Semantic Web-based Group Formation for E-learning

Asma Ounnas

School of Electronics and Computer Science
University of Southampton, UK
ao05r@ecs.soton.ac.uk

1. The Research Problem

Collaboration has long been considered an effective approach to learning. However, forming optimal groups can be a time consuming and complex task. Teachers often need to set some constraints for the grouping based on the aim of the collaborative task. To achieve optimal grouping, the formation needs to satisfy these constraints, even when the list of students is unknown. In this research, we investigate the use of Semantic Web technologies to assist teachers in overcoming this problem. In particular we investigate the following problems with forming groups:

- Describing the students: how do we describe the students in a way that is meaningful to the group formation.
- Specifying and satisfying the constraints: how do we model the constraints for the different collaboration goals, and how can we form optimal groups by satisfying the given constraints.
- Avoiding *"Orphan Students"*: when assigning students to groups, some students remain unassigned to any groups, often because some constraint has been violated. This problem is known as the orphan students.
- Handling the formation with incomplete data: If the students do not provide the relevant data, how do we ensure that the formation is still efficient.

We know that the Semantic Web aims at providing a promising foundation for enriching resources with well defined meanings and inferring new data from existing one. Therefore, we study the use of Semantic Web technologies and mainly ontologies and deduction rules in describing students and handling incomplete data in constraint-based group formation. The challenges of the research reside in applying the potential of the Semantic Web is real life applications such as forming groups for learning; especially with the growing use of collaboration applications over the Web.

2. Related Work

Group formation is a well-known problem in various disciplines including Psychology, Philosophy, social studies, Economics, and Education. In learning, different applications have been developed to automate the process of allocating participants to groups as tool for Computer Supported Collaborative Work. Most of the existing applications follow a self-selecting formation approach, where the learner

selects the potential learners that can assist him or her in achieving the learning goal, and the selected learners get to accept or reject joining the group [1], [2]. These systems usually model the learners' context, experience, and previous performance in the subject of the collaboration. Similar systems employ agents to negotiate the participation in the groups [3], which facilitates the dynamic formation of the groups (coalition formation) through agents' communication and decision-making [4]. In terms of constraint-based formation, we argue that:

- Existing systems only model a fixed set of constraints.
- Most of these systems are based on self-selecting group formation [1], [3] which is not the most efficient approach in forming teams, as it does not ensure balanced grouping, and usually end with some students being unassigned to any group (orphan students).
- Most systems use *Opportunistic Group Formation* concepts where the system initiates the collaboration and sets up a learning goal for the learner [2], [1], [4], [3]. OGF ensures the satisfaction of the participants in the group through negotiation, but does not discuss the efficiency of the negotiation if all students are grouped simultaneously. Furthermore, OGF is usually more beneficial in short-term groups.
- In [5] and [6], the authors introduced tools that assign all the students in the class to groups simultaneously. However, although these applications only model a limited number of constraints, their evaluation showed that manual corrections to the results were needed due to the appearance of orphan students in the generated formation.

3. Expected Contribution

The theoretical contribution of this research is a study of the feasibility and usefulness of employing Semantic Web technologies for group formation within a complex domain such as learning, and particularly in handling incomplete data. To overcome the complexity of allocating students to groups, we provide a framework to assist the teacher in forming groups based on their chosen set of constraints. The framework handles the group formation process based on the following concepts:

Modeling the students' features: We model a large range of features that can be considered for different group formations using a number of domain ontologies, which can form a reliable dynamic learner profile [7]. In this context, semantic modeling provides meaningful descriptions of the students and the relationships between them. Examples of the modeled features are: personal details, course details, interests, team roles, preferences, friends, collaborators, trust ranking, and so on [7]. The ontologies we use are based on *Friend Of A Friend* (FOAF), an existing ontology that describes people for building social groupings. This allows us to identify the relations between the participants in order to form groups from social networks. This feature allows the teacher to be aware of the social connections between the students and therefore controlling the group dynamics, or detecting plagiarism.

Negotiating the group formation: We model the students' allocation problem as a *Constraint Satisfaction Problem (CSP)* [8] with strong and weak constraints. The negotiation process can be then handled by a constraint satisfaction solver. We emphasize that, in this research, we are not concerned with proving that any particular

set of constraints leads to better results in terms of the performance of the groups; neither do we claim that any particular algorithm leads to best grouping.

Handling Incomplete Data: The semantic representation of students' data, to which the instructor constraints can be mapped to, allows inferences to generate more data. We use domain ontologies and deductions rules for substituting missing data with data mined from the Web. For example, if the information about whether student John is a leader or not is missing, and we know from John's web page that he is a captain of the football team, then we can infer that John is a leader; or if we are grouping student by skills, and we don't know Sarah's skills, but we know that Sarah has a high grade in discrete mathematics and Sarah has a high grade in Logic then we can infer that Sarah will perform well in formal methods.

Calculating the group formation quality: To evaluate the generated group formation, we provide a metrics framework for calculating its quality in terms of the satisfied constraints [9], and hence the collaboration goals set by the teacher. Using these metrics, the teacher can check the confidence of the group formation framework in generating the groups.

In general, although it is applied to learning, this research can be employed in other domains as a solution to any type of constrained group formation. When completed, the system will form a standard semantic technology that allows groups of users to be generated based on a set of constraints and a range of information about themselves.

Research Methodology

In the early stage of the research, we run an observational study to analyze the different constraints teachers consider when forming groups. We studied the possible students' features that can be relevant to forming different types of groups by investigating the available literature on collaborative learning theories [7], and asking teachers what constraints they employ for different educational goals. We then gave a class of 66 undergraduate students some questionnaires to monitor the data they provide for the grouping and their satisfaction with the groups at the end of course. The study enabled us to realize the depth of the problem and the pedagogical issues that accompany it. We then modeled the group formation problem as a constraint satisfaction problem to be implemented as the semantic (group formation) framework [9], [10]. We also reviewed the different techniques for evaluating group formation, and provided a model for the formation quality metrics framework in terms of the formation goals and hence the constraints satisfaction [9]. For the next step, we started implementing the group formation framework [10] based on the following:

The Student Interface: The student can enter their data through a web-based form composed of the student's personal data, a list of their friends, their interests and preferences, and information about their course such as the modules they are taking.

The Ontology: We created an ontology called *Semantic Learner Profile* (SLP) that extends FOAF with a description of a large range of student's personal, social, and academic data such as learning styles and collaborators. We also use the trust ontology (http://trust.mindswap.org/ont/trust.owl) to allow the students to rank their trust towards each other in specific topics. Following the vision of James Hendler in

reusing and sharing small ontological components instead of large complex ontologies [11], we intend to enrich our learner profile with more features by employing other domain ontologies (competency and interest topics ontologies). Once the student submits the profile data through the interface, an RDF file is created (FOAF+SLP), and processed using Jena, a Semantic Web inference engine.

The Instructor Interface: Through this web-based interface, the instructor is given a degree of freedom in selecting the constraints they care about for the formation they are initiating. They are provided with an option that enables them to set a priority value for each constraint. Ranking the importance of the constraints enables the application to manage compromises based on these priorities. The constraints are then written as a constraint satisfaction problem in the group generator.

The Group Generator: The generator is based on a DLV solver, an implementation of disjunctive logic programming, used for knowledge representation and reasoning. DLV's native language is *Disjunctive Datalog* extended with constraints, true negation and queries [12]. DLV performs a simple forward checking algorithm [8] on the data provided to process the groups. The use of strong and prioritized weak constraints in DLV enables the framework to always generate a solution with all students allocated even if some of the weak constraints are violated [13]. This avoids the orphan students' problem. The solver returns the optimal group formation that minimizes the number of violated constraints and returns the list of the violated constraints, which can be used in calculating the group formation quality.

For our future work, we plan to add a module to the architecture of the framework that mines data from web pages and connect it to the ontology and a set of deduction rules to infer the missing data from the knowledge base. In this case, if the needed data is incomplete, the system will substitute the necessary data and subsequently feed it to the solver. So far, we evaluated the framework with two classes of undergraduate students. However, since the teachers had a maximum of three constraints, the framework returned a best model in both cases with violation of one constraint for one group in both courses. Future evaluation of the framework will include running it with different scenarios on simulated classes of students. The simulated data will be based on the population statistics collected from our observational study. The framework will be tested with various constraints, different in content and number. Since groups can also be generated from social networks, a range of constraints will be based on the social connections between the learners. We intend to use the metrics framework we introduced in [9] to record the formation quality for the evaluation. Once the framework is refined with deduction rules, the evaluation of its performance with incomplete data will be compared to its performance with complete data (and no deduction rules), and its performance with incomplete data (and no deduction rules). For handling incomplete data, due to privacy issues, we aim at using students' web pages from the University of Southampton as a base for our mining.

Conclusion and Future Work

In this research, we propose an approach to learner group formation, based upon satisfying the constraints of the person forming the groups by reasoning over possibly

incomplete semantic data about the potential participants. We are currently evaluating the semantic (group formation) framework with complete data. Within the next few months, we intend to implement extensions to allow for handling incomplete data, and for forming groups from social networks. The research can then be fully evaluated and results published to the community with more results in more depth.

We believe that by reasoning on learners' profiles and the teacher's constraints, we can achieve a powerful foundation for automated group formation. The use of Semantic Web technology demonstrates the powerful characteristics of the Semantic Web that can be put in use to facilitate daily life tasks such as allocating students to groups for collaborative learning. The use of the Semantic Web in this domain can be extended to other areas of group formation such as forming teams within organizations, sports, or even military. The research can also be extended to other constraint satisfaction problems where data is key to the solution of the problem. The interoperability of the Semantic Web facilitates the use of such an application in different platforms and systems, even when the participants are geographically distributed. For this, the only challenge to applying this research in other areas is the development of domain ontologies and deduction rules.

References

1. Hoppe, H.U. Use of multiple student modeling to parametrize group learning, In proc of AI-ED 95, AACE, Charlottesville, VA, USA, 1995, pp. 234-241.
2. Wessner, M. and H-R. Pfister, Group formation in computer-supported collaborative learning, Proc. of ACM SIGGROUP, Boulder, Colorado, USA, 2001, pp. 24 - 31.
3. Inaba, A., T. Supnithi, M. Ikeda, R. Mizoguchi, and J.I. Toyoda. How Can We Form Effective Collaborative Learning Groups? In ITS, Montréal, Canada, 2000, pp. 282-291.
4. Soh, L.K., N. Khandaker, X. Liu, and H. Jiang. A Computer-Supported Cooperative Learning System with Multiagent Intelligence. In AAMAS'06, Japan, 2006, pp. 1556-1563.
5. Redmond, M.A., A computer program to aid assignment of student project groups, In proc of ACM SIGCSE, Charlotte, NC, USA, 2001, pp. 134-138.
6. Tobar, C.M., de Freitas, R.L. A support tool for student group definition. In Proceedings of the 37th ASEE/IEEE Frontiers in Education Conference. 2007.
7. Ounnas, A., Davis, H. C. and Millard, D. E. Semantic Modeling for Group Formation. In: PING Workshop at the 11th UM2007 conference, Corfu, Greece, 2007
8. Kumar, V. Algorithms for Constraint Satisfaction Problems: A Survey. In AI Magazine, Vol 13. Issue 1, 1992, pp. 32-44.
9. Ounnas, A., Millard, D. and Davis, H. A Metrics Framework for Evaluating Group Formation. In Proc of ACM Group'07, Sanibel Island, Florida, USA, 2007, pp 221-224.
10. Ounnas, A., Davis, H. C. and Millard, D. E. A Framework for Semantic Group Formation. To appear in the 8th IEEE ICALT'08, Santander, Cantabria, Spain. 2008.
11. Hendler, J., Agents and the Semantic Web. IEEE Intelligent Systems, 2001 pp.30-37.
12. Leone, N., Pfeifer, G., Faber, W., eiter, T., Gottlob, G., Perr, S., Scarcello, F. The DLV system for knowledge representation and reasoning. In ACM Transactions on Computational Logic (TOCL), Vol 7, Issue 3, 2006, pp. 499 - 562.
13. Buccafurri, F., Leone, N., Rullo, P. Strong and weak constraints in disjunctive Datalog. In Proc of the 4th International Conference on Logic Programming and Nonmonotonic Reasoning. In LNCS, Vol. 1265. 1997, pp. 2 -17.