4OR
https://doi.org/10.1007/s10288-020-00466-7

Sports timetabling: theoretical results and new insights in algorithm performance

David Van Bulck

This preprint has not undergone peer review or any post-submission improvements or corrections. The Version of Record of this article is published in 4OR, and is available online at `https://doi.org/10.1007/s10288-020-00466-7`.

This is a summary of the author's Ph.D. thesis supervised by Dries Goossens and defended on 3 September 2020 at Ghent University. The thesis is written in English and is available from the author upon request at david.vanbulck@ugent.be and from `https://doi.org/10.13140/RG.2.2.13928.49929`. This work derives theoretical results and insights in algorithm performance with regard to the construction of practical and fair sports timetables which define who plays whom, when, and where.

Sports timetabling is a relatively young field, with the first pioneering papers that appeared in the 1970s and 1980s. In the past 50 years, operations research has brought great advancements to sports timetabling, but the lack of a unified sports timetabling problem formulation also made that the description of constraints and objectives is currently highly inconsistent and that problem instances are rarely shared. *A first contribution* of this thesis is therefore to develop a three-field notation to describe a sports timetabling problem by means of the tournament format, the constraints in use, and the objective. We use this notation to classify most of the problems presented in the literature, propose XML-based file templates, and maintain an online data repository with real-life problem instances and their best known solutions from over fifteen different countries and eight different sports (see `www.sportscheduling.ugent.be/RobinX/`).

Decomposition methods are common in sports timetabling and break down the timetabling process into different sub-problems. In the first-break-then-schedule method, the first sub-problem is to determine the home-away pattern (HAP) set which defines when teams must play at home, play away, or have an off day. Alternatively, in the first-off-day-then-schedule method, the first sub-problem is to determine the game-off-day pattern (GOP) set, which regulates when teams must play (home or away) or have an off day. Since the efficiency of decomposition methods heavily depends on the ability to avoid infeasible sub-problems early on, an important question is whether a tournament can be scheduled given an HAP or GOP set. The decision problem of verifying whether a given HAP or GOP set is feasible, is respectively known as the HAP and GOP set feasibility problem. Bao proposes necessary conditions for the HAP and GOP set feasibility

D. Van Bulck (david.vanbulck@ugent.be)
Ghent University, Faculty of Economics and Business Administration

and asks whether these conditions are sufficient when the timetable is time-relaxed, i.e. there are more time slots than strictly needed (see R. Bao, 'Time relaxed round robin tournament and the NBA scheduling problem', PhD thesis, Cleveland State University, 2009.). As a *second contribution*, this thesis answers this question negatively, analyses and proposes necessary conditions, and shows that the time-relaxed HAP and GOP set feasibility problems are $\mathcal{NP}$-complete. Nevertheless, we provide a polynomial time algorithm for the GOP set feasibility problem when the number of off days per team is less than three, and we conjecture the problem to be polynomially solvable if the number of off days per team is less than four. Whereas the HAP set defines the time slots on which teams *must* play home or away, venue and player availability constraints regulate when teams *can* play home or away. This thesis shows that time-relaxed timetabling remains difficult as soon as player or venue availability is considered, and therefore settles the complexity status for a considerable number of problems studied in the literature that involve availability constraints. Finally, we employ list-edge colouring techniques to derive sufficient conditions on the availability of teams and venues for a feasible timetable to exist.

Despite their flexibility to take into account availability constraints, time-relaxed timetables have the drawback that the rest period between teams' consecutive games can vary considerably, and the difference in the number of games played at any point in the season can become large. For fairness reasons, organizers may also want to timetable home and away games alternately. As a *third contribution*, this thesis proposes exact mathematical models and three types of heuristics that can handle these fairness issues. First, we propose relax-and-fix based heuristics using team- and time-based variable groupings, and find that team-based variable groupings work better than time-based groupings. Second, we propose an adaptive large neighbourhood method that repeatedly destroys and repairs a timetable. Third, we propose a memetic algorithm backed by a local search component that solves a transportation problem to schedule or reschedule all home games of a team. We use the heuristics to generate timetables for a real-life non-professional indoor football competition, and look at the trade-off between a timetable that is as fair as possible for the league overall, versus a timetable that equitably splits its unfair aspects over the teams.

Any sports competition needs a timetable but not all timetabling algorithms are equally suitable to construct this timetable: depending on the features of the competition, an algorithm may perform well for one type of competition but not for another. Given the performance differences of algorithms, organizers face the important decision what algorithm to select to generate timetables. This decision is complicated by the fact that an algorithm needs to be chosen long before the specifics of the competition to be planned are known. As a *fourth contribution*, we propose an algorithm recommendation system which predicts what algorithm is likely to perform best when given the features of the competition the organizers face. In particular, we show how to project problem instances in a two-dimensional feature space, use integer programming to generate new problem instances in the part of the feature space where insufficient data is available, and use machine learning techniques to predict the best performing algorithm based on the coordinates in the two-dimensional space. We demonstrate the effectiveness of the proposed techniques by generating algorithm predictions for time-relaxed sports timetabling with availability constraints and rest time optimisation.

This preprint has not undergone peer review or any post-submission improvements or corrections. The Version of Record of this article is published in 4OR, and is available online at `https://doi.org/10.1007/s10288-020-00466-7`.