Towards Fair and Efficient Assignments of Students to Projects

Marco Chiarandini · Rolf Fagerberg · Stefano Gualandi

We consider the following problem in project assignment. We are given a set P of *project topics* and a set S of *students*. For each topic $i \in P$ a limited number t_i of teams can be created. If a team for topic i is created, the number of students assigned to it must be between a minimum and maximum bound, l_i and u_i, respectively. Students express preferences for the topics by ranking a subset of project topics. For each student $s \in S$ this ranking is given by an ordered set $r(s) = (p^{(1)}, \ldots, p^{(q_s)})$, $p^{(i)} \in P$, $1 \leq i \leq q_s \leq n$. Moreover students can register in *groups* of at most ℓ persons, if they want to be assigned to the same team. That is, students are partitioned in groups, $G = \{g_1, \ldots, g_{m'}\}$, $\bigcup_i g_i = S$, $g_i \cap g_j = \emptyset, \forall i \neq j$ and $|g_i| \in \{1, 2, \ldots, \ell\}$. Students in the same group g_i have the same preference set, that is, $r(s_1) = r(s_2)$ for all $s_1, s_2 \in g_i$.

We wish to find an assignment of students to project teams, $\sigma : S \rightarrow \{1, 2, \ldots, \sum_{i \in P} t_i\}$ such that students are assigned to exactly one project from their preference set and team bounds and group requirements are satisfied. If no assignment satisfying these constraints exists, we will modify the input data and iterate. If more than one assignment can be found, we wish to choose one among them according to the criteria of fairness and collective welfare.

Preference sets can be transformed into a *score* matrix $V \in \mathbb{N}_0^{|S| \times |\mathcal{P}|}$, where each element v_{si} represents how student s ranks project i. The score 1 is set for the topics that are ranked first and q_s for the topics that are ranked in the q_s-th position. If a project topic i is not in the preference list of the student s, the corresponding value v_{si} will never be used and it can be set to any number, for example, zero.

M. Chiarandini · Rolf Fagerberg
Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark, E-mail: {marco|rolf}@imada.sdu.dk

S. Gualandi
Department of Mathematics, Università di Pavia, via Ferrata, 1 I-27100, Pavia, Italy, E-mail: stefano.gualandi@unipv.it

The quality of an assignment σ that satisfies all constraints is determined by a vector $\mathbf{v} = (v_{1,\sigma(1)}, \ldots, v_{m,\sigma(m)})$, $v_{s,\sigma(s)} > 0, \forall s \in S$, and by the distribution of students over ranks $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_\Delta)$, where $\Delta = \max\{q_s \mid s \in S\}$ and $\delta_{i=1..\Delta} = |\{s \in S \mid v_{s,\sigma(s)} = i\}|$. Students will prefer assignments over others on the basis of their *individual utility*, that is, their score in the vector $\mathbf{v}$. In the decision-making process of an ad hoc committee that has to solve the allocation problem under limited resources, the focus will be on the *collective welfare*. From this viewpoint, the interest is on assignments that are Pareto optimal or *efficient* with respect to their profile vectors $\mathbf{v}$. The two most prominent ways to aggregate a profile of preference relations into a collective preference relation are the *classical utilitarian ordering* and the *egalitarian ordering* [4]. For two feasible assignments σ^1 and σ^2, the former assigns a weight to each value, $w : \{1, 2, \ldots, \Delta\} \to \mathbb{Z}^+$ and compares $\sum_s w(v_{s,\sigma^1(s)})$ with $\sum_s w(v_{s,\sigma^2(s)})$, and the latter uses the *leximin order*, which consists in reordering the two vectors $\mathbf{v}^1$ and $\mathbf{v}^2$ by increasing coordinates and comparing them lexicographically. Both relations define a strict weak order.

Taking only efficiency into account it is possible to create examples where the overall satisfaction is high to the disadvantage of a few students. The individual welfare or *fairness* criterion aims at ensuring that no student is disadvantaged to the benefit of others. A way to achieve this is by searching for the assignment that minimizes the worst rank, that is, $\min\max\{v_{s,\sigma(s)}|s \in S\}$. This is also known as the *minimax* criterion [5].

On the other hand, the minimax criterion alone makes no use of additional information to decide among assignments with the same guarantees on the maximum scores in the vector $\mathbf{v}$. For example, the two vectors $(1, 3, 3, 3)$ and $(1, 2, 2, 3)$ are not distinguishable. An approach that takes both fairness and collective welfare into account consists in first optimizing according to the minimax criterion and then, restricted to only minimax optimal solutions, optimizing collective welfare using the weighted value order. An alternative approach that overcomes the drawback of the minimax criterion and conciliates egalitarianism and Pareto-efficiency is the *leximin* order, which subsumes the minimax criterion.

Without common registrations and no minimum number of students per project team, the problem under a classical utilitarian approach can be formulated as a particular case of minimum cost flow. With all constraints the problem is instead strongly NP-hard as it is a special case of the multiple knapsack problem or the generalized assignment problem that are optimization versions of the strongly NP-complete 3-partition problem [1].

In [3] the authors used a genetic algorithm approach to solve a similar problem but without project topics, lower bounds on team sizes, and group registrations. The issue of fairness is addressed by means of the multiple solutions returned by the genetic algorithm when it solves a formulation with weighted sum. Our work is an adaptation of the study by Garg et al. [2] in the context of conference management for assigning papers to referees. The problem there treated is similar to ours, but with the further issue that referees, contrary to students, can receive more than one paper and a load balancing criterion has to be included. Garg et al. show that the leximin approach under lexicographic and weighted score orders are both NP-hard for $\Delta \geq 3$. Then they give an approximation algorithm for the general case.

We have designed a randomized greedy algorithm that implements a lottery approach. The procedure is appealing from the point of view of fairness. Then we studied different formulations of the problem in integer linear programming (ILP) and constraint programming (CP). We compared these methods on real life instances with up to 300 students and 102 project teams in 80 different topics. Results showed that ILP models based on a *distribution approach* to handle the lexicographic order solve the problem in a matter of seconds and the assignments found outperform those of the lottery approach in terms of both feasibility and quality. The CP models studied so far are instead not yet competitive.

We use our solutions in practice at the Faculty of Science of the University of Southern Denmark, where in the first year of their education students must undertake a group project of the duration of one quarter. Students are left free to rank project topics independently from the specific branch of science in which they will later specialize. In the past we used the lottery algorithm and had to interact with the students because initially no feasible solution was found. Since 2011, we use the ILP model based on lexicographic optimization and assignments that satisfy students and the administrating committee are found immediately even when the lottery algorithm would not find any.

References

1. Garey, M.R., Johnson, D.S.: Computers and Intractability: A Guide to the Theory of $\mathcal{NP}$-Completeness. Freeman, San Francisco, CA, USA (1979)
2. Garg, N., Kavitha, T., Kumar, A., Mehlhorn, K., Mestre, J.: Assigning papers to referees. Algorithmica **58**(1), 119–136 (2010). DOI 10.1007/s00453-009-9386-0
3. Harper, P.R., de Senna, V., Vieira, I.T., Shahani, A.K.: A genetic algorithm for the project assignment problem. Computers & Operations Research **32**(5), 1255 – 1265 (2005). DOI 10.1016/j.cor.2003.11.003
4. Moulin, H.: Fair Division and Collective Welfare. The MIT Press (2003)
5. Yager, R.R.: Constrained OWA aggregation. Fuzzy Sets and Systems **81**(1), 89 – 101 (1996). DOI 10.1016/0165-0114(95)00242-1