Algorithms and Machine Learning Techniques in Collaborative Group Formation

Rosanna Costaguta(✉)

Facultad de Ciencias Exactas y Tecnologías (FCEyT),
Instituto de Investigación en Informática y Sistemas de Información (IIISI),
Universidad Nacional de Santiago del Estero (UNSE), Avda. Belgrano (S) 1912,
Santiago del Estero 4200, Argentina
rosanna@unse.edu.ar

Abstract. Computer Supported Collaborative Learning (CSCL) is an emerging interdisciplinary research area that deals with the formation of students groups to work and to learn together in an educational context. One of the factors that affect successful collaborative learning is the group composition. This paper surveys the most relevant researches carried out in this field to date. For each one it describes the applied criterion to form learning groups and the way in which the grouping criterion is applied. These researches are compared and some conclusions are outlined.

1 Introduction

Collaborative learning (CL) can be seen as teaching methods in which students work in small groups to help them learn from each other [1]. The technological advance occurred in recent decades allowed the CL adopt computational tools that facilitate collaboration, coordination and communication transforming it in Computer Supported Collaborative Learning (CSCL). Today its use is diffused in the area of education, and there are numerous studies showing that can be advantageously applied.

A learning group is defined as a structure formed by people who interact to achieve specific learning objectives through their participation [2]. In CSCL there are different approaches to form groups, it is possible randomly select the members, let them self-select themselves, or choose them by certain criterion established by the teacher, and also do it manually (by the teacher) or automatic (by the system). Each of these alternatives has certain disadvantages. Randomization can generate very unbalanced groups that are unlikely to be effective; the self-selection can cause discrimination among students with poor social relationship; and manually creating unviable when the number of students is high or when the selection criterions are complex.

The objective of this work is to present some background on proposed approaches to the formation of groups in CLSL, analyzing in them the algorithms and machine learning techniques that apply. This paper is organized as follows: Sect. 2 describes the experiences surveyed in the formation of groups, Sect. 3 performs an analysis and comparison of the same, and finally, Sect. 4 presents some conclusion.

O. Pichardo Lagunas et al. (Eds.): MICAI 2015, Part II, LNAI 9414, pp. 249–258, 2015.
DOI: 10.1007/978-3-319-27101-9_18

2 Applied Approches in Collaborative Group Formation

The Supnithi *et al.* [3] suggest the formation of opportunistic group using two ontologies, one of negotiation and other collaborative learning. For authors opportunistic group is one that is dynamically generated when a situation where it is desirable that the student migrate individual learning to collaborative, there formed a group with members who share a goal of learning is detected. In this proposal the students start working individually with a software agent that monitors the actions of the student and updates his student profile. This agent is able to recognize when your student would benefit by changing to the mode of collaborative work. In such cases, the agent initiates a process of negotiation with the agents of other students of the course to form a group. The agent begins by setting a goal of learning for the group and a role for the student group. The agents of all other students negotiate with him using the ontology created for it (each agent considers the information contained in the profile of the student and estimate the benefits that could get, if it participate of the group that it was called). If negotiation is successful every student is informed about the target of the group learning and the role which must assume, and a communication channel is opened for the use of members. When one of the students affirms to have reached the goal, the agents close the communication channel and update the student profile. The authors claim to have developed the ontologies but still no experimental data.

Balmaceda *et al.* [4] suggest the use of an assistant agent to form collaborative groups based on three characteristics that may affect the group performance: the psychological styles, team roles and social relationships. The psychological styles considered are those proposed by Myers-Briggs (extroversion/introversion, sensing/intuitive, thinker/sentimental, judgment/perception). Respect to roles, they take the life cycle phases of collaborative work and the eight team roles proposed by Mumma, respecting the paired appearance of these by phase. Zheong [5] consider that the formation of groups is a constraint satisfaction problem; they take every place in the group as a variable, the list of course students as the domain of these variables, and derived from Mumma roles and the styles from Myers-Briggs the restrictions that them have to satisfy. Some examples of restrictions proposed by the authors are: the student may participate in only one group in each group all roles must appear, psychological styles must be balanced. The preferences of each student regarding the manifestation of roles and psychological styles are stored in your student profile. To validate the proposal's authors developed a pilot without assessing the performance of the groups created experience.

Zheong [5] proposes a generator groups by applying data mining techniques (clustering), analyzes the interactions of students to extract interaction patterns. These patterns and the rating assigned by the teacher are used to differentiate between effective and weak groups. Based on this distinction, and using decision trees, establish composition rules of groups used by the generator groups. Students are represented by a set of personal characteristics, and the first cluster is carried out considering these characteristics. Then, by having the teacher's grade and the level of interaction revealed new patterns that allow you to refine the formation of groups and produce new groupings are extracted. The authors plan to implement and validate the proposal.

Henry [6] describes a program that automatically performs the grouping of students. Initially students are surveyed to capture information necessary for the creation of his student profiles (preferred programming language, number of postponements, subject notes, whit which partner it prefer to work and which not, etc.). The teacher must indicate the desired size for the groups, and whether homogeneous or heterogeneous want based on a particular feature. The authors state that they use the described software for years but they do not have experimental results.

Hoppe [7] establishes three criterions to perform the grouping: a complementary criterion where a student with high competence in a topic is grouped with other of low competition, a competitive approach where students are grouped with similar profiles, and finally, a selection criterion of problem where no member of the group can solve the problem alone but a group integrating with other members who possess the knowledge as required. There is little documentation of experimentation by the author.

Wessner and Pfister [8] present a group formation process composed of three stages: initiation, pairs identification and negotiation. The initiation of a collaborative situation can arise for student choice or decision of the teacher. When the student proposes the initiative, the system searches the profiles of students other pupils who meet the requirements to work with him, providing a list of these or telling you that there are no matches. The student can choose his partner from the list, or cancel the training. In the case where a candidate selected, the system will consult whether to accept it. If the group agree is created, but rather the student may choose another candidate. The group created a communication channel opens. Wessner and Pfister [8] also propose the concept of point of cooperation (PoC), understood as an opportunity to collaborate included in the system. Also, classify PoC in: generic PoC (GPoC), which are all the facilities of cooperation provided by the system that may or may not be used by students (mail, chat, etc.); spontaneous PoC (SPoC), incorporated into the course but not linked to a particular position (seek help from the teacher, find a partner for dialogue, etc.) activities, and intentional PoC (IPOC), logical collaborative activities and didactically integrated course in a given thereof (forum enabled by the teacher after the development of certain units, chat to discuss a given concept, etc.) point. The authors implemented their concept of PoC in the L3 environment. There is a working mode in collaboration that allows synchronous and asynchronous POCI with formation of groups by the teacher manually or automatically by the system. For automatic formation the system considers the course as a sequential list of units and calculates the learning distance among students by calculating the difference between the units in which students find themselves.

Duque Medina *et al.* [9] suggest identifying indicators of collaboration and then apply some of these criterions to form groups: concentration, grouping students who have similar values in certain indicators, and dispersion, grouping students differ in the values of certain indicators. Obviously, while a criterion generates homogeneous groups the other produces heterogeneous groups. However, the authors experimented with the combination of both criterions in the COLLECE system. For this they calculated indicators of the student: work (it measures the dedication) and discussion (it measures the level of participation), indicators of the group: coordination (it measures the extent to which students agree to share the charge and the workspace), and speed (it measures the time taken on the task), and indicators of solution: correctness and validity. Then they generated the groups considering the concentration criterion for indicators for student and dispersion criterion for indicators of the solution, achieving homogeneous groups in some aspects and heterogeneous in others.

Cocea and Magoulis [10] proposed case-based reasoning combined with data mining techniques (clustering) to form the groups. The authors experimented with the proposal by using the learning environment eXpresser in the field of mathematics generalization. With case-based reasoning can recognize solving strategies applied and the clustering can detect students who applied similar strategies. With this information the teacher defines the constitution of the groups.

Sukstrienwong [11] proposes the use of a genetic algorithm for forming heterogeneous groups. The author considers each student attributes defined by the current rating and score on the previous year, by way of representing the academic and educational skills of the student, respectively. Averaging the values of the members for the two attributes it is calculated by a chromosome also two attributes per group. The authors describe an experiment and they analyze the results.

Lin *et al.* [12] consider the formation of collaborative groups as a problem of multi-grouping and to fix propose using the particle swarm optimization. For this purpose define two grouping criterions: the level of understanding of each student on a given topic, and the level of student interest on that topic. With these criterions, the authors calculate the difference in the level of understanding between groups and the maximum distance in the level of interest among groups. The first time the indicators are calculated on historical data and are updated for later groupings. Experimental data show the viability of the proposal.

Yannibelli and Amandi [13] propose setting up well-balanced groups to the nine roles of equipment known as Belbin roles. An unbalanced group is one where these roles do not appear naturally or where the same role is manifested by different members. The authors suggest the existence of three indicators and the implementation of a genetic algorithm. The first indicator shows whether each of the roles appear naturally in each group. The second indicator is applied to each group and calculate the balance level roles based on the first indicator. The third indicator maximizes the average balance levels in all groups, and is taken as evaluation function indicator for the genetic algorithm. The constitutive genes of chromosome correspond to students of the course. The initial population is given by random permutations of the value of the genes, that is, with the same students positioned at various locations within different chromosomes.

Martin and Paredes [14] propose using learning styles by Felder and Silverman as a main feature for grouping students. The authors propose the creation of groups which combine different styles in the same proportion, making a temporary shape with a criterion of homogeneous styles and then regrouping with a criterion of heterogeneous styles. There are no experimental results.

Carro *et al.* [15] propose the use of grouping rules looking for students with similar characteristics to integrate the same group. Initially to form the groups is considered one of the following characteristics: learning style, knowledge level and frequency of interaction. Then there is the possibility of sub-groupings within groups formed, considering some of the other characteristics or other criterions (for example, considering the collaboration wishes of students, with whom they want to work or with who do not want to).

Liu *et al.* [16] propose perform an intelligent grouping of students based on their learning styles. For this, first they get the learning style of each student using the styles of Felder and Silverman. The teacher provides the necessary grouping parameters (number of groups to form and number of members per group). Students are arranged in descending order according to the score obtained for the style, and then the ranked list is segmented into as many equal parts as students should have each group, and finally, the groups are generated by assigning randomly from each segment to one student group. There are no experimental results.

Barati Jozán *et al.* [17] propose the use of a genetic algorithm and two evaluation functions: one intragroup and intergroup other. While the first measures the quality of each group, the second compares the competition between groups created. The authors seek heterogeneity intergroup and intragroup homogeneity. The genetic algorithm considers each group as a chromosome and students as genes. The length of chromosome indicates the number of group members. For define the initial population the students are randomly distributed within groups to form. The experiences analyzed are on simulated data.

Razmerita and Brun [18] propose perform homogeneous groupings using data mining techniques (clustering) on the data of students who are judged adequate. The authors suggest evaluating the performance (individual and group) of these groups to make changes to the groupings made. There are no experimental results.

Ounnas *et al.* [19] propose to create well-balanced groups to the nine roles of Belbin, which implies a certain presence of roles in the group. The authors using ontologies for modeling the characteristics of the students, which includes personal, social and academic data (learning style, favorite subjects, preferred partners, leading role, supporting role, etc.). The negotiation is presented as a constraint satisfaction problem, for example, looking for heterogeneity in learning style or homogeneity in the favorite subjects. The teacher shows how many students want to group. The authors conducted experiences with real students and also with simulated data.

Wang *et al.* [20] introduce a heterogeneous grouping system called DIANA. This system uses a genetic algorithm to form groups heterogeneous with same size and same level of diversity. For this genetic algorithm one chromosome represents one group and each gene within a chromosome represents one student. The tool uses the students thinking styles collected from questionnaires to create groups with 3 to 7 members. DIANA was tested with real students.

3 Analysis and Comparison of the Applied Approaches

To make the comparison between the different approaches of clusters of students in CSCL environments, the following questions were raised: 1 - What students features involved in the grouping process are considered?, 2 - What techniques or algorithms are applied especially for grouping?, 3 - Form groups is the decision of one person (teacher or student) or the same system?, 4 – In the formed groups, the characteristics of the students assume similar values (homogeneous groups), different (heterogeneous groups), or there are some with similar characteristics and other with differing values (mixed groups)?, 5 – The used algorithm raises the possibility of regrouping looking for improvement ?, 6 - Are there experimental results?.

Considering the first question, we can say that all the approaches analyzed use own characteristics of the students involved. In some cases there are similarities in the aspects evaluated, in [3, 4, 13, 19] are considered the team roles, in [4, 14–16] learning styles, in [3, 5, 9, 11, 17] teacher qualifications, in [12, 18] the topic or level of interest, in [7, 8, 12, 15, 17, 18] the level of knowledge or understanding, in [6, 7, 18, 19] private personal data, and in [4, 5, 9, 15, 17] social relations, the level of interaction or communication. In other cases there are no similarities, for example, only in [3] takes the target learning feature, in [10] the style of problem solving, and in [20] the thinking style. It was also noted that in [3–6] the authors identify in their proposals a student model as the place where all these characteristics are stored.

Considering the second question, we can say that the techniques and algorithms applied are varied, although in some cases there are similarities. In [3, 19] ontologies are proposed, in [3, 4] software agents are used, in [5, 10, 18] clustering is used, in [7–9, 12, 15] rules or specific grouping criterions are defined, in [4, 12, 19] grouping is solved as a constraint satisfaction problem, in [14, 18] segmentation system is applied, and in [11, 13, 17, 20] genetic algorithms are proposed. The decision trees are used only in [5].

Considering the third question, we can say that most of the analyzed approaches perform the formation of groups at the request of teacher organizes the course of CSCL [4–20]. Only in [3, 8] there are possibilities of automatic grouping initiative (by systems), and in [8] the possibility of forming groups on the initiative of the students themselves are also offered.

Considering the fourth question, we can say that the approaches of heterogeneous groups and approaches of homogeneous groups exist in the same amount. The proposed approaches in [3, 5, 7, 11, 16, 19, 20] only generate heterogeneous groups, whereas in [8, 10, 13, 17, 18] only homogeneous, and in [6, 9, 12, 15] it is possible to choose between the two categories. Furthermore, in particular, in [4, 9, 12, 14, 15] exists the possibility of forming mixed groups, using simultaneously criterions to homogeneity and heterogeneity.

Considering the fifth question, we can say that only in [5, 12] there are possibilities of iterative regrouping looking for more efficient formation of groups.

Finally, considering the sixth question, we can say that considerable number of analyzed approaches does not support their proposals with experimental results; this occurs in [3, 5, 6, 8, 14–16, 18].

In Table 1 the questions and answers are synthetized.

Table 1. Comparison of grouping approaches

Ref.	Questions					
	1	*2*	*3*	*4*	*5*	*6*
[3]	Team roles, Qualifications, Learning goals	Ontologies, Agents	System	Heterogeneous	No	No

(*Continued*)

Table 1. (*Continued*)

Ref.	Questions					
	1	*2*	*3*	*4*	*5*	*6*
[4]	Psychologies styles, Team roles, Social, relationships	Agents, Constraints satisfaction	Professor	Mixed	No	Yes
[5]	Qualifications, Interaction level	Clustering, Decision trees	Professor	Heterogeneous	Yes	No
[6]	Collaboration preferences, Personal information	Grouping criterions	Professor	Homogeneous Heterogeneous	No	No
[7]	Knowledge, Capacities	Grouping criterions	Professor	Heterogeneous	No	Yes
[8]	Knowledge	Grouping criterions	Professor, Students, System	Homogeneous	No	No
[9]	Communication level, Quality	Grouping criterions	Professor	Homogeneous Heterogeneous Mixed	No	Yes
[10]	Style of problem resolution	Clustering, Case-based reasoning	Professor	Homogeneous	No	Yes
[11]	Qualifications	Genetic algorithms	Professor	Heterogeneous	No	Yes
[12]	Interest level Learning level	Grouping criterions	Professor	Homogeneous Heterogeneous Mixed	Yes	Yes
[13]	Team roles	Genetic algorithms	Professor	Homogeneous	No	Yes
[14]	Learning styles	Rank and segmentation	Professor	Mixed	No	No
[15]	Learning styles, Knowledge level, Interaction style, Opinions, Collaboration preferences	Grouping criterions	Professor	Homogeneous Heterogeneous Mixed	No	No
[16]	Learning styles	Rank and segmentation	Professor	Heterogeneous	No	No
[17]	Social characteristics, Qualifications	Genetic algorithms	Professor	Homogeneous	No	Yes

(*Continued*)

Table 1. (*Continued*)

Ref.	Questions					
	1	*2*	*3*	*4*	*5*	*6*
[18]	Interest topics, Knowledge level, Country	Clustering	Professor	Homogeneous	No	No
[19]	Team roles Sex	Ontologies, Constraint satisfaction	Professor	Heterogeneous	No	Yes
[20]	Thinking styles	Genetic algorithms	Professor	Heterogeneous	No	Yes

4 Conclusions

In many areas of science and industry success depends on individual skills to be a productive member of a group that people can demonstrate, this is also true for ACSC. So far, the formation of collaborative groups are made based on personal information of the students of the course which is available on systems, and is usually contained in the profiles or student models (data such as sex, age, level of knowledge, main interests, preferences, learning styles, grades obtained, level of participation, etc.). Thus, these data are evaluated to select the members of the group so that everyone benefits potentially working together. In making this selection sometimes complementarity is encouraged (when there is heterogeneity in the group), in other cases competitiveness (when there is homogeneity among members), and others are looking for both (when there is homogeneity between some characteristics of the members and heterogeneity in others).

Several approaches of group formation have been proposed and machine learning techniques that are varied include: genetic algorithms, agents, clustering, optimization of restrictions, individual grouping criterion, etc. The factors that guide the grouping are also varied: psychological styles, learning styles, social relations, level of knowledge, level of participation, etc. Many of the proposed approaches to the formation of groups have been tested, but most points to validate the effectiveness of clustering algorithm rather than evaluating the effects on the performance of the group formed with this algorithm.

Predominantly the formation of groups in CSCL is performed at the request of the teacher who also indicates the parameters under which the grouping algorithm will perform its task. In general, these parameters are the number of groups to be formed, the number of members that each group should have. In some cases the teacher should indicate the type of group to form (homogeneous or heterogeneous) and the student characteristics for the selection of members. The examples in which these tasks are performed automatically or delegated to software agents are few.

The current perspectives for ubiquitous computing and its relationship to intelligent systems augur the emergence of new approaches to the formation of groups of CSCL including components related to the context of the student. Examples of contextual variables in CSCL that could be considered in future grouping algorithms are: emotional parameters, noise, climate, temperature, availability of devices, proximity of others, etc.

Acknowledgements. This study was partially supported by research projects PICTO UNSE 2012-0016 and SECYT UNSE 23-C089.

References

1. Slavin, R.: Cooperative Learning: Theory, Research and Practice. Pearson, London (1995)
2. Souto, M.: Didáctica de lo grupal. Ministerio de Educación y Justicia. In: INPAD (1990)
3. Supnithi, T., Inaba, R., Ikeda, M., Toyoda, J., Mizoguchi, R.: Learning goal ontology supported by learning theories for opportunistic group formation. In: Proceedings of 9th World Conference on Artificial Intelligence in Education, France (1999)
4. Balmaceda, J., Schiaffino, S., Diaz-Pace, J.: Using constraint satisfaction to aid group formation in CSCL. Revista Inteligencia Artif. **17**(53), 35–45 (2014)
5. Zheong, Z.: A dynamic group composition method to refine collaborative learning group formation. In: Proceedings of 6th International Conference on Educational Data Mining, pp. 360–361 (2013)
6. Henry, T.: Creating effective student groups: an introduction to groupformation.org. In: Proceeding of 44th. ACM Technical Symposium on Computer Science Education, pp. 645–650, USA (2013)
7. Hoppe, H.: The use of multiple student modeling to parametrize group learning. In: Proceedings of 7th World Conference on Artificial Intelligence in Education, USA (1995)
8. Wessner, M., Pfister, H.: Group formation in computer-supported collaborative learning. In: Proceedings of International ACM SIGGROUP Conference on Supporting Group Work, pp. 24–31, USA (2001)
9. Duque Medina, R., Gómez-Peréz, D., Nieto-Reyes, A., Bravo Santos, C.: A method to form learners groups in computer-supported collaborative learning systems. In: Proceedings of First International Conference on Technological Ecosystem for Enhancing Multiculturality, pp. 261–266, Spain (2013)
10. Cocea, M., Magoulis, G.: User behaviour-driven group formation throuh case-based reasoning and clustering. Expert Syst. Appl. **39**, 8756–8768 (2012)
11. Sukstrienwong, A.: A genetic algorithm approach for forming heterogeneous groups of students. Int. J. Appl. Eng. Res. (IJAER) **9**(3), 297–311 (2014)
12. Lin, Y., Huang, Y., Cheng, S.: An automatic group composition system for composing collaborative learning groups using enhanced particle swarm optimization. Comput. Educ. **55**, 1483–1493 (2010)
13. Yannibelli, V., Amandi, A.: A deterministic crowding evolutionary algorithm to form learning teams in a collaborative learning context. Expert Syst. Appl. **39**, 8584–8592 (2012)
14. Martin, E., Paredes, P.: Using learning styles for dynamic group formation in adaptive collaborative hypermedia systems. In: Proceedings of First International Workshop on Adaptive Hypermedia and Collaborative Web-based Systems (AHCW), pp. 188–198 (2004)
15. Carro, R.M., Ortigosa, A., Martín, E., Schlichter, J.: Dynamic generation of adaptive web-based collaborative courses. In: Favela, J., Decouchant, D. (eds.) CRIWG 2003. LNCS, vol. 2806, pp. 191–198. Springer, Heidelberg (2003)

16. Liu, S., Joy, M., Griffiths, N.: iGLS: intelligent grouping for online collaborative learning. In: Proceedings 9th IEEE International Conference on Advanced Learning Technologies (ICALT), Latvia (2009)
17. Barati Jozán, M., Taghiyareh, F., Faili, H.: An inversion-based genetic algorithm for grouping of students. In: Proceedings of 7th International Conference on Virtual Learning, pp. 152–161, Rumania (2012)
18. Razmerita, I., Brun, A.: Collaborative learning in heterogeneous classes. towards a group formation methodology. In: Proceedings of 3rd International Conference on Computer Supported Education, The Netherlands (2011)
19. Ounnas, A., Davis, H., Millard, D.: A framework for semantic group formation in education. Educ. Technol. Society **12**(4), 43–55 (2009)
20. Wang, D., Lin, S., Sun, C.: DIANA: a computer-supported heterogeneous grouping system for teachers to conduct successful small learning groups. Comput. Hum. Behav. **23**, 1997–2010 (2007)