Adaptive group formation to promote desired behaviours

Amir Mujkanovic[a], David Lowe[b], and Keith Willey[a]
Centre for Real-time Information Networks – Faculty of Engineering and IT – University of Technology Sydney[a],
School of Information Technologies – Faculty of Engineering and IT – The University of Sydney[b]
Amir.Mujkanovic@uts.edu.au, David.Lowe@sydney.edu.au, Keith.Willey@uts.edu.au

BACKGROUND

There is substantial literature that shows the benefits of collaborative work, though these benefits vary enormously with circumstances. Irrespective of their structure and composition, groups usually exist for a particular reason and implicitly or explicitly target one or more outcomes. The achievements of group outcomes depend on many factors, including the individual behaviour of each group member. These behaviours are, in turn, affected by the individual characteristics, the context and the group composition. Constructing groups in a way that maximises the achievement of a specific outcome is complex with the optimal group composition depending on the attributes of the group members. Previous work has in most cases considered group formation based on one particular attribute, such as learning style, gender, personality, etc. Less common are instances of group formation rules being adjusted systematically to accommodate changes in an individual's attributes or disposition.

PURPOSE

This paper considers how the multi-factorial nature of group performance and the variations in desired behaviour across different circumstances can be addressed within a consistent framework.

DESIGN/METHOD

The methodology consisted of two main stages. In the first stage, a simulation was encoded in MatLab to assess the conceptual approach of progressively updating rules for group formation. The method uses an unsupervised learning algorithm and correlation factors between quantifiable group characteristics (average age, degree of motivation, etc.) and resultant behaviours of the groups that are actually formed (level of dialogue, interface interactions, etc.) to update the rules used for group formation, and hence progressively construct groups that are more likely to behave in desired ways. The second stage involved an evaluation of this approach in a real world scenario using remotely accessible laboratories where engineering students voluntarily participated in a study in April 2012.

RESULTS

The simulation results show that under certain conditions the desired behaviour chosen with the intention of improving specific learning outcomes can be optimized and that groups can be constructed that are more likely to exhibit desired behaviour. The paper also reports preliminary evidence that shows the feasibility of this approach in selecting group participants in an engineering class to promote a desired outcome in this case independent learning.

CONCLUSIONS

This study demonstrates the feasibility of using a set of individual characteristics of group members to form groups that are more likely to have desired group behaviours and that these characteristics can be monitored and updated to dynamically alter group formation to account for changes in any individual's characteristics. This has potential to allow groups formation decisions to be made dynamically to achieve a desired outcome, for example promote collaborative learning.

KEYWORDS

Group formation, collaboration, group formation rules, dynamic rules changes

Introduction

There is substantial literature that shows the benefits of collaborative work within groups, though these benefits vary enormously with circumstances (Dillenbourg, 1999; Roberts, 2005).

Collaborative learning generally refers to a situation where "two or more … people learn or attempt to learn something together". (Dillenbourg, 1999) In this paper we are interested in small groups (3 – 5 students), which perform laboratory experiments together in remotely accessible laboratories. Groups can be geographically detached from the experimentation equipment and group members might even be at different locations and possibly in different time zones. (Gomes & Bogosyan, 2009; Gravier, Fayolle, Bayard, & Ates, 2008)

Panitz (1999) points out benefits of collaborative learning, including academic, social and psychological aspects. Advancing critical thinking, better engagement of students in the learning process, and improved "problem-solving techniques" are regarded as academic benefits of collaborative learning. Social benefits include aspects such as development of conflict resolution skills. Students also develop understanding and learn how different ethnic groups approach work. A very important advantage of collaborative learning is that students can help each other when teaching staff are not available and therefore gain team-working skills which could be very important for later employment (Roberts, 2005). In our domain of interest – remotely accessible laboratories – the same assumptions regarding benefits of collaboration are made.

Throughout our lives we will regularly find ourselves as part of diverse groups or teams Cartwright (1960). Examples include business teams, project teams, club teams – for example soccer team, family, educational teams, etc. The group members will vary in their characteristics such as age, previous knowledge, motivation, problem understanding, etc. Irrespective of a group's diversity, the groups usually exist for a particular reason and implicitly or explicitly target one or more outcomes (Donelson, 1983). Generally, we can distinguish between task oriented groups and groups that are natural, such as families (Adair, 1986). For example, a business team might exist to make financial profit, or a project team could exist to achieve a particular task. A soccer team typically exists to win a match or a competition.

It is worth noting that in different teams different outcomes are important. For example, in project teams it could be argued that the project outcomes are more important than what individual team members achieve. In educational teams, it is typically more important what every member has learned and less important what the group as a whole has achieved. This is also important when grading each individual.

Constructing groups to achieve a specific outcome is complex with the optimal group composition depending on the attributes of the group members (Cole, 2001). In our literature review we have previously discussed that group learning outcomes may depend on many factors including characteristics of each member and the group composition, the task context, the individual behaviour of each group member (Mujkanovic, Lowe, Guetl, & Kostulski, 2011) In addition, we have proposed an algorithmic approach to vary the group composition and hence promote a desired behaviour (Mujkanovic, Lowe, Willey, & Guetl, 2012).

In this paper we are interested in considering the potential of adapting the formation of the groups based on knowledge gained from a correlation analysis between those characteristics of groups that we can control (such as the gender balance within the group) and the resultant behaviours of the groups that we form, particularly the behaviours that may be desired by teaching staff. Both the behaviour and the characteristics are quantitative measurements during an experiment using remote laboratories. Examples include the level of dialogue, and the number of interface interactions with the laboratory interface. The work presented in this

paper is an empirical evaluation of the algorithmic approach that is proposed in our previous work (Mujkanovic, Lowe, Willey, & Guetl, 2012) The emphasis lies in showing the general feasibility and that a desired behaviour of a student group can be encouraged, but not assured.

The remainder of the paper is organized as follows. We discuss related work in the background section then we present the methodological approach including data collection and data analysis and we finally conclude with a discussion and further work.

Background

Previous work has mostly considered group formation based on one particular attribute, such as learning style, gender, personality, etc. For example, Alfonseca, Carro, Martín, Ortigosa, & Paredes (2006) studied whether the students' learning style affects their learning outcomes and how group formation can be automated. Christodoulopoulos & Papanikolaou (2007) have introduced a tool for group formation that uses a fuzzy c-means method. In their approach it is possible to use up to three characteristics to form heterogeneous and homogenous groups.

Cocea & Magoulas (2010) use a clustering technique combined with a similarity coefficient to form groups of students. Crespo, Pardo, Pérez, & Kloos (2005) discuss a fuzzy classification approach to match reviewers and authors of scientific papers for a review process. Inaba, Supnithi, Ikeda, Mizoguchi, & Toyoda (2000) consider in their paper how an effective group with guaranteed learning benefits for each members can be formed.

Martin & Paredes, (2004) use a questionnaire base method to assign students to groups based on their learning style. In their approach they use a visual and verbal workspace. In an approach that uses the semantic web technologies Ounnas, Davis, & Millard (2007) form groups of students based on instructor desire. Pollalis & Mavrommatis (2009) use an array based clustering method to form groups with matching learning material. Rubens, Vilenius, & Okamoto (2009) introduce a method to form groups of learners by matching their learning objectives with knowledge skills of each member. Thammano & Moolwong (2010) present in their paper a computational technique that underlies the theory of social group formation and investigate the performance of three different neural network types. Zakrzewska (2009) uses an unsupervised clustering method to form groups of students with similar learning style.

The above research approaches have typically focused on dynamically creating groups, but have done so based on the optimisation of a specific characteristic or combination of a small number of characteristics. In contrast to this previous work we are interested in considering the case where we may not know, *a priori*, which group characteristics may lead to the desired behaviours. We argue that by analysing initial groups we can however dynamically learn which characteristics to optimise, and subsequently form groups that do indeed optimise those characteristics and hence become more like to exhibit desired behaviours. We propose a dynamic rule based approach that identifies the most significant characteristics of the group and progressively uses these to form subsequent groups that are more likely to exhibit the desired behaviours.

This approach can be understood by considering a simple (but hypothetical) scenario where we are creating student groups to carry out an online laboratory exercise. In allocating students to groups, we may be able to measure (and control) group characteristics such as the gender balance, age similarity, spread of home addresses, and range of previous academic performance of the members of the group. We may desire (for an educational reason that is beyond the scope of this paper to discuss) to form groups that maximise the level of interaction that the group members have with the laboratory apparatus, as measured by the number of interface control actions that occur. We do not know *a priori* what type of groups to form. However by evaluating the performance of the initial groups, as measured by

their level of interactions, we might progressively learn that forming groups that minimise the gender variation and maximise the range of previous academic performance will lead to the best performance. In other words, we are agnostic with regard to the specific learning behaviours being optimised and the group characteristics we can control, but rather are focused on how we can use machine-learning techniques to link these in a constructive way.

Methodology

Based on the previous discussion we propose a methodology that consists of two major phases. In the first phase we have previously developed a simulation (Mujkanovic et al., 2012) that provided a general theoretical approach to group formation. In this simulation we demonstrated the viability of the conceptual approach (Mujkanovic et al., 2011) and showed that it is theoretically feasible to construct groups that exhibit a behaviour that is desired by the teaching staff. The technique used for the group formation rule update is based on using multiple regression and rules that are represented as weights associated with the regression factors of each characteristic.

In the second phase reported here we have conducted a case study involving 110 students enrolled in an introductory ICT Engineering subject. The students participated voluntarily in the study. In line with good ethical practice students remained anonymous during the study and their performance during the case study had no impact on their course marks.

In the case study we were particularly interested in the range of student characteristics that could be determined (and hence used in making decisions regarding allocation of students into groups) and subsequently in the performance of the constructed groups during a particular remote laboratory session. Student groups undertook a laboratory exercise based on the use of a shake table[1] apparatus, which was accessed remotely across the Internet. It is worth noting that the use of a remotely accessible laboratory apparatus meant that we were able to quantifiably measure a broader range of group behaviours than might normally have been the case.

We began the case study by determining students' characteristics (see below) through an online survey. The entire class was then subdivided into two equal sized cohorts, and the students in cohort 1 were randomly allocated to student groups. All groups from cohort 1 then performed the laboratory experiment and their behaviour (see Table II) was captured every five minutes for the entire experiment duration. For logistical reasons the experimentation time had to be limited to 15 minutes for each group. Remote laboratories provide a mediated interface, which makes it easy to measure student online behaviour, however current remote laboratories have limited opportunities for measuring the attributes needed. Therefore, in this work all groups were present in the class and performed the laboratory experiment face-to-face and the measurements are self-reported.

The teaching staff then chose which of the measured behaviours whilst cohort 1 undertook the experiment should be optimized. The data from cohort 1 was then used as a training set for the purposes of learning what combination of group characteristics were more likely to lead to the desired group behaviours. We then used this knowledge to systematically allocate students from cohort 2 to groups. The cohort 2 groups then carried out the experiment and again we captured their behaviours. Finally, we assessed whether or not there was a significant improvement in the desired group behaviours.

---

[1] The shake table apparatus can be used to analyse the behaviour of a building during an earthquake

Data Collection

As previously discussed there were two steps in our data collection. In the first step we have determined students' characteristics. Examples include age, ATAR (Australian Tertiary Admission Rank), previous access to remote laboratories, motivation, previous knowledge, etc. In the second step we measured 5 different group behaviours during the laboratory session: level of dialogue; level of interface interactions; level of lecturer assistance; level of peer assistance; and provision of help.

Determination of group characteristics
Each characteristic was measured on a Likert scale between 1 and 5 and is based on the students' self-perception. As an example we can consider the question: *"Do you understand the problem statement of the laboratory experiment that you are going to perform?". Possible responses to this question* were: - 1 for *yes, I fully understand*, 2 for *Yes, to some extent*, 3 for *not sure*, 4 for *I will follow the experiment instructions*, and 5 for *I have no idea what to do*.

We acknowledge that the response options are imperfect as they overlap in meaning and we assume that equal numerical differences are also equal in meaning. For example, the difference between *1 Yes, I fully understand* and *2 Yes, to some extent* is the same as between *3 not sure*, *4 I will follow the experiment instructions* and *4* and *5 I have no idea what to do.*

Group characteristics were then simply determined by calculating a set of metrics derived from combinations of the group member characteristics. As previously mentioned the core of this work is a system-learning algorithm that learns which student characteristics would encourage a desired behaviour. No conclusions are drawn on the usefulness of particular student characteristics and student behaviour. The collected data is used to evaluate the algorithm and is therefore not essential in this research. Table I presents illustrative individual and group characteristics for the first 5 groups in cohort 1.

Determination of the group behaviour

Each group member was asked to fill out a form that captured the individual behaviour during the entire experiment. Based on their perception they answered 1 for *extremely high*, 2 for *high*, 3 for *moderate*, 4 for *low* and 5 for *extremely low*. We expected that the students' behaviour would vary throughout the experiment (e.g. students might talk more in the beginning to clarify the problem statement) therefore we have measured their behaviour every five minutes for the entire duration of the experiment. The group behaviour, as shown in Table II, is then derived from a simple sum of each individual's behaviour over the entire experimentation period (though more complex variations could also be considered).

Data Analysis

Determining the group metrics

For each individual characteristic *C* a number of metrics has been determined. These metrics *m* include the minimum *min(C)*, the maximum *max(C)*, the average *mean(C)*, and the difference between the maximum and the minimum *difference = max(C) - min(C)* of each value as illustrated in Table I.

Table I. Problem understanding, Group IDs, and Group Metrics

Group ID	Individual characteristics			Group metrics			
	Participant 1	Participant 2	Participant 3	Min	Max	Average	Difference
1	4.00	2.00	1.00	1.00	4.00	2.33	3.00
2	2.00	3.00		2.00	3.00	2.50	1.00
3	2.00	2.00	2.00	2.00	2.00	2.00	0.00
4	2.00	2.00	2.00	2.00	2.00	2.00	0.00
5	3.00	5.00	3.00	3.00	5.00	3.67	2.00

Table II illustrates the collected behaviour of five groups after performing the laboratory experiment. The behaviour can vary between 1 and 5 where 1 represent an extremely high level and 5 a very low of the corresponding behaviour.

Table II. Group behaviour

Group ID	Dialogue	Interface interactions	Lecturer assistance	Peer assistance	Provision of help
1	2.67	2.67	4.11	3.56	2.67
2	2.78	2.67	1.00	1.89	2.00
3	1.44	1.22	2.78	1.67	1.33
4	2.00	2.44	4.11	2.33	2.44
5	3.44	3.11	3.11	3.78	3.22

Calculation of the Pearson correlation factors

Once the data on *group characteristics* and *group behaviours* was collected, correlation analysis has been applied to determine correlations factors. An example of the correlations for one characteristic (the overall degree of understanding of the problem by the group members) is presented in Table III.

Table III. Pearson correlation factors between group metrics and group behaviour

Group Behaviour	Min (problem understanding)	Max (problem understanding)	Mean (problem understanding)	Difference (problem understanding)
Dialogue	0.18	0.55	0.51	0.42
Interface interactions	0.22	0.45	0.47	0.30
Lecturer assistance	0.12	0.07	0.05	-0.01
Peer assistance	0.24	0.60	0.52	0.44
Provision of help	0.44	0.60	0.69	0.31

Selection of the behaviour to optimize

In our particular case study, the lecturer's objective was to develop students more into independent learners, and hence desired groups that exhibited a high level of peer assistance. The lecturing staff selected then this behaviour as the optimization criteria for a systematic group formation of cohort 2.

In order to find the optimal group all possible combinations using the binomial coefficient were calculated. We have chosen to maximize the difference in problem understanding of

each group and expected then a higher correlation factor of subsequent groups. Out of 2300 possible group combinations we have formed the ones with the highest score as illustrated in Table IV. The same table also shows all systematically formed groups including the group's score and the correspondent behaviour of each group that has been measured during the laboratory experiment. The group score is simply the difference of problem understanding of each group member of the potential group. Only groups with the highest score are accepted. Note that in this live study no threshold was set and all groups were accepted to perform the laboratory experiment. For this scenario (when there are no groups that pass a threshold) there is a mechanism in the simulation that adjusts the score of each group depending on the time a group has been waiting in queue. Even if the score is below the threshold, each group will be able to perform the experiment.

Table IV. Cluster 2 Systematically formed groups, score and group behaviour

	Group Members				Behaviours				
Group ID	Participant ID	Participant ID	Participant ID	Score	Dialogue	Interface interactions	Lecturer assistance	Peer assistance	Provision of help
17	49	51	56	4	2.33	2.33	4.00	4.78	3.67
18	50	58	69	2	3.89	3.33	2.00	1.78	2.78
19	53	59	75	1	1.78	1.78	2.44	2.22	1.44
20	55	62	83	1	3.00	2.56	4.89	3.56	3.00
21	60	63	77	1	1.78	2.56	2.56	2.00	2.22
22	64	65	79	1	1.78	1.89	3.67	2.56	2.22
23	66	68	86	1	2.78	2.67	3.67	3.00	3.33
24	71	87	88	1	2.78	2.33	4.33	3.33	2.78

By correlating the behaviour of the systematically formed groups it can be seen that the correlation factor has increased from 0.44 to 0.61. Interestingly, even a higher increase of the correlation factor between provision of help and the difference in problem understand can be noted. As shown in Table V all other factors decreased by a significant amount. It is worth noting that despite the smaller sample in the second cluster the Pearson correlation factors have changed as intended.

Table V. Pearson correlation factor of systematically formed groups

Behaviour	Difference (problem understanding)
Dialogue	0.14
Interface interactions	0.16
Lecturer assistance	0.02
Peer assistance	**0.61**
Provision of help	0.58

Discussion and Further Work

This paper presented a real world study of our approach to correlation based group formation. It shows evidence that group characteristics and group behaviour can be used to form groups and encourage a desired behaviour. There is a high potential for this method in the area of large online course where students participate in online courses from multiple countries.

The data presented in this paper is self-reported. It would be useful if remote laboratories had user accounts so that the meta-data could be determined from their profiles. In addition, the group behaviour could be automatically monitored and saved into log files. This data from both user profiles and log files could then be used for an automated and even more sophisticated group formation.

For logistical reasons the student behaviour in this study the group behaviour was measured over a 15 minutes period. This could be extended and monitored over a longer period of time and saved in log files.

The preliminary results discussed in this paper suggest that there is a higher likelihood to achieve the desired learning outcomes by systematically encouraging the behaviour that could address a desired learning outcome.

Conclusions

This study demonstrates the feasibility of using a set of individual characteristics of group members to form groups that are more likely to have desired group behaviours and that these characteristics can be monitored and updated to dynamically alter group formation to account for changes in any individual's characteristics. This has potential to allow group formation decisions to be made dynamically to achieve a desired outcome, for example promote collaborative learning.

In our study we could find increasing correlation factors between the variable that were intended to increase where all other correlation factors decreased. It is however unclear why this happened. This approach to group formation has a high potential for application in massive online courses.

References

Adair, J. (1986). *Effective teambuilding: How to make a winning team.* London: Pan Books.

Alfonseca, E., Carro, R. M., Martín, E., Ortigosa, A., & Paredes, P. (2006). The impact of learning styles on student grouping for collaborative learning: a case study. *User Modelling and User-Adapted Interaction*, *16*(3-4), 377–401. doi:10.1007/s11257-006-9012-7

Cartwright, D. (1960). *Group Dynamics.* Harper & Row. Retrieved from http://books.google.com.au/books?id=0vcNAAAAQAAJ

Christodoulopoulos, C. E., & Papanikolaou, K. a. (2007). A Group Formation Tool in an E-Learning Context. *19th IEEE International Conference on Tools with Artificial Intelligence(ICTAI 2007)* (Vol. 2, pp. 117–123). IEEE. doi:10.1109/ICTAI.2007.155

Cocea, M., & Magoulas, G. D. (2010). Group formation for collaboration in exploratory learning using group technology techniques. *Knowledge-Based and Intelligent Information* (pp. 103–113). Retrieved from http://www.springerlink.com/index/60007R2N41641475.pdf

Cole, G. A. (2001). *Organisational Behaviour*. Continuum. Retrieved from http://books.google.com.au/books?id=K5EKfJaZqgwC

Crespo, R. M., Pardo, A., Pérez, J. P. S., & Kloos, C. D. (2005). An algorithm for peer review matching using student profiles based on fuzzy classification and genetic algorithms. *Innovations in Applied Artificial Intelligence*, 685–694. Retrieved from http://www.springerlink.com/index/TYC6N77QD0MXKVF8.pdf

Dillenbourg, P. (1999). *Collaborative Learning: Cognitive and Computational Approaches*. Pergamon. Retrieved from http://books.google.com.au/books?id=Zq4wQgAACAAJ

Donelson, R. F. (1983). *Group Dynamics* (2nd ed.). Pacific Grove, California: Brooks/Cole Publishing Company.

Gomes, L., & Bogosyan, S. Current Trends in Remote Laboratories. , 56 IEEE Transactions on Industrial Electronics 4744–4756 (2009). IEEE. doi:10.1109/TIE.2009.2033293

Gravier, C., Fayolle, J., Bayard, B., & Ates, M. (2008). State of the art about remote laboratories paradigms-foundations of ongoing mutations. *International Journal of Online Engineering (iJOE)*, *4*(1), 19–25. Retrieved from http://hal.archives-ouvertes.fr/docs/00/31/96/12/PDF/rlab_proposedguideline.pdf

Inaba, A., Supnithi, T., Ikeda, M., Mizoguchi, R., & Toyoda, J. (2000). How can we form effective collaborative learning groups? *Intelligent Tutoring Systems* (pp. 282–291). Springer. Retrieved from http://www.springerlink.com/index/00XNX7Y4T7UYA1VQ.pdf

Martin, E., & Paredes, P. (2004). Using learning styles for dynamic group formation in adaptive collaborative hypermedia systems. *Proceedings of the First International Workshop on Adaptive Hypermedia and Collaborative Web-based Systems (AHCW 2004)* (pp. 188–198). Citeseer. Retrieved from
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.106.9315&rep=rep1&type=pdf

Mujkanovic, A., Lowe, D., Guetl, C., & Kostulski, T. (2011). An architecture for automated group formation within remote laboratories. *Remote Engineering & Virtual Instrumentation* (pp. 91–100). Brasov, Romania.

Mujkanovic, A., Lowe, D., Willey, K., & Guetl, C. (2012). Unsupervised Learning Algorithm for Adaptive Group Formation: Collaborative Learning Support in Remotely Accessible Laboratories. *IEEE International Conference on Information Society* (pp. 59–66). London, UK: IEEE.

Ounnas, A., Davis, H. C., & Millard, D. E. (2007). Towards Semantic Group Formation. *Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007)*, (Icalt), 825–827. doi:10.1109/ICALT.2007.268

Panitz, T. (1999). *The case for student centered instruction via collaborative learning paradigms*. Retrieved from http://home.capecod.net/~tpanitz/tedsarticles/coopbenefits.htm

Pollalis, Y. A., & Mavrommatis, G. (2009). Using similarity measures for collaborating groups formation: A model for distance learning environments. *European Journal of Operational Research*, *193*(2), 626–636. doi:10.1016/j.ejor.2007.11.053

Roberts, T. (2005). *Computer-supported collaborative learning in higher education*. Hershey: Idea Group Publishing. doi:10.4018/978-1-59140-408-8

Rubens, N., Vilenius, M., & Okamoto, T. (2009). Automatic Group Formation for Informal Collaborative Learning. *IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, *3*, 231–234. doi:10.1109/WI-IAT.2009.270

Thammano, A., & Moolwong, J. (2010). A new computational intelligence technique based on human group formation. *Expert Systems with Applications*, *37*(2), 1628–1634. doi:10.1016/j.eswa.2009.06.046

Zakrzewska, D. (2009). Cluster Analysis in Personalized E-Learning. *Intelligent Systems for Knowledge Management*, *252*, 229–250.

Acknowledgements

Copyright statement