Indoor football scheduling

Dries Goossens · Frits Spieksma

Abstract This paper deals with a real-life scheduling problem from an amateur indoor football league. The league consists of a number of divisions, in each of which a double round robin tournament is played. The goal is to develop a schedule which avoids close successions of matches involving the same team. This scheduling problem is interesting, because matches are not planned in rounds. Instead, each team has a number of time slots available to play its home games. Furthermore, in contrast to professional leagues, alternating home and away matches is hardly relevant. We present a mathematical programming formulation and a heuristic based on tabu search, which resulted in high-quality schedules that have been adopted in practice.

Keywords scheduling · non-professional · indoor football · time-relaxed · tabu search

1 Introduction

In this paper, we discuss the sports scheduling problem faced by the “Liefhebbers Zaalvoetbal Cup (LZV Cup)”, an amateur indoor football league founded in 2002. This league currently involves 87 teams, all situated in the vicinity of Leuven (Belgium), grouped in 6 divisions. The LZV Cup focusses on teams that
consist of friends, is open to all ages, and considers fair play of paramount importance. The matches are played without referees, since, according to the organizers, “referees are expensive, make mistakes, and invite players to explore the borders of sportsmanship” (see also http://www.lzvcup.be [in Dutch]).

Academic interest for scheduling amateur leagues is rather limited, compared to the attention that professional leagues receive. Perhaps this can be explained by the fact that a lot less money is involved with non-professional leagues. In addition, scheduling constraints coming from broadcasting rights, security forces, public transport, media and fans usually are non-existing. This does not imply however that amateur scheduling problems are less challenging than their professional counterparts. Indeed, stadium or ground availability is in general more limited, because the team’s venue tends to be shared with other teams or sports disciplines. Moreover, practical considerations for the players are far more important, since they have other activities (e.g. family, work) as well.

In our indoor football scheduling problem, there are multiple divisions. In each division, each team plays against each other team twice. The teams provide dates on which they can play a home game, and dates on which they cannot play at all. The league organizers are not worried by (long) series of consecutive home games (or away games), but do not want a close succession of two matches featuring the same teams. The goal is to develop a schedule for each of the divisions, where each team has a balanced spread of their matches over the season.

In section 2, we give a formal description of the problem, followed by an overview of the literature on related problems in section 3. We provide a mathematical formulation in section 4, which is used to tackle the problem with Ilog Cplex. In section 5, we develop a heuristic approach, based on tabu search. Computational experiments with both methods for real-life problem instances we solved for the LZV Cup are described in section 6. The paper ends with conclusions and future work in section 7.

2 Problem description

In this section, we provide a formal problem description, and we introduce the notation used in the remainder of this paper. The teams in the LZV Cup are grouped according to their strength into a division. In each division, a double round robin tournament is played, i.e. each team meets each other team twice (once at its home venue, and once at the opponent’s venue). Apart from the number of teams, the problem description is identical for all divisions. A division has a set of teams $T$, with $|T| = n$, and a set of time slots $S = \{1, 2, ..., |S|\}$, ranging from the first day of the season till the last. All
matches should be played within this time frame.

Each team $i \in T$ provides a list of time slots $H_i \subseteq S$ for which their home ground is available. Home games for a team can only be scheduled on time slots from this list. Obviously, if each match is to be scheduled, each team should at least provide as many time slots as it has opponents, i.e. $|H_i| \geq n - 1$. This list is called the *home game set*. Some teams may have a slot on the same weekday every other week; other teams may have a more irregular home game set. Each team can also provide a list of calender days $A_i \subseteq S$ on which it doesn’t want to play a match; we call this list the *forbidden game set*. Teams can use this list to avoid matches in the Christmas and New Year period, on holidays, during an exam period, etc. The forbidden game set implies that on all days not in the list, the team is able to play an away game. We assume that $H_i \cap A_i = \emptyset$. A team is not allowed to play twice on the same day, or more than twice in a period of $R_{\text{max}}$ days. Finally, there should also be at least $m$ calendar days between two matches featuring the same teams. Notice that it is allowed to meet an opponent for the second time, before all other opponents have been faced once.

In summary, we have the following constraints:

- each team plays a home match against each other team exactly once [C1]
- home team availabilities $H_i$ are respected [C2]
- away team unavailabilities $A_i$ are respected [C3]
- at least $m$ days between two matches with the same teams [C4]
- each team plays at most one game per day [C5]
- each team plays at most 2 games in a period of $R_{\text{max}}$ days [C6]

The goal is to develop a schedule with for each team a balanced spread of their matches over the season. More in particular, teams wish to avoid having two matches in a period of $R_{\text{max}}$ days or less. We use $p_r$ to denote the penalty incurred for every pair of consecutive matches played by a team within period of $r \in R = \{2, 3, \ldots, R_{\text{max}}\}$ days. Obviously, having 2 matches in 2 days is considered more unpleasant than having 2 matches in 4 days. The main idea behind this, is that most players prefer not to fully spend their weekend with their sport. Moreover, matches packed together could also lead to injuries. If a team has more than $R_{\text{max}}$ days between two consecutive matches, we assume that the league organizers no longer care, and consider any number greater than $R_{\text{max}}$ as equally adequate. Constraint C1 is in fact interpreted as a soft constraint, i.e. it is possible not to schedule a match, but only at a high cost. This guarantees feasibility of any instance (e.g. in case some team does not provide enough time slots for which their home ground is available). In practice, if a match cannot be scheduled, the league organizers leave it to the home team to find a suitable date and location to play the match (if the home team fails to organize the match, they lose the game).
3 Related work

A large number of sports scheduling papers deal with professional leagues (see e.g. [2], [4], and [6]; for a complete overview we refer to [9]). Academic interest in problems faced by amateur leagues is far more rare. In general, there are two main differences between professional and amateur leagues. Firstly, in amateur leagues matches are typically not planned in rounds. Scheduling problems in sports leagues can be divided into two types: temporally constrained and temporally relaxed problems. In the first type, the matches are grouped in rounds, and no more than the minimum number of rounds required to schedule all the matches is available. For leagues with an even number of teams, this means that each team plays exactly once in each round. Nowadays, all European professional football schedules are temporally constrained [7]. With temporally relaxed problems, the number of rounds is larger than the minimum number of rounds needed. In this case, some teams will have one or more rounds without a match. In the cricket world cup as reported in [1], 9 teams play a single round robin tournament, resulting in 36 games that need to be scheduled in a 26-day period. In this tournament, several constraints and practical considerations need to be respected, resulting in a schedule that is suitable for a worldwide TV audience. The Australian state cricket season also provides a temporally relaxed scheduling problem, where one of the constraints is that matches between the 6 states must be scheduled around predetermined international and test match fixture dates [15]. Scheduling the National Hockey League, as discussed in [3], involves planning 82 games per team in a period of 28 weeks. In addition to other constraints, no team should not play games on three days straight, nor should it play more than three games in five consecutive days. In our problem, the slots themselves can be seen as rounds, making the schedule extremely temporally relaxed.

Secondly, successions of home or away matches (breaks) are hardly relevant in amateur leagues. The main reason is that there are usually few spectators, and the home advantage is quite limited. Only extremely long series of consecutive home or away matches could be problematic. In our problem, however, they are unlikely, because a team’s home game set is usually well balanced over the season, and the number of days where its home venue is available is not much larger than the number of home games. In professional leagues, however, alternation of home and away matches is usually the most important constraint. This is illustrated by the popularity of the so-called “first-break, then schedule” approach (see e.g. [11]), and the attention for the break minimization problem (see e.g. [12]). A break in this alternating sequence is not desirable for the spectators, and less spectators tend to show up for the second or third home match in a row [5].

The concept of a list of dates on which no game should be scheduled for some team has been introduced in [13]; they call it suspension dates. The authors discuss a scheduling problem from a regional amateur table-tennis
federation in Germany, involving more than 30 leagues. This scheduling problem is quite similar to ours. In each league a double round robin tournament is played; the schedule is temporally relaxed. All games have to be scheduled within a given period, while each team may be involved in at most one game per day. Home teams provide a list of permitted dates, and away teams can also specify a number of dates on which they are not available. The teams also specify the number of days-off they want between two successive games. Unlike in our problem, the season is split in two halves, such that each team meets each other team once in each half. Furthermore, to be able to make a meaningful ranking, the season is subdivided into six time periods of equal length, and the number of matches that each team has to play in each period is constrained by a lower and an upper bound. The authors solve this scheduling problem with a permutation based genetic algorithm for which feasibility preserving operators are defined. In a follow-up paper [14], the authors propose a memetic algorithm, backed by a constraint propagation based heuristic and use a co-evolutionary approach.

Knust [10] also starts from the table-tennis scheduling problem discussed in [13], but adds a number of constraints (e.g. some matches should be played on weekend days (instead of weekdays), and some matches should be scheduled in specific time intervals). More importantly, for each team home and away matches should be scheduled alternately (i.e. breaks should be avoided). Knust [10] models the problem as a multi-mode resource-constrained project scheduling problem, for which a 2-stage heuristic solution algorithm is proposed, involving local search and genetic algorithms.

4 A mathematical formulation

In this section, we write the scheduling problem more formally as an integer problem. Our main decision variable is $x_{ij,s}$, which is 1 if team $i$ plays a home game against team $j \neq i$ on slot $s$, and 0 otherwise. The variable $y_{ist}$ is 1 if team $i$ plays a match on slot $s$, followed by its next match on slot $t$, for each $s$ and $t$ such that $t > s$ and $t - s + 1 \leq R_{max}$, and 0 otherwise. The variable $u_{ij}$ is 1 if no home game of team $i$ against team $j$ is scheduled, and 0 otherwise. Each unscheduled match results in a penalty $P$. We can now write the following formulation for our problem:

$$
\text{minimize} \sum_{i \in T} \sum_{j \in T: j \neq i} Pu_{ij} + \sum_{i \in T} \sum_{s \in S} \sum_{t \in S} p^{t-s+1} y_{ist}
$$
subject to

\[
\sum_{s' \in H_i \setminus A_j} x_{ijs} + u_{ij} = 1 \quad \forall i, j \in T : i \neq j \quad (1)
\]

\[
\sum_{j \in T} (x_{ijs} + x_{jis}) \leq 1 \quad \forall i \in T, s \in S \quad (2)
\]

\[
y_{ist} \leq \sum_{j \in T} (x_{ijs} + x_{jis}) \quad \forall i \in T, s, t \in S \quad (3)
\]

\[
y_{ist} \leq \sum_{j \in T} (x_{ijt} + x_{jlt}) \quad \forall i \in T, s, t \in S \quad (4)
\]

\[
\sum_{j \in T} (x_{ijs} + x_{jis}) + \sum_{j \in T} (x_{ijt} + x_{jlt}) - \sum_{j \in T} \sum_{k = s + 1}^{s + R_{max} - 1} (x_{ijk} + x_{jik}) - 1 \leq y_{ist} \quad \forall i \in T, s, t \in S \quad (5)
\]

\[
x_{ijs} + x_{jis'} \leq 1 \quad \forall i, j \in T, s \in H_i, s' \in H_j : |s - s'| < m \quad (6)
\]

\[
\sum_{j \in T} \sum_{k = s}^{s + R_{max} - 1} (x_{ijk} + x_{jik}) \leq 2 \quad \forall i \in T, s \in S \quad (7)
\]

\[
x_{ijs} = 0 \quad \forall i, j \in T, s \notin H_i \cup s \in A_j \quad (8)
\]

\[
x_{ijs} \in \{0, 1\} \quad \forall i, j \in T, s \in S \quad (9)
\]

\[
y_{ist} \in \{0, 1\} \quad \forall i \in T, s, t \in S \quad (10)
\]

\[
u_{ij} \in \{0, 1\} \quad \forall i, j \in T : i \neq j \quad (11)
\]

The objective function minimizes the number of unscheduled matches, and penalizes each pair of matches scheduled within \(R_{\text{max}}\) days. The first set of constraints ensures that each team meets each other team exactly once in a home game, unless the match is not scheduled. Consequently, each team will meet each other team exactly once in an away game as well, and these constraints are sufficient to construct a double round robin tournament \([C1]\). The next set of constraints make sure that each team plays at most once per time slot \([C5]\). Constraints (3)-(5) keep track of the number of days between two consecutive matches featuring the same team. The next set of constraints puts at least \(m\) calendar days between the two confrontations of a pair of teams \([C4]\). Constraints (7) enforce that a team plays at most two matches in a period of \(R_{\text{max}}\) slots \([C6]\). Constraints (8) make sure that there is no match between two teams on a particular time slot if the home team does not have its venue available \([C2]\), or if the away team marked this time slot in its forbidden game set \([C3]\). The final sets of constraints state that all variables are binary.

5 A heuristic approach

This section describes our heuristic approach, which is based on tabu search. The core component of our algorithm consists of solving a transportation problem, which schedules (or reschedules) all home games of a team \(i \in T\). This
All slots in team i’s home game set $H_i$.

Opponents of team i
day 5
day 19
team a
team b
day 258
team n
q

Fig. 1 The graph $G_i$ (example)

transportation problem is explained in section 5.1. Our heuristic approach consists of three phases: the construction phase (section 5.2), the tabu phase (section 5.3), and the perturbation phase (section 5.4).

5.1 Transportation problem

For any given team $i$, we construct a bipartite graph $G_i = (U, V, E)$ as follows. We have a set of supply nodes $U$, containing a node with supply equal to 1 for each slot $s \in H_i$, i.e. the home team set of team $i$, and a node $q$ with supply equal to $n - 1$, corresponding to an unscheduled slot. The set of demand nodes $V$ has a node with demand equal to 1 for each opponent of team $i$, i.e. $T \setminus \{i\}$, and a node corresponding to a dummy team. The demand of this last node is such that total supply equals total demand. Figure 1 represents an example of $G_i$.

The costs for each edge in $E$ are set as follows. The costs on the edges between the dummy team node and any node in $U \setminus \{q\}$ are zero (dashed edges in Figure 1). The costs on the edges from node $q$ to the non-dummy nodes are equal to $P$ (dotted edges in Figure 1). Finally, an edge from a node $u \in U$ corresponding to a home slot $s \in H_i$ and a node $v \in V$ corresponding to a team $j \in T \setminus \{i\}$ has a cost that corresponds with inserting a home game of team $i$ against $j$ on time slot $s$ in the current schedule (more details follow in Sections 5.2 and 5.3). If no matches involving team $i$ or $j$ have been scheduled so far, then the cost is zero. These edges correspond to the solid edges in Figure 1, and we will refer to them as such in the remainder of this text. Notice that the graph need not be complete: indeed, if the time slot corresponding to node $u$ is in team $j$’s forbidden game set, then there is no edge between $u$ and $j$. Solving this transportation problem will schedule (or reschedule) the home games of team $i$; if flow is sent from node $q$ to some opponent $j$, the home game of $i$ against $j$ is not scheduled. Notice that by construction, this problem is always feasible.
5.2 Construction phase

In the construction phase, we solve the transportation problem sequentially, for each team \( i \in T \). The order of the teams is determined by \( H_i \); we start with the team with the lowest number of available home game slots. The end result of the construction phase is a schedule (where some matches possibly remain unscheduled).

Initially, no matches have been scheduled, and hence the cost on the solid edges is zero. During the construction phase, the schedule will gradually be filled with matches, and the costs on the solid edges will increase accordingly. Indeed, the cost of an edge from a node \( u \in U \) to a node \( v \in V \), corresponding to scheduling the home game of team \( i \) against \( j \) on time slot \( s \), will depend on the previous and next match of \( i \), and the previous and next match of \( j \) in the preliminary schedule, with respect to time slot \( s \). For instance, assume that a match of team \( b \) has been scheduled on day 21, and an away game of team \( i \) has been planned on day 18. In this case, the cost of the edge between day 19 and team \( b \) in Figure 1 is set to \( p^1 + p^2 \), which corresponds to the increase in objective function value of the mathematical formulation described in section 4 if a match between \( i \) and team \( b \) is inserted on day 19.

Furthermore, an edge from \( u \) to \( v \), which corresponds to planning team \( i \)'s home game against team \( j \) on time slot \( s \) is removed if at least one of the teams \( i \) and \( j \) already has a match scheduled on slot \( s \), or if the match \( j - i \) is already scheduled within \( m \) days of slot \( s \).

Notice that there is in fact not a full correspondence between the objective function in section 4 and the costs as presented in this section. Indeed, when scheduling the home games of team \( i \), we do not take into account costs related to scheduling two successive home games of \( i \) in less than \( R_{\text{max}} \) days (only away matches of team \( i \) are considered for this). Consequently, the total cost resulting from solving the transportation problem is an underestimation of the cost in the problem description section. In practice, however, this has little or no effect, since teams almost never specify two home game slots with less than \( R_{\text{max}} \) days in between (for the majority of the teams, the home ground is available on a fixed weekday, every other week).

5.3 Tabu phase

Tabu search is a heuristic search procedure which goes back to Glover [8] and has proven its value in countless applications. In our implementation, the tabu phase works with a tabu list of length 5, and is initially empty. We randomly pick a team \( i \) which is not in tabu list, and add it to the tabu list. Next we remove all the home games of this team from the current schedule, and solve
the transportation problem for this team.

If the resulting schedule is better than previous schedule, we accept the new schedule and continue the tabu search phase with a new randomly picked team (different from \( i \)). In the other case, we impose changes to the home game assignment of team \( i \). We do this by sequentially resolving the transportation problem, each time with a different solid edge that was part of the previous schedule removed from the graph. From these solutions we select the best one, and accept the resulting schedule. Notice that this schedule may be worse than the previous schedule. Also in this case, we continue the tabu search phase with a new randomly picked team (different from \( i \)).

5.4 Perturbation phase

In order to escape local optima, we apply the following perturbation of the current schedule if 500 iterations without improvement of the best schedule found so far occur. We randomly remove 10% of the scheduled matches, in order to open up some space in the schedule, and continue the tabu search phase.

6 Computational results

We solved the indoor football scheduling problem for all divisions for the seasons 2009–2010 till 2012–2013, which corresponds with 18 instances. Most divisions have 15 teams, although some division have 13 or 14 teams. The season is played from September 1st to May 31st, which results in \(|S| = 273\) (leap years excepted). The home game set of a team has on average 4.4 slots more than what is needed for the league. However, in two instances, a team provided less home game slots than it has opponents, which inevitably leads to unscheduled matches. On average, teams ask not to play a match on 17 days. However, it turns out that 19 teams have a forbidden game set that contains more than the allowed 28 days. This is tolerated, since most of these teams also provide a home game set that largely exceeds the requirements. In the opinion of the league organizers, it suffices to have 3 days between two successive matches for a team (i.e. \( R_{max} = 4 \)). The penalties were chosen as follows: \( p^2 = 10 \), \( p^3 = 3 \), \( p^4 = 1 \). We set \( P = 1,000 \) in order to maximize the number of scheduled matches, and to be able to clearly distinguish the contribution of unscheduled matches from matches in close succession in the objective function value.

We implemented the formulation provided in section 4 using IBM Ilog Cplex, version 12.2. Notice that constraints (10) and (11) can be relaxed by stating that all \( y \) and \( u \) variables should be between 0 and 1. Indeed, given
Table 1 Results for real-life instances from seasons 2009–2010 till 2012–2013

<table>
<thead>
<tr>
<th>Instance</th>
<th>Teams</th>
<th>IP formulation</th>
<th>Heuristic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Best solution</td>
<td>Lower bound</td>
</tr>
<tr>
<td>1</td>
<td>14</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>57</td>
<td>43</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>68</td>
<td>19</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>3000</td>
<td>3000</td>
</tr>
<tr>
<td>6</td>
<td>15</td>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>15</td>
<td>43</td>
<td>20</td>
</tr>
<tr>
<td>8</td>
<td>15</td>
<td>2087</td>
<td>2067</td>
</tr>
<tr>
<td>9</td>
<td>13</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>14</td>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>15</td>
<td>50</td>
<td>12</td>
</tr>
<tr>
<td>13</td>
<td>15</td>
<td>29</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>15</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>15</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>16</td>
<td>15</td>
<td>113</td>
<td>60</td>
</tr>
<tr>
<td>17</td>
<td>15</td>
<td>60</td>
<td>9</td>
</tr>
<tr>
<td>18</td>
<td>15</td>
<td>2010</td>
<td>2000</td>
</tr>
</tbody>
</table>

The objective function, the integrality conditions on the $x$ variables (9) are sufficient to ensure that the $y$ and $u$ variables are 0 or 1. All models were run on a Windows XP based system, with 2 Intel Core 2.8GHz processors, with a maximum computation time of 5000 seconds. The heuristic was implemented using C++ and run on the same machine, however with a maximal computation time of 500 seconds. The transportation problems were solved to optimality using an implementation of Kuhn-Munkres algorithm.

Table 1 presents our computational results. The first two columns provide the instance number and the number of teams in this division. The next two columns show the best found solution and lower bound found within the given computation time using the IP formulation. Only 5 of the 18 instances were solved to optimality; for other instances the maximal number of matches was scheduled, given team availabilities. The computation times in the fifth column indicate that if Cplex manages to find and prove optimality, this usually happens rather quickly. The final column shows the best found solution by our heuristic. In 4 cases, the heuristic approach found an optimal solution, however, for 6 instances, the heuristic failed to schedule the maximal number of matches. This is not as bad as it may appear in terms of objective function value, since in practice, a date for an unscheduled match is settled through negotiations under the responsibility of the home team. It is striking that in 7 cases, the heuristic resulted in a better solution than Cplex, despite being given 10 times less computation time.
7 Conclusions and future research

In this paper, we described and solved a sports scheduling problem for the LZV Cup, an amateur indoor football competition. This scheduling problem is interesting, because matches are not planned in rounds. Instead, each team has a number of time slots available to play its home games; away teams can specify days on which they are not able to play. Furthermore, successions of home or away matches are irrelevant. The goal is to balance each team’s matches over the season, in the sense that there should be no close succession of matches involving the same team.

We developed an integer programming formulation and a heuristic approach, and used them to generate schedules for all divisions for the seasons 2009–2010 till 2012–2013. Overall, the performance of the tabu search based heuristic is comparable to that of Cplex, however, the reduced computation time, and the absence of expensive licenses make the heuristic implementation more suitable for (amateur) competitions such as the LZV Cup. These schedules were approved by the league organizers and have been implemented in practice, much to the satisfaction of the participating teams. In rare occasions where not all matches could be scheduled, the organizers appreciated that our approach makes it clear with which team to put the responsibility to find a solution.

Some future work remains. First of all, it would be interesting to integrate the planning of the cup competition in the scheduling process. Indeed, if the cup and the league are planned together instead of sequentially, an even better spread of the matches could be accomplished. Indeed, all teams take part in a cup competition, creating dependencies between the division schedules. Currently this matter is handled by scheduling the first round matches of the cup competition beforehand, simply by using the first time slot for which the home team has its venue available and which is not mentioned in the visitor’s forbidden game set. If no such time slot exists between September 1st and October 15th, then we invert the home advantage. If this still does not result in a solution, we remove this cup match from the schedule, and leave it up to both opponents to find a suitable time slot themselves (e.g. by finding another venue). With this procedure, the divisions can be scheduled independently from each other. Whereas simultaneous scheduling of all divisions and the cup could clearly improve the quality of the schedules, a formulation like the one provided in section 4 may not be tractable in this case, even for advanced IP solvers. It could also be interesting to make an educated guess about which teams will survive the first and the following rounds in the cup, such that for these strong teams, we can leave gaps in their league schedule such that future cup matches can be fit in more easily. Finally, it would also be interesting to test the performance of the heuristic on a number of instances from other (amateur) indoor sports.
References