Integer Linear Optimization for Solving the Team Formation Problem: Multi-Team Mentoring Optimization

Gabriel Vinícius dos Santos and
Maristela Oliveira dos Santos
Institute of Mathematical and Computer Sciences - ICMC
University of São Paulo - USP - Brazil
gabrielvs.mdo@gmail.com mari@icmc.usp.br

Objectives

The field of optimization has been extensively studied due to its wide range of applications, including product storage and workforce scheduling. The main focus of this work is the development of a heuristic applied to the Teamwork with Mentoring Problem (TPM), which was proposed in the 2022 Google Hash Code competition [1]. The TPM involves the formation of teams of contributors for projects, taking into consideration the skills required for each project and the skills of the contributors. The objective is to select a team that maximizes the total project scores. There are strict constraints, such as limiting a contributor to one project at a time and the availability of contributors after project completion. Additionally, professionals can assist each other with mentoring, and skills can improve if they do not meet the minimum requirements. Multiple projects can be in progress simultaneously, with contributors occupying at most one role in each project. Projects have scores and deadlines, with the score decreasing by one point per day after the deadline. The planning horizon is defined by the date when all projects have a score of zero. Flexible constraints are introduced to maximize the final score. The solution approach to the problem involves the use of heuristics to find initial solutions. Subsequently, an algorithm is employed to improve the solutions, with the main focus of the research being a deep understanding of TPM and its resolution, as well as generating good initial solutions and comparing them to the state of the art. This multidisciplinary approach underscores the importance of optimization in addressing complex real-world challenges. This way, the objective of this work is to create an algorithm that effectively solves the problem, streamlining processes and maximizing score.

Methods and Procedures

For the TMP, it was possible to create an initial solution algorithm, adhering to the strict constraints. The algorithm proposes an initial solution with various heuristics since studying and working with different approaches provide a better understanding of the problem. Project sorting heuristics include: Shortest Processing Time (**SPT**), Longest Processing Time (**LPT**), Earliest Due Date (**EDD**), Minimum Slack Time (**MST**), Critical Ratio (**CR**), Slack (**SLK**) [2], and Manpower (**MPW**), which takes into account the relationship between the deadline and the processing time of each project. Meanwhile, the contributor selection heuristics used include: Nearest Skill Value (**Nst**), Farthest Skill Value (**Fst**), and First to Find (**Ftf**), which prioritizes selecting the first contributor where a skill level increase is observed. The greedy algorithm employed in problem resolution utilizes input data from instances [1] and yields the best combination of heuristics. Initially, the algorithm was applied to test cases with fewer projects and contributors, allowing for a comprehensive study of the problem. After studying the problem and identifying the most effective heuristic combinations, these were applied to more complex test cases and compared against the best-known scores up to that point. Additionally, a local perturbation algorithm, inspired by other articles [3], was developed and tested on simpler test cases.

Results

For the results presented in this project, Table 1 displays the best results obtained for the six available instances [1]. The ”total”row shows the highest accumulated value following a single algorithm, while each individual instance in the state of the art has its value obtained by different algorithms.

Tabela 1: Instances [1], State of the Art Solutions [1], and Solutions Found in This Work.

Inst.	State of the Art	Solution - Project
a	33	33
b	1.005.020	855.630
c	288.508	199.049
d	674.945	173.626
e	1.650.488	1.610.072
f	1.194.515	543.471
total	4.220.236	3.381.881

Table 2 presents, for each instance, a result comparison metric called GAP, which indicates how far the solution is from the state of the art. This table also shows which heuristics were used to obtain the found solution. The GAP can be defined as follows:

$$\text{GAP} = \frac{\text{Best Solution} - \text{My Solution}}{\text{Best Solution}} 100\% \quad (1)$$

Tabela 2: Instances [1], GAP of Solutions, Heuristics Used in the Best Solutions for Each Instance.

Inst.	GAP	Heuríst. 1	Heuríst. 2
a	0%	SPT	Ftf
b	14,86%	MST	Ftf
c	31,00%	LPT	Ftf
d	74,28%	SPT	Ftf
e	24,49%	CR	Ftf
f	54,50%	SLK	Ftf
Total	19,87%	-	Ftf

Finally, Table 3 presents the results of the local improvement algorithm for the related instances. Additionally, we have GAP' as a metric to evaluate how much the solution was improved. GAP' can be defined as follows:

$$\text{GAP'} = \frac{\text{New Solution} - \text{First Solution}}{\text{New Solution}} 100\% \quad (2)$$

Tabela 3: Instances [1], First best solution found, Second best Solution found after improvement algorithm, Improvement of Solutions (GAP')

Test	1st Sol.	2nd Sol.	GAP'
b - SPT	800981	800981	0%
b - LPT	800975	800975	0%
b - EDD	800918	769399	3,94%
b - MST	800991	769472	3,94%
b - CR	800991	800991	0%
b - SLK	800930	769411	3,94%
b - MPW	800989	800989	0%

Conclusions

In this study, the results indicated that project sorting heuristics had no significant impact on the outcomes, while the "First to Find (**Ftf**)"heuristic for contributor selection played a crucial role in obtaining good initial solutions. The analysis also revealed that the characteristics of TMP instances, such as the number of contributors, projects, and skills, influenced the algorithm's performance. Only one instance ("d") showed a considerable deviation from the best-known result, as also observed in similar studies [3]. It became evident that the swapping algorithm was not as effective as the initial solution approach. A suggested future proposal is to develop a perturbation algorithm for the initial solution, inspired by literature works [3], with the aim of surpassing the State of the Art. Thus, this study brought a new approach to TMP, focusing on the construction of an effective initial solution. The source code of the utilized algorithm can be found at: [https://github.com/GVS2001/TPM_Teamwork-with-Mentoring-Problem].

References

[1] GOOGLE. Mentorship and teamwork - data, February 2022.

[2] Tatiane Carvalho Silva and Helio Yochihiro Fuchigami. Desenvolvimento e análise de regras de sequenciamento da produção para minimização de medidas de atraso em flow shop com tempos de setup. *Revista Produção Online*, 18(2):454–478, jun. 2018.

[3] Uran Lajçi and Elvir Misini. Exploring local search metaheuristics for optimizing the mentorship and teamwork problem. 05 2023.