Project on Optimization

Jacob Puthipiroj

Introduction

The sports industry is huge. Even within a one league of one single sport, revenues can amount to over billions of dollars per season. The investment of capital in creating a marketable sports tournament is similarly enormous: the booking of stadiums and television slots, the salaries of coaches and players, the pre-game advertising and post-game roundup. Executives have every incentive to maximize the profits of each game while adhering to constraints imposed by the industry. Consequently, the once mundane scheduling of games − the specification of when, where, and which teams play each other − becomes a mathematical optimization problem capable of saving millions of dollars.

In this paper, I explore a constraint-programming approach to this optimization problem in the context of a scheduling a Major League Soccer tournament. The setup is as follows:

Objective Function: The total distance traveled altogether by teams, to be minimized.
Decision Variables: The trip start dates for each team.
Constraints: Each team cannot play more than two consecutive games away from home. Each team must play all other teams exactly once.

As shown in the final spreadsheet, this implementation led to the saving of over 10,000 kilometers of travel for a 5-team problem. When scaled up to real situations, this could result in saving millions in transportation and accommodation costs, as well as latently boosting revenues by allowing better performance from well-rested athletes.

Model Uncertainty and Sensitivity Analysis

Unlike blending problems, where the certainty assumption of certain parameters are dubious and sensitivity analysis is helpful in understanding the stability of optimal solutions, tournament scheduling problems have relatively fixed parameters. It is incoherent, say, to alter the distance matrix in theory, as cities cannot move in practice. Interestingly, Kostuk (2012) broadens the definition of sensitivity analysis in the TPP as 'how does the optimal schedule distance change given different schedule constraints'?

The paper explicitly mentions an 'iterative process arriving a definitive schedule', whereby three or four different bounce back and forth between them as the details are fine-tuned to incorporate new constraints[1]. Some of these constraints were based on team popularity, e.g. the four teams with the lowest drawing power should not play each other in the first 5 rounds because attendance would be low, while each of the three most popular teams must play exactly one classic matchup at home. The authors did not report optimal solution distances given any, or even all of these constraints, but instead reported that the value of the feasible solution in the objective function was 5.6% better than in the relaxed problem. Surprisingly, constraining the problem lead to an *improvement*, rather than a reduction in quality of the solution. This was partially

---

[1]#modelSensitivityandUncertainty: I address the problem of performing useful sensitivity analysis in fixed-parameter problems and provide a workaround by broadening the definition of sensitivity analysis, while citing the relevant literature.

due to the nature of the objective function itself; instead of the total distance traveled by all teams, a unit of distance to be minimized (as in the implementation), the authors aimed to maximize the 'concentration of games between teams within the same group towards the final round of the tournament'.

Implementation Challenges

This was a difficult problem; deliberately so in order to simulate a real-world problem that would be impossible to solve by hand. There were many problems encountered:

1. **Understanding the correct approach:** Instead of considering one game at a time (between two teams), I intuited that each team would have its own optimal (minimum) schedule, and hopefully the total optimal schedule will contain all optimal sub-schedules.

2. **Constructing the single-team solution:** This is itself a linear optimization problem, with the most difficult part being the constraint against two-cycles. The was resolved by creating a variable for the total number of trips going in and out of each city pair, limited to 1 at most.

3. **Combining the single-team solutions**: A major breakthrough came when realizing that reverse trips were fundamentally different from forward trips, and helped to break the symmetry of the solution. Instead of 2 trips × 10 days = 20 decision variables, using 4 trips (2 forward, 2 reverse) × 10 days allowed easy and accurate encoding of certain difficult constraints, including the "no 3 consecutive home/away games" rule.[2]

Generalizable Lessons

Obtain Optimal Patterns by combining Optimal Subpatterns

Constraint Programming can accompany Integer Programming in order to make it much more powerful. The article specifically mentions this in terms of feasible 'patterns' obtained from each team which were either explicitly or implicitly assigned. This was directly used in the implementation. The optimal tour for Denver, for example, is as follows:
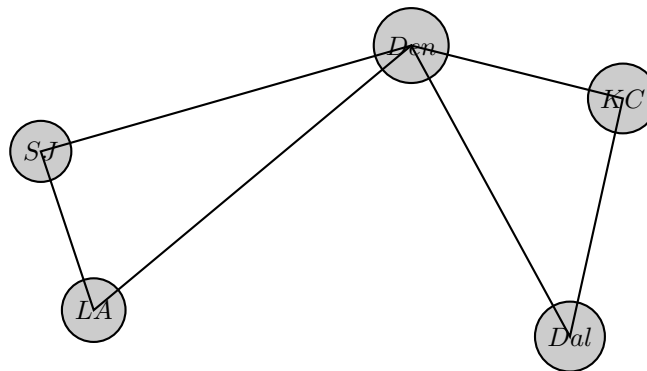
The pattern is clear: make one trip to the two closest cities, and another to the two furthest cities[3]. By hardcoding this pattern into the solution, it is easy to see that, say, for KC, its optimal tour will be one of

---

[2]#breakitdown: I effectively decompose a problem into tractable, clear, and well defined components (the whole schedule into simply the schedule for each individual city); accurately combines component solutions into a full solution (by finding linear combinations of each of the 5 subtours on different days).

[3]#dataviz: This visualization creates new, significant insights into the construction of optimal subtours for each team.

(Den-Dal or Dal-Den) $\times$(SJ-KC or KC-SJ) = 4 total paths. The final schedule are thus linear combinations of $4 \times 5 = 20$ different paths - clearly, a computationally tractable number.[45]

Insights Regarding LOs

The TPP is a 'notoriously difficult problem', but educationally rewarding. Not only did I successfully minimize the objective function - the schedule distance - to within 4.6% of the theoretical minimum, but did so within the context of realistic constraints - that teams may not spend more than 2 consecutive games away from home. In fact, this problem was only solveable in Excel *because* the constraints allowed the search space to be reduced via constraint programming.

This assignment also provided an interesting take on the concept of #modelSensitivityandUncertainty by broadening the definition of 'Sensitivity Analysis' to include fixed-parameter/roughly deterministic problems. In fact, the nature of the problem revealed how closely the two LOs interact, since the primary method for reducing model sensitivity to the uncertainties of the real world was through the addition of model constraints!

---

[4]#entObjectivesandConstraints: The constraint that that each team cannot play more than two consecutive games away from home would normally be a soft constraint (should be obeyed), rather than a hard one (must be obeyed). However, by making it a hard constraint, and turning the problem into a constraint programming one, I find an optimal feasible solution for the soft constraint instance.

[5]#constraints: Actually, not only did I find a solution that respects the constraints, I would not have been able to find an optimal solution (at least in Sheets) without it. This constraint was a blessing, because it significantly reduced the search space from $20!/2^{10} \approx 10^{15}$ to about 10^5.

Implementation Appendix

1 The Single Team Problem

Decision Variables

The games are defined using $\forall i \neq j \in I$ (the set of teams) as the following binary variables:

$$x_{ij} = \begin{cases} 1 & \text{if team } i \text{ plays against home team } j \\ 0 & \text{otherwise} \end{cases}$$

Variables

$d_{ij} =$ the distance between city i and home team j

Objective Function

For a single team i, the objective is the total distance for it to travel to all the cities

$$\sum_{i=1}^{5} \sum_{i=1}^{5} d_{ij} x_{ij}, \quad x_{ii} = 0$$

In practice, d_{ii} was made large enough that x_{ii} was never chosen, effectively making $x_{ii} = 0$.

Constraint Function

Team i must enter all other cities eactly once and leave all other cities exactly once.

$$\sum_{j \neq i} x_{ij} = 1 \text{ for all } i \neq 1$$

$$\sum_{i \neq j} x_{ij} = 1 \text{ for all } j \neq 1$$

However, team i will exit its city exactly twice, and return to its city exactly twice.

$$\sum_{j \neq i} x_{ij} = 2 \text{ for } i = 1$$

$$\sum_{i \neq j} x_{ij} = 1 \text{ for } j = 1$$

In order to avoid 2-cycles, each pair of cities i and j can have at most one enter or exit status.

$$x_{ij} + x_{ji} \leq 1$$

2 The Multiple Team Problem

Decision Variables

In this problem, there is a distinct difference between going on a tour, and going on it in reverse. *i.e*, $(a \to b \to c \to a) \neq (a \to c \to b \to a)$

$$y_{ijk} = \begin{cases} 1 & \text{if team } i \text{ starts trip } j \text{ on day } k \\ 0 & \text{otherwise} \end{cases}$$

$$z_{ijk} = \begin{cases} 1 & \text{if team } i \text{ starts trip } j \text{ in reverse on day } k \\ 0 & \text{otherwise} \end{cases}$$

Variables

$d_{ij} =$ the distance between city i and home team j

Objective Function

For all teams, the objective is the total distance traveled to all cities.

$$\sum_{j=1}^{5} \sum_{i=1}^{5} \sum_{k=1}^{10} d_{ij} y_{ijk} + \sum_{j=1}^{5} \sum_{i=1}^{5} \sum_{k=1}^{10} d_{ij} z_{ijk}$$

Constraint Functions

Exactly one forward/reverse trip is allowed per j.

$$\sum_{k=1}^{10} y_{ijk} + \sum_{k=1}^{10} z_{ijk} \leq 1 \text{ for all } i = 1, ..., 5, j = 1, 2$$

Teams may not have two road trips in a row.

$$\sum_{k=l}^{l+2 \bmod 10} (y_{i1k} + y_{i2k} + z_{i1k} + z_{i2k}) \leq 1 \text{ for } l = 1, ..., 10$$

Home teams must be present when visitors arrive.

$$\sum_{l \neq i, i \text{ first in } j(i,l)} (y_{lj(i,l)k} + z_{lj(i,l)(k-1)\bmod 10}) \quad +$$

$$\sum_{l \neq i, i \text{ second in } j(i,l)} (y_{lj(i,l)(k-1)\bmod 10} + z_{lj(i,l)k})$$

$$\leq 1 - \sum_{m=(k-1)\bmod 10} \sum_{j=1,2} (y_{ijm} + z_{ijm})$$

$$\text{for } k = 1, ..., 10; i = 1, ..., 5$$

References

[1] James C Bean and John R Birge. Reducing travelling costs and player fatigue in the national basketball association. *Interfaces*, 10(3):98–102, 1980.

[2] Rishiraj Bhattacharyya. A note on complexity of traveling tournament problem. *Optimization Online*, 2480, 2009.

[3] John R Birge. Scheduling a professional sports league in microsoft® excel: Showing students the value of good modeling and solution techniques. *INFORMS Transactions on Education*, 5(1):56–66, 2004.

[4] Federico Della Croce and Dario Oliveri. Scheduling the italian football league: An ilp-based approach. *Computers & Operations Research*, 33(7):1963–1974, 2006.

[5] Guillermo Durán, Mario Guajardo, Jaime Miranda, Denis Sauré, Sebastián Souyris, Andres Weintraub, and Rodrigo Wolf. Scheduling the chilean soccer league by integer programming. *Interfaces*, 37(6):539–552, 2007.

[6] Kelly Easton, George Nemhauser, and Michael Trick. The traveling tournament problem description and benchmarks. In *International Conference on Principles and Practice of Constraint Programming*, pages 580–584. Springer, 2001.

[7] Kelly Easton, George Nemhauser, and Michael Trick. Solving the travelling tournament problem: A combined integer programming and constraint programming approach. In *International Conference on the Practice and Theory of Automated Timetabling*, pages 100–109. Springer, 2002.

[8] Jacques A Ferland and Charles Fleurent. Computer aided scheduling for a sport league. *INFOR: Information Systems and Operational Research*, 29(1):14–25, 1991.

[9] Kent J Kostuk and Keith A Willoughby. A decision support system for scheduling the canadian football league. *Interfaces*, 42(3):286–295, 2012.

[10] George L Nemhauser and Michael A Trick. Scheduling a major college basketball conference. *Operations Research*, 46(1):1–8, 1998.

[11] Edo Stephan Poort. *Aspects of sensitivity analysis for the traveling salesman problem*. PhD thesis, [University Library Groningen][Host], 1997.

[12] Robert A Russell and Janny MY Leung. Devising a cost effective schedule for a baseball league. *Operations Research*, 42(4):614–625, 1994.

[13] Michael A Trick. Using sports scheduling to teach integer programming. *INFORMS Transactions on Education*, 5(1):10–17, 2004.