incorporated into the GFMOA. Because GFMOA exclusively employs numeric resources, we use the ordinal encoding method to turn categorical characteristics into ordinal numbers [19].

3.2. Manual Methodology for Group Formation

The teacher utilized three criteria to establish the final groupings based on the information gathered from the students:

1. Divide the students into groups based on the professional profile stated, balancing the various talents. This criterion ensures a diversity of talents for the project's progress.
2. Consider the experience time and graduation year when creating groups with various experiences to avoid an imbalance in professional experience levels.
3. Adoption of the MBTI profile and availability so that organizations may better understand each member's psychological profile and commitment to the project throughout.

Since 2007, the instructor's group creation process has been applied, and the groups have proven to be effective in terms of student performance (final grades). Thus, automating the replication of this technology would be a critical step toward its scalability.

Table 2
Student Profile Form fields.

Feature	Options
City and State from the student	Open text variable
Company that the student work (if applicable)	Open text variable
Under-graduated course	Open text variable
Graduation year	Open text variable
Professional experience	1 to 3 years (Low) 4 to 7 years (Medium) more than 8 years (High)
Time availability to participate in group activities	Only class hours 5 hours 10 hours More than 10 hours Exclusive dedication
Professional profile	Developer (DES) Systems Analyst (ANA) Test Analyst (TEST) Project Manager (PM)
Typological classification of Myers-Briggs, MBTI [20]	Introvert (I) Extravert (E)

3.3. Automatic Methodology for Group Formation

3.3.1. Optimization Process

The multi-objective method presented in [16] was used in this work for the automated group construction of face-to-face classes in a postgraduate degree in software engineering. The goal is to see if the GFMOA can construct groups with similar characteristics to those produced by the manual technique. The GFMOA seeks to maximize two goals: inter-homogeneity between groups and intra-heterogeneity within each group. This is accomplished by taking into consideration the same factors used by the teacher (see section 3.2). The sections that follow describe the key aspects of the algorithm that was employed.

3.3.2. Calculation of inter-group homogeneity ($InterHom$)

The measure of inter-group homogeneity is obtained through the steps described below: the matrix $M_{students}$ represents the course run, in which each line represents a student in the class, and the columns correspond to their features, normalized between 0 and 1. Let define $VT = (\overline{C_1}, \overline{C_2}, ..., \overline{C_n})$, where $\overline{C_i}$ is the average of the values for each feature i among all the students.

Similarly, given the matrix $M_{students,j}$ whose rows represent each students of the j-th group, and $\vec{VI_j} = (\overline{X_{j,1}}, \overline{X_{j,2}}, ..., \overline{X_{j,n}})$ the average of the values each i features in the j group. The value of inter-group homogeneity between the $InterHom$ groups is given by:

$$InterHom = \sum_{j=1}^{g}[(\overline{C_1} - \overline{X_{j,1}})^2 + (\overline{C_2} - \overline{X_{j,2}})^2 + ... + (\overline{C_n} - \overline{X_{j,n}})^2]. \quad (1)$$

$InterHom$ corresponds to the sum of the squares of the differences between VT and VI, so that the lower the value of $InterHom$, the greater the similarity of each group in relation to the others.

3.3.3. Calculation of intra-group heterogeneity ($IntraHet$)

Let MD be a square matrix of dimensions *student* $\times$ *student*, representing the distances (Euclidean) of each student in relation to the others. That is, each position $D_{a,b}$ of MD is given by:

$$D_{a,b} = \sqrt[2]{(C_{a,1} - C_{b,1})^2 + (C_{a,2} - C_{b,2})^2 + ... + (C_{a,n} - C_{b,n})^2}. \tag{2}$$

Where $\overline{D_{a,b}}$ is the average of the distances of each student in a group regarding colleagues on the same team. The value of the total intra-group heterogeneity is given by the average of the distances of each group, that is:

$$IntraHet = \frac{\sum_{j=1}^{g} D_{a,b}}{g}. \tag{3}$$

3.3.4. Algorithm Configuration

In this study, we used the same algorithm (NSGA-II) and parameters validated in [16] and described in table 3, where it presents the details of each parameter adopted. The experiments were performed in the following set up:

1. Operational system Arch Linux
2. CPU Intel Core i5-7300HQ
3. 8GB de RAM
4. GPU GeForce GTX 1050 de 4GB.

Table 3
NSGA II's parameters.

Hyperparamemter	Value
Initialization method	Uniform random
Selection	Roulette
Crossover	Partially matched
Mutation	Swap
Mutation rate	20%
Chromosome Mutation rate	15%
Crossover rate	50%

4. Results

The findings were categorized into three sections: (i) Group formation in relation to the professional profile; (ii) Group formation in relation to professional experience; (iii) Group formation in relation to the MBTI.

Table 4
Manual and automatic group formation in relation to the professional profile divided into Developer(DEV), Systems Analyst (ANA), Test Analyst (TES) and Project Manager (PM).

Class	Automatic Formation	Manual Formation	Simil.
2019.1a	ANA, DEV, DEV, DEV, PM, TES	ANA, ANA, ANA, DEV	50.0%
	ANA, DEV, PM, PM, TES, TES	DEV, DEV, PM, PM, TES, TES	83.3%
	ANA, DEV, DEV, PM, TES, TES	DEV, DEV, DEV, PM, PM, TES	66.6%
	DEV, DEV, DEV, DEV, DEV, TES	DEV, DEV, DEV, DEV, TES, TES, TES	71.4%
	DEV, DEV, PM, PM, PM, TES	DEV, DEV, DEV, PM, PM, PM, TES	85.7%
2019.1b	ANA, ANA, DEV, DEV, DEV, DEV, PM	ANA, DEV, DEV, DEV, DEV, DEV, PM	85.7%
	DEV, DEV, DEV, DEV, DEV, DEV, TES	ANA, DEV, DEV, DEV, DEV, DEV, DEV	85.7%
	ANA, DEV, DEV, DEV, DEV, PM, TES	ANA, DEV, DEV, DEV, PM, TES, TES	85.7%
2019.2	DEV, DEV, DEV, DEV, DEV, PM, TES	DEV, DEV, DEV, DEV, PM, TES, TES	85.7%
	DEV, DEV, DEV, DEV, PM, PM, TES	DEV, DEV, DEV, DEV, DEV, PM, PM	85.7%
2020.1	ANA, ANA, DEV, DEV, DEV, DEV, PM	ANA, ANA, ANA, DEV, DEV, DEV, PM	85.7%
	ANA, ANA, ANA, ANA, ANA, DEV, PM	ANA, ANA, ANA, DEV, DEV, DEV, PM	71.4%
	ANA, ANA, DEV, DEV, DEV, DEV, PM	ANA, ANA, DEV, DEV, DEV, PM, PM	85, 7%
	ANA, DEV, DEV, DEV, PM, PM, PM	ANA, ANA, DEV, DEV, DEV, PM, PM	85.7%

In this study we considered characteristics such as professional, psychological and experience profile. In addition, we also evaluate the GFMOA in terms of processing time. The experiment carried out in this article considered classes between 14 and 30 students, with the GFMOA taking 120 seconds to automatically form groups for the class with the largest number of students, which enhances the use of the algorithm in practice.

4.1. Group formation in relation to the professional profile

Table 4 displays the groupings produced by the automatic and manual procedures when the professional profile is taken into account. Each student in the group is represented by their profile, which includes Developer (DEV), Systems Analyst (ANA), Test Analyst (TES), and Project Manager (PM) (PM).

As shown in Table 4, the automatic formation was able to generate heterogeneous groups internally and homogeneous among themselves, in all classes. Regarding the similarity with the groups formed by the manual approach, those underlined stretches represent the identical elements between the groups considering their components' profile. In 2019.1a, the instructor allowed groups of different sizes. As the GFMOA formed groups of the same size (5 groups of 6 students), the variability allowed in manual formation penalized similarity in two cases (50.0% and 66.6% similarity). However, the other groups showed similarities above 70%, reaching up to 85.7%. In the case of 2019.1b and 2019.2, groups of the same size were formed. Consequently, GFMOA's performance was better, achieving 85.7% similarity across all groups formed. It means that the algorithm made the wrong choice in only one of the seven possible profiles per group. In 2020.1, of the four groups formed, three showed 85.7% similarity with manual formation. Only one group presented two divergences comparing the automatic and manual outcomes, obtaining 71.4% similarity with manual formation. On average, the similarity in terms of the professional profile is 79.58%, with a standard deviation of 10.84%.

Table 5
Manual and automatic group formation in relation to professional experience divided into High (H), Medium (M) and Low (L).

Class	Automatic Formation	Manual Formation	Similarity
2019.1a	H, H, H, L, M	H, H, H, M	100%
	H, H, L, L, M, M	H, H, L, M, M, M	83.3%
	H, H, H, L, M, M	H, H, L, M, M, M	83,3%
	H, H, L, M, M, M	H, H, L, M, M, M, L	85,7%
	H, H, L, L, M, M	H, H, H, L, L, L, M	71.4%
2019.1b	H, H, L, L, L, M, M	H, H, H, L, L, M, M	85.7%
	H, H, H, H, L, M, M	H, H, H, L, M, M, M	85.7%
	H, H, H, L, L, M, M	H, H, H, L, L, L, M	85.7%
2019.2	H, L, L, L, M, M, M	H, H, L, L, L, M, M	85.7%
	H, H, H, H, L, M, M	H, H, H, L, M, M, M	85.7%
2020.1	H, L, L, L, M, M, M	H, H, L, L, L, M, M	85.7%
	H, H, H, L, M, M, M	H, H, L, L, L, M, M	71.4%
	H, L, L, M, M, M, M	H, H, L, M, M, M, M	85.7%
	H, H, L, L, M, M, M	H, L, M, M, M, M, M	71.4%

4.2. Group formation in relation to professional experience

In addition to the analysis of the professional profile, Table 5 presents the results for the professional experiences divided into High (H), Medium(M) and Low(L); again the underlined stretches represent the identical elements between groups. In this case, the average similarity reached 83.31% (with a standard deviation of 7.59%), which is higher than the professional profile result. The main difference was in the 2019.1a class. In general, the algorithm made one wrong choice per group. The lower accurate group achieved 71.4% of similarity in the class of 2020.1. On the other hand, the GFMOA reached 85.7% of similarity for all groups in 2019.1b and 2019.2.

4.3. Group formation in relation to the MBTI

Finally, we also investigated the automatic and manual group matching in terms of the MBTI, which divides the students into Introvert (I) and Extravert (E). Table 6 shows the results for this category. In this case, the average similarity was the higher one, reaching 87.57%, with

Table 6
Manual and automatic group formation in relation to the MBTI divided into Introversion (I) and Extroversion (E).

Class	Automatic Formation	Manual Formation	Similarity
2019.1a	E, I, I, I, I, I	E, I, I, I	100%
	E, E, I, I, I, I	E, I, I, I, I, I	83.3%
	E, E, I, I, I, I	E, E, I, I, I, I	100%
	E, E, I, I, I, I	E, I, I, I, I, I, I	85.7%
	E, E, I, I, I, I	E, E, E, E, I, I, I	71.4%
2019.1b	E, E, I, I, I, I, I	E, E, E, I, I, I, I	85.7%
	E, E, E, I, I, I, I	E, I, I, I, I, I, I	71.4%
	E, E, I, I, I, I, I	E, E, E, I, I, I, I	85.7%
2019.2	E, E, E, I, I, I, I	E, E, E, I, I, I, I	100%
	E, E, E, E, I, I, I	E, E, E, E, I, I, I	100%
2020.1	E, E, E, E, I, I, I	E, E, I, I, I, I, I	71.4%
	E, E, I, I, I, I, I	E, E, E, I, I, I, I	85.7%
	E, E, E, I, I, I, I	E, E, E, E, I, I, I	85.7%
	E, E, E, E, I, I, I	E, E, E, E, I, I, I	100%

a standard deviation of 11.06%. This result was expected as MBTI divides the students into only two groups, while the professional profile and experience divides the students into four and three groups, respectively. The groups created in the 2019.2 semesters reached a perfect match between the automatic and manual approaches. On the other hand, the 2019.1b semester reached its worse results. The course offerings in 2019.1a and 2020.1 obtained similar results creating groups with 100% and 71.4% of similarity in the best and worse case, respectively.

5. Conclusion

This work proposes the study and evaluation of the group formation algorithm proposed by [16] in a master's course in software engineering, in which the classes are formed by an experienced instructor, considering technical, professional and psychological aspects . The method was evaluated using the same information made available to the instructor, taking into account the criteria of intergroup homogeneity and intragroup heterogeneity. The results showed that the automatically formed groups had an average similarity of $83.46\%\pm9.83$ with the manually formed groups. Thus, it is possible to state that the proposed methodology can be used to reproduce a well-established methodology for manual formation of groups from a pedagogical point of view.

As a future work, it was intended to expand the process of characterization and description of student profiles in greater detail. In addition, in terms of assessment, we intend to use different approaches, including a) a questionnaire that will be applied after the group activity to measure student satisfaction and b) conducting a random control trail [21] to measure the effects of different groups training approaches on individual and group performance. Finally, we intend to develop a recommendation system to support the instructors' decision on the best groups for specific activities.

References

[1] B. Zheng, M. Niiya, M. Warschauer, Wikis and collaborative learning in higher education, Technology, Pedagogy and Education 24 (2015) 357–374.

[2] G. Chen, V. Rolim, R. F. Mello, D. Gašević, Let's shine together! a comparative study between learning analytics and educational data mining, in: Proceedings of the Tenth International Conference on Learning Analytics & Knowledge, 2020, pp. 544–553.

[3] C. Odo, J. Masthoff, N. Beacham, Group formation for collaborative learning, in: International Conference on Artificial Intelligence in Education, Springer, 2019, pp. 206–212.

[4] P. Dillenbourg, Collaborative learning: Cognitive and computational approaches. advances in learning and instruction series., ERIC, 1999.

[5] S. Järvelä, P. A. Kirschner, A. Hadwin, H. Järvenoja, J. Malmberg, M. Miller, J. Laru, Socially shared regulation of learning in cscl: Understanding and prompting individual- and group-level shared regulatory activities, International Journal of Computer-Supported Collaborative Learning 11 (2016) 263–280.

[6] A. Müller, H. Bellhäuser, J. Konert, R. Röpke, Effects of group formation on student satisfaction and performance: A field experiment, Small Group Research 0 (2021) 1046496420988592. doi:10.1177/1046496420988592.

[7] C.-C. Liu, C.-C. Tsai, An analysis of peer interaction patterns as discoursed by on-line small group problem-solving activity, Computers & Education 50 (2008) 627–639.

[8] I. Monteverde, D. Ramos, B. Gadelha, E. Oliveira, P. do Nascimento, Formação de grupos em ambientes virtuais de aprendizagem: Uma revisão sistemática da literatura, in: Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE), volume 27, 2016, p. 310.

[9] N. Maqtary, A. Mohsen, K. Bechkoum, Group formation techniques in computer-supported collaborative learning: A systematic literature review, Technology, Knowledge and Learning 24 (2019) 169–190.

[10] A. Krouska, C. Troussas, M. Virvou, Applying genetic algorithms for student grouping in collaborative learning: a synthetic literature review, Intelligent Decision Technologies 13 (2019) 395–406.

[11] C. A. da Silva, D. C. Rosa, P. B. Miranda, F. R. Cordeiro, T. Si, A. C. Nascimento, R. F. Mello, P. S. de Mattos Neto, A multi-objective grammatical evolution framework to generate convolutional neural network architectures, in: 2021 IEEE Congress on Evolutionary Computation (CEC), IEEE, 2021, pp. 2187–2194.

[12] J. Moreno, D. A. Ovalle, R. M. Vicari, A genetic algorithm approach for group formation in

collaborative learning considering multiple student characteristics, Computers & Education 58 (2012) 560–569.

[13] M.-I. Dascalu, C.-N. Bodea, M. Lytras, P. O. De Pablos, A. Burlacu, Improving e-learning communities through optimal composition of multidisciplinary learning groups, Computers in Human Behavior 30 (2014) 362–371.

[14] Y. Zheng, C. Li, S. Liu, W. Lu, An improved genetic approach for composing optimal collaborative learning groups, Knowledge-Based Systems 139 (2018) 214–225.

[15] S. Garshasbi, Y. Mohammadi, S. Graf, S. Garshasbi, J. Shen, Optimal learning group formation: A multi-objective heuristic search strategy for enhancing inter-group homogeneity and intra-group heterogeneity, Expert Systems with Applications 118 (2019) 506–521.

[16] P. B. Miranda, R. F. Mello, A. C. Nascimento, A multi-objective optimization approach for the group formation problem, Expert Systems with Applications 162 (2020) 113828.

[17] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: Nsga-ii, IEEE transactions on evolutionary computation 6 (2002) 182–197.

[18] C.-L. Hwang, A. S. M. Masud, Multiple objective decision making—methods and applications: a state-of-the-art survey, volume 164, Springer Science & Business Media, 2012.

[19] K. Potdar, T. S. Pardawala, C. D. Pai, A comparative study of categorical variable encoding techniques for neural network classifiers, International journal of computer applications 175 (2017) 7–9.

[20] A. Furnham, Myers-Briggs Type Indicator (MBTI), Springer International Publishing, Cham, 2017, pp. 1–4.

[21] H. Bellhäuser, J. Konert, A. Müller, R. Röpke, Who is the perfect match? effects of algorithmic learning group formation using personality traits, Journal of Interactive Media (i-com) 17 (2018) 65–77. doi:https://doi.org/10.1515/icom-2018-0004.