Simulating a collective intelligence approach to student team formation

Juan M. Alberola, Elena del Val, Victor Sanchez-Anguix, Vicente Julian

Departament de Sistemes Informàtics i Computació,
Universitat Politècnica de València,
Camí de Vera s/n. 46022, València. Spain
{jalberola,edelval,sanguix,vinglada}@dsic.upv.es

Abstract. Teamwork is now a critical competence in the higher education area, and it has become a critical task in educational and management environments. Unfortunately, looking for optimal or near optimal teams is a costly task for humans due to the exponential number of outcomes. For this reason, in this paper we present a computer-aided policy that facilitates the automatic generation of near optimal teams based on collective intelligence, coalition structure generation, and Bayesian learning. We carried out simulations in hypothetic classroom scenarios that show that the policy is capable of converging towards the optimal solution as long as students do not have great difficulties evaluating others.

1 Introduction

In the last few years, educational organizations have shown a growing interest in shifting towards teaching paradigms that promote teamwork. In fact, teamwork is now considered as a prominent general competence in the European Higher Education Area [12, 15]. The inclusion of teamwork as a general competence responds to two main reasons: (i) it is strongly related to cooperative learning, a methodology that has been reported to enhance learning in the classroom [10, 9]; (ii) nowadays, most successful business and engineering projects are carried out by small multidisciplinary teams [18, 4]. Nevertheless, teamwork is a *double-edged sword* that can bring both positive and negative consequences [7], making the task of team formation a complex one. Several factors like personality, expertise, competitiveness, and human behavior can interfere with the performance of the team [13].

Therefore, it is of crucial importance to identify teams that can perform correctly. Several scholars have studied under what circumstances positive teamwork can emerge [16, 2]. One of these studies is the well-known Belbin's role inventory [2]. Belbin identifies nine behavior patterns (i.e., roles) that are useful for teams: plants, resource investigators, coordinators, shapers, monitors, teamworkers, implementers, finishers, and specialists. As some studies suggest, teams usually benefit from having a varied mix of roles [2, 8].

Team formation can also be a cognitive complex task. For instance, a class formed by 30 students can yield up to 767746 different teams of sizes 3, 4, 5 and 6[1], and the

[1] It corresponds to the sum of the number of combinations of size 3, 4, 5, and 6 ($\sum_{i=3}^{6}\binom{30}{i}$).

number of different partitions of the class in disjoints teams grows exponentially with the number of possible teams. The problem is especially complicated if teachers want to find an optimal or near optimal team allocation. Professionals in education would certainly benefit from policies and tools that guide the team formation process.

In this paper, we simulate a computer-aided policy for forming disjoint teams of students based on collective intelligence, coalition formation, and bayesian learning. More specifically, students grade other teammates based on Belbin's role inventory after the completion of each group activity. The information regarding the predominant roles of each student is then updated via Bayesian learning. Teams are formed following a coalition formation mechanism that employs the information inferred via the Bayesian learning process.

The remainder of the article is organized as follows. Section 2 presents the policy workflow. Section 3 describes the formalization of the policy and the Bayesian learning method. Section 4 shows an experimental evaluation in order to validate the policy. Section 5 details some previous works related to our proposal. Finally, Section 6 presents some concluding remarks.

2 Policy Overview

As stated, the proposed policy relies on collective intelligence, coalition formation, and Bayesian learning to form proper distributions of students teams. This policy is aided by a software application that prevents teachers from carrying out the costly task of dividing students into optimal or near optimal teams. The workflow of tasks followed by the policy can be observed in Figure 1. Next, we describe the workflow in more detail.

1. At the start of the course, there is no prior information regarding the natural preferences of each student for one of Belbin's roles. The policy starts with the first group activity of the course.
2. The teacher indicates to the computer tool that there is a new group activity. Then, students are divided into disjoint teams for the task at hand. As we will observe in Section 3, the problem of dividing students into optimal and disjoint teams is casted to a coalition structure generation algorithm.
3. After that, the activity is carried out.
4. Once the activity is finished, each student should log in the software application. Then, he/she is presented with a full description of Belbin's roles. At that point, the student should classify each teammate into a role. The information is then gathered by the application and stored in a database.
5. After the peer evaluation, the application updates the information for each student. Then, Bayesian learning is applied to determine which roles are more predominant for each student (Section 3.2).
6. The process continues (go to step 2) until the course has finished.

3 Policy formalization

In this section, we formalize our policy for dividing students into disjoint teams. First, we describe how dividing students into optimal teams is equivalent to a coalition struc-

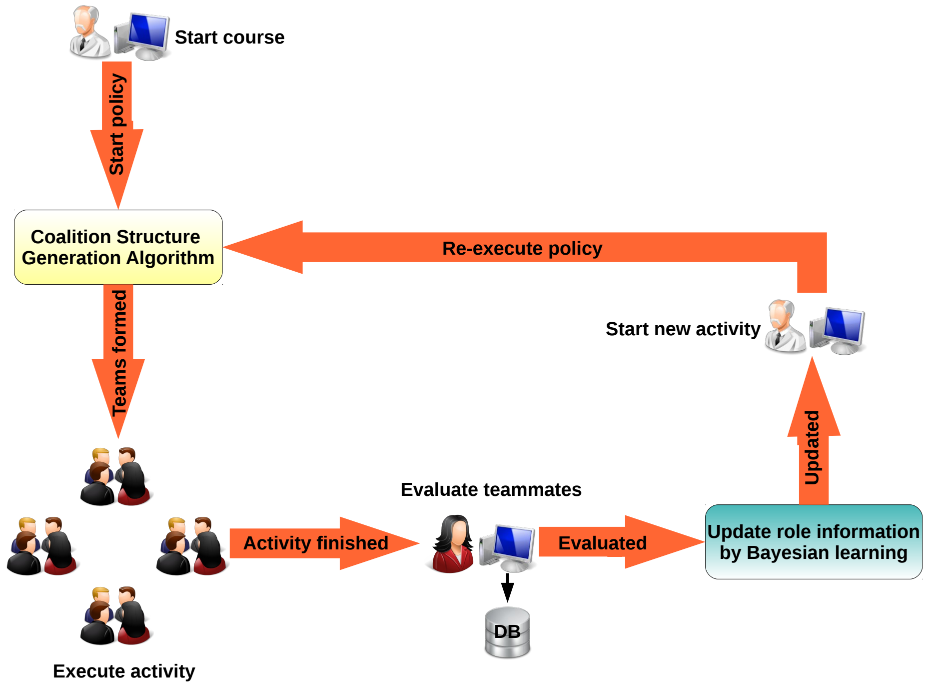

Fig. 1: An overview of the solution applied for grouping students into teams.

ture generation problem. Then, we describe how Bayesian learning is employed in order to update the information of the classroom.

3.1 Student team formation as a coalition structure generation problem

Let $A = \{a_i, ..., a_n\}$, be a set of students, and $R = \{r_1, r_2, ..., r_m\}$ be the set of roles that the student can play (in our case it is the set of Belbin's roles), and let $role_i$ denote the true predominant role of a_i.

A subset $T \in A$ is called a *team*, and a *team structure* $S = \{T_1, T_2, ..., T_k\}$ is a partition of disjoint teams such that $\bigcup_{\forall T_j \in S} T_j = A$ and $S \in 2^A$. The goal of the application is determining an optimal team structure for the classroom $\underset{S \in 2^A}{argmax}\ v(S)$, where $v(S)$ is a evaluation function for the team structure. In this study, we will consider that the quality of each team is independent of other teams. Hence, we can calculate the value of the team structure as $v(S) = \sum_{T_j \in S} v(T_j)$. The value of a team $v(T_j)$ can be calculated attending to the predominant role that each student $a_i \in T_j$ has $(role(a_i))$. Let $|T_j| = k$ denote the size of the team and $\pi_j = \{r'_1, ..., r'_k\}$ with $\forall r'_i \in R$ be a vector with the true predominant role of each team member. In that case, $v(T_j) = v(\pi_j)$. According to different studies [8], the team should benefit from having a balanced distributions of roles (i.e., one person per role). This score can be provided by an expert.

Unfortunately, it is not possible to accurately know the predominant role of each team member π_j and therefore $v(\pi_j)$ cannot be calculated with precision. However, it is possible for us to calculate an estimation of the value of the coalition given the history of evaluations H that is gathered from the students during the course. Let $\pi' = \{role_1 = r'_1, ..., role_k = r'_k\}$ be a vector containing a set of hypotheses for the predominant roles of each team member, and Π be the set of all possible vectors of hypotheses for predominant roles of T_j. In that case, we can calculate the expected value of a team given the history of evaluations as:

$$\hat{v}(T_j|H) = \sum_{\pi' \in \Pi} p(\pi'|H) \times v(\pi') = \sum_{\pi' \in \Pi} v(\pi') \times \prod_{a_i \in T_j} p(role_i = r'_i|H) \quad (1)$$

where and $p(\pi'|H)$ represents the probability for π' to be the real role distribution in T_j given the history of evaluations H. Each $p(\pi'|H)$ can be divided into its $p(role_i = r'_i|H)$ since we assume that the role of each student is conditionally independent given the history of evaluations. Therefore, our team formation problem at each iteration is casted out to one problem that follows the next expression:

$$\underset{S \in 2^A}{argmax} \sum_{T \in S} \hat{v}(T|H) \quad (2)$$

It turns out that partitioning a set students into disjoint teams while optimizing a social welfare function corresponds to the formalization of coalition structure generation problems. For our simulation experiments, we formalize the coalition structure generation problem as a linear programming problem [14] and solve it with the commercial software *ILOG CPLEX 12.5* [2].

3.2 Bayesian learning

After every activity, students evaluate their peers by stating the most predominant role of each of his/her teammates. Then, new information becomes available regarding the most predominant role of each student and the history of evaluations H grows. Hence, at each iteration we can update information regarding the probability for an agent a_i to have r'_i as his/her most predominant role given the evaluation history $p(role_i = r'_i|H)$. We employ Bayesian learning for this matter :

$$p(role_i = r'_i|H) = \frac{p(H|role_i = r'_i) \times p(role_i = r'_i)}{\sum_{r \in R} p(H|role_i = r) \times p(role_i = r)} \quad (3)$$

where $p(H|role_i = r'_i)$ is the likelihood function and $p(role_i = r'_i)$ is the prior probability for the hypothesis. For the likelihood function, we can calculate it as $p(H|role_i = r'_i) = \frac{\#\{r'_i \in H_i\}}{|H_i|}$, where H_i denotes the peer evaluations about agent a_i, and $\#\{r'_i \in H_i\}$ indicates the number of times that r'_i appears as evaluation in H_i. As for the prior probability, we calculate it as $p(role_i = r'_i) = \frac{\#\{r'_i \in H\}}{|H|}$. Laplace smoothing [17] is employed to ensure that the likelihood for each role hypothesis can be calculated in the first iterations.

[2] http://www.ibm.com/software/commerce/optimization/cplex-optimizer/

4 Simulation experiments

In this section, we simulate different classroom scenarios to study the behavior of the proposed policy throughout different group activities carried out in a course. First, we describe the general experimental setting, and then we describe the different experiments that we carried out and its results.

4.1 Experimental setting

For the experiments, we simulate a classroom with $|A| = 20$ students and we employ all of the Belbin's role except the specialist $|R| = 8$, since we consider that no student has specialized knowledge for the subject at hand.

The objective of policy is obtaining the optimal team structure. According to the Belbin's taxonomy, this value is higher when a team is composed by heterogenous roles, i.e. when the predominant roles played by its teammates are different between themselves. As we stated in Equation 2, the expected value of a team structure is defined as the aggregation of the expected values of each individual team T given the history of evaluations H. As for the expected value of teams (see Equation 1), it depends on the aggregation of the probability of each possible role distribution π_j multiplied by the evaluation by an expert of such role distribution $v(\pi_j)$. For these simulations, we define $v(\pi_j) = \frac{MAX_v}{2^{\gamma-1}}$, where $MAX_v = \frac{\#\{\text{different } r \in \pi_j\}}{|T_j|}$ is the number of different roles in π_j divided by the number of team members, and γ is the number of teammates playing repeated predominant roles. This way, we penalize those teams with less diversity.

In the policy, students classify other teammates according to Belbin's taxonomy after the finalization of each activity. This process is simulated via Bernoulli distributions. Similarly to [1], each team member classifies other teammate into its corresponding predominant role according to a probability $\rho = [0..1]$. This probability is associated to the number of teammates with the same predominant role. In more detail, the probability of classifying a teammate into its predominant role is higher when no other teammate has the same predominant role. We set this probability to decrease with the number of teammates playing the same predominant role.

4.2 Results

In the first experiment we study scenarios where the distribution of predominant roles among students is uniform and students are grouped into teams of size 4. The simulation of each scenario is repeated 10 times to capture stochastic variations. In Figure 2 we show scenarios with different classifying probabilities $\rho = \{0.125, 0.25, 0.50, 0.75\}$. On the one hand, $\rho = 0.125$ represents a scenario where team members classify other team members randomly [3] and $\rho = 0.25$ represents a scenario where students have difficulties to classify other teammates. On the other hand, $\rho = 0.5$ represents a scenario where there is an average difficulty for students to classify other team members, and $\rho = 0.75$ represents a scenario where team member easily classify other team members. The y-axis shows the average team structure value normalized between 0 and 1.

[3] It is equivalent to a random classification since $\rho = \frac{1}{|R|} = \frac{1}{8}$.

We represent the average expected value of the solution found by the learning policy. Additionally, we also draw the value of the best team structure that can be found in the scenario.

As it can be observed, the real value keeps fluctuating when $\rho = 0.125$ since students provide no reliable information regarding other students. However, when $\rho = 0.25$, the real value found by the policy gradually converges towards the optimal solution. With just 5 iterations, the quality of the solution has improved a 10% over the initial solution. When $\rho = 0.5$ and $\rho = 0.75$, it can be observed that the real value converges in two iterations to the optimal value, approximately improving the initial solution a 30%. Since ρ is higher, the information provided by students can be considered reliable and the policy finds an optimal team structure faster. One important finding is that even with just 5 iterations, which represents a reasonable number of activities for a course, the policy is able to find reasonable good results even when students have difficulties evaluating other students.

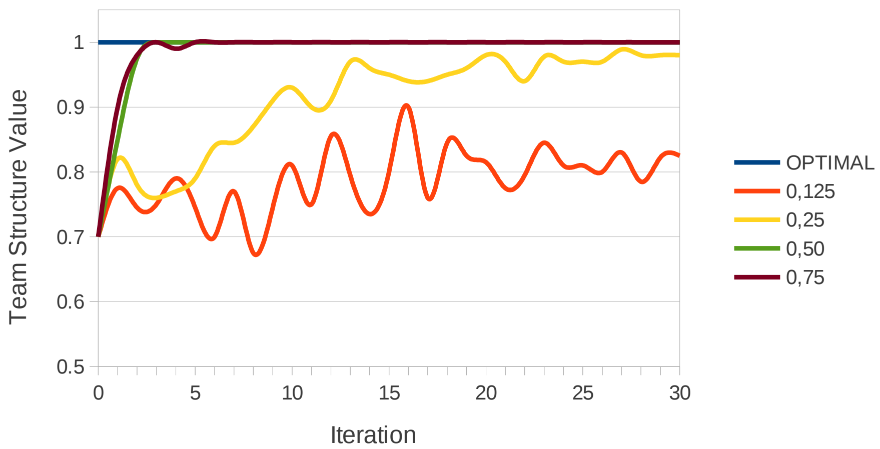

Fig. 2: Average team structure value found by the proposed policy when the role distribution among students is uniform.

In the second experiment we test scenarios where the distribution of predominant roles is not uniform. More specifically, we set 3 out of the 8 roles to account for 50% of the student population. We tested scenarios where students have difficulties to evaluate other students $\rho = 0.25$ and scenarios where students have an average difficulty evaluating others $\rho = 0.5$. The rest of the parameters are set to the same value than in the previous experiment. We can observe the results for this experiment in Figure 3. For comparison purposes, we also include the analogous results for uniform distributions of roles among students. It can be observed that when roles are uniformly distributed, the policy converges more quickly towards the optimal solution than the analogous case in the non-uniform scenario. This can be explained due to the fact that the probability of putting students with the same role in a team is higher, then reducing the probability for students to classify other students correctly ρ. When students have difficulty to grade others ($\rho = 0.25$) the convergence is slow in the first iterations due to the fact that many students with the same role where placed together in the initial solutions, thus decreasing ρ. It is not until 15 iterations that the policy solution has improved approximately a 10% with respect to the initial solution. Nevertheless, the policy still converges towards the optimal solution even when students have an average difficulty to grade others $\rho = 0.5$. In just 5 iterations, it has been able to enhance the solution a 15% with respect to the initial solution. This result also suggests that the policy is applicable in a classroom as long as it is not highly difficult for students to evaluate others.

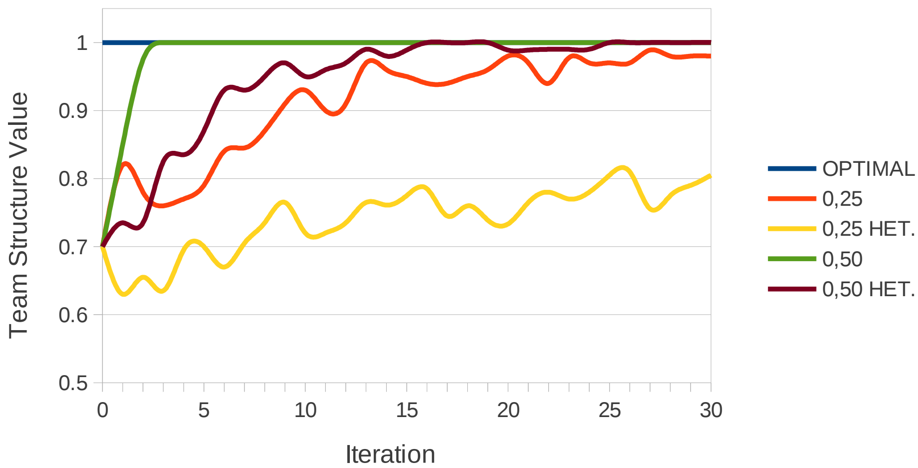

Fig. 3: Average team structure value found by the proposed policy when the role distribution among students is non-uniform.

5 Related Work

When there is a large number of students and different grouping criteria, the task of forming collaborative learning teams to promote successful outputs is considered a NP-hard problem. Over the past years, several approaches have been proposed in order to deal with this goal.

The majority of the proposals try to create heterogeneous teams since there is a direct relationship between the performance of a team and the balance level among the roles. Christodoulopoulos et al. [3] present a web-based group formation tool that facilitates the creation of homogeneous and heterogeneous groups based on three criteria.

	initial info.	algorithm	feedback	clusters	population
[3]	3 attributes	Fuzzy C-mean random selection	negotiation	het./hom	18
[5]	personal traits	ant colony optimization	-	het.	500
[19]	*thinking styles* (questionaries)	genetic alg.	final students satisfaction	het	66
[20]	personal knowledge social network	genetic alg.	-	het.	45
[11]	understanding level interest of students	particle swarm optimization	-	hom	15-2000
[21]	self-evaluation of roles	crowding evolutionary alg.	-	het.	18-3000

Table 1: Comparison of approaches that deal with the problem of forming collaborative learning teams. The features considered are: the initial information available about the students, the algorithm used, the inclusion of feedback in the coalition formation process, population in each cluster (heterogeneous or homogeneous), and number of students used in the experiments.

This tools allows the instructor to manually modify the groups and allows the students to negotiate the grouping. The creation of homogeneous groups is based on a Fuzzy C-Means algorithm, and the creation of heterogeneous groups is based on a random selection algorithm. The tool also provides an option to negotiate the teams proposed with the students. However, this negotiation consists on a direct interaction with the teacher. Other approaches use bio-inspired algorithms. Graf et al. [5] present an Ant Colony Optimization approach that provides heterogeneous groups based on personal traits of students. The groups are formed by four students with low, average, and high students scores. The algorithm tries to maximize the diversity of the group while keeping a similar degree of heterogeneity of all the groups.

Genetic or evolutionary algorithms are also commonly used to solve the NP-hard problem of forming collaborative learning teams. Wang et al. [19] present an approach for automatic team formation based on *thinking styles* [6] that determines the features of the students. They consider an heterogeneous group formation. The algorithm translates the features of the students into points in a two-dimensional space and then, they are classified into categories. The algorithm uses a genetic algorithm to create the optimal group formation based on the categories of the students. The experiments consider 66 students and groups of 3 people. At the end of the process, the coalitions are evaluated by the students through a questionary. However, this evaluation is not used as feedback to improve the group formation. Lin et al. [11] present a system to assist instructors to form collaborative learning groups that provide outcomes to all their members. The algorithm considers two criteria: information about understanding levels and interests of the students. Particle swarm optimization is used in the group composition algorithm to deal with the complexity of the problem. The main drawback of this proposal is that the groups consist of homogeneous students. Yannibelli and Amandi [21] propose a crowding evolutionary algorithm to deal with the complexity of the problem of creating collaborative learning teams. The algorithm balances the roles of the members and the number of members of each group. Belbin's roles [2] are considered in the experiments and, in order to assign a role to each student they initially use the Team Role Self-Perception Inventory.

The problem of team formation is also present in the context of human resource management. Wi et al [20] present a framework to deal with the team formation in R&D-oriented institutes. The authors propose a genetic algorithm that uses a fuzzy model to take into consideration information about the candidates related to their knowledge and expertise about certain topics related to certain project. Moreover, the algorithm considers information about the position of the candidates in a social network in order to see their suitability for project management positions.

Our proposal for collaborative learning teams formations improves previous approaches in several ways. First, our proposal does not have previous information about the abilities, attributes, or roles played by the students. The only previous information is the set of roles that could appear in a team. Using these roles, team members can provide an estimation about the roles played by other team members. Second, our proposal provides a more reliable role assignment since it considers the opinion of other members instead of a personal evaluation. Finally, in each iteration of the algorithm, the solution is improved with the feedback received from direct interactions among students in each team.

6 Conclusions

In this paper we have presented a computer-aided policy for generating teams of students a classroom. The policy is based on Belbin's role taxonomy [2], collective intelligence, coalition structure generation algorithms, and Bayesian learning. After the execution of a class group class activity, students classify other teammates according to Belbin's role taxonomy. Then, the information regarding the predominant role of each student is updated via Bayesian learning. This information is then used by the coalition structure generation algorithm to calculate the next team structure. The simulations have shown that, as long as students do not have great difficulties classifying others, the policy is capable of improving the quality of team structures in a few iterations and gradually converging towards the optimal solution.

We simulated different scenarios in order to test different environmental conditions. The results are encouraging enough to continue this research. As a future work, we plan to extend the experiments in order to consider large populations of students and environmental conditions, such as scenarios where some roles are more important than others. In addition, we also intend to study whether or not the inclusion of more attributes in the classification problem can improve the performance of the policy.

Acknowledgements This work is supported by TIN2011-27652-C03-01 and TIN2012-36586-C03-01 of the Spanish government and FPU AP2008-00601 granted to Elena del Val.

References

1. J. M. Alberola, V. Julian, and A. Garcia-Fornes. Multidimensional Adaptation in MAS Organizations. *IEEE Trans. Syst. Man Cybern. Part B-Cybern.*, pages 1–12, 2012.
2. R. M. Belbin. *Team roles at work*. Routledge, 2010.
3. C. E. Christodoulopoulos and K. A. Papanikolaou. A group formation tool in an e-learning context. *2012 IEEE 24th International Conference on Tools with Artificial Intelligence*, 2:117–123, 2007.
4. M. F. De Vries. High-performance teams: Lessons from the pygmies. *Organizational Dynamics*, 27(3):66–77, 2000.
5. S. Graf and R. Bekele. Forming heterogeneous groups for intelligent collaborative learning systems with ant colony optimization. In *Proceedings of the 8th international conference on Intelligent Tutoring Systems*, pages 217–226. Springer, 2006.
6. E. Grigorenko and R. Sternberg. Styles of thinking, abilities, and academic performance. *Exceptional children*, 63(3):295–312, 1997.
7. R. S. Hansen. Benefits and problems with student teams: Suggestions for improving team projects. *Journal of Education for Business*, 82(1):11–19, 2006.
8. M. Higgs, U. Pewina, and J. Ploch. Influence of team composition and task complexity on team performance. *Team Performance Management*, 11(7/8):227–250, 2005.
9. D. W. Johnson and R. T. Johnson. An educational psychology success story: Social interdependence theory and cooperative learning. *Educational researcher*, 38(5):365–379, 2009.
10. D. W. Johnson, R. T. Johnson, and K. Smith. The state of cooperative learning in postsecondary and professional settings. *Educational Psychology Review*, 19(1):15–29, 2007.
11. Y.-T. Lin, Y.-M. Huang, and S.-C. Cheng. An automatic group composition system for composing collaborative learning groups using enhanced particle swarm optimization. *Computers & Education*, 55(4):1483 – 1493, 2010.
12. F. Maffioli and G. Augusti. Tuning engineering education into the european higher education orchestra. *European Journal of Engineering Education*, 28(3):251–273, 2003.
13. F. P. Morgeson, M. H. Reider, and M. A. Campion. Selecting individuals in team settings: The importance of social skills, personality characteristics, and teamwork knowledge. *Personnel Psychology*, 58(3):583–611, 2005.
14. N. Ohta, V. Conitzer, R. Ichimura, Y. Sakurai, A. Iwasaki, and M. Yokoo. Coalition structure generation utilizing compact characteristic function representations. In *Principles and Practice of Constraint Programming - CP 2009*, volume 5732, pages 623–638. 2009.
15. S. Pajares, V. Sanchez-Anguix, A. Torreño, and S. Esparcia. A novel teaching-learning strategy for teamwork based on agreement technologies. *IJCA Proceedings on Design and Evaluation of Digital Content for Education (DEDCE)*, (1):21–30, 2011.
16. V. Ratcheva and S. Vyakarnam. Exploring team formation processes in virtual partnerships. *Integrated Manufacturing Systems*, 12:512–523, 2001.
17. S. J. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, 2010.
18. P. Tarricone and J. Luca. Employees, teamwork and social interdependence–a formula for successful business? *Team Performance Management*, 8(3/4):54–59, 2002.
19. D.-Y. Wang, S. S. Lin, and C.-T. Sun. Diana: A computer-supported heterogeneous grouping system for teachers to conduct successful small learning groups. *Computers in Human Behavior*, 23(4):1997 – 2010, 2007.
20. H. Wi, S. Oh, J. Mun, and M. Jung. A team formation model based on knowledge and collaboration. *Expert Systems with Applications*, 36(5):9121 – 9134, 2009.
21. V. Yannibelli and A. Amandi. A deterministic crowding evolutionary algorithm to form learning teams in a collaborative learning context. *Expert Systems with Applications*, 39(10):8584 – 8592, 2012.