OPTIMIZING SCHEDULES FOR CHILEAN SOCCER

COURSE PROJECT
ISYE 6501

JULY 2020

Introduction

While looking at the world map as a child, Chile always stood out to me as unique. Why was it so long and skinny? If you had to travel from the north to the south in a car, how long would it take? Why didn't they just make it a rectangle instead? It was as if I knew, even then, that an optimization problem would be needed to figure out how to travel efficiently around Chile.

Because of this prior intellectual curiosity, figuring out a way to efficiently optimize the fan experience for the Chilean soccer team seemed like a fun analytics problem to dive into.

Background

Soccer (football) is the most popular sport in Chile. In fact, the national team won the Copa América tournament in both 2015 and 2016. The top division in the top league in Chile is called the Chilean Primera División or the First Division. The First Division consists of 18 teams. In total, there are five leagues or divisions in Chilean soccer.

In 2004, the season was scheduled in a split-season format. Half of the season was played in the spring and half of the season was played in the fall. Each team played every other team in the First Division once in the spring and once in the fall. At the end of the season, the top two teams advanced to a playoff where they would play for the championship. The two worst teams in the First Division would be relegated to the Second Division at the end of the season while the two best teams in the Second Division would move up to the First Division.

Problem

In 2004, match attendance and general public interest were waning in the Chilean Soccer League, which is run by the Chilean Professional Soccer Association (Asociación Nacional De Fútbol Profesional) (ANFP). Part of the problem was the schedule that they were using. Here were some of the problems with the schedule.

1. Travel time was not considered, and, at times, teams had to travel significantly more than other teams. Teams that logged more travel time tended to perform worse than teams who logged less travel time due to fatigue.
2. The weaker teams would often have to play the top three teams (Colo-Colo, Universidad Católica, and Universidad de Chile) back to back or even three times in a row.
3. Some teams would play three road games in a row. This resulted in a drop in revenue during those road trips.
4. Some weaker teams were not scheduled to host the top three teams. These games, while difficult, would draw more fans and increase revenue for the weaker teams.
5. The schedule was made by doing a random drawing to assign slot numbers. Then the league used the canonical one-factorization method to make the schedule. This method, while effective, takes no outside factors into consideration. In fact, the method is more than a century old!

On another note, this problem was not reserved to South American soccer. A husband and wife team made the Major League Baseball schedule from 1981-2005 using a giant wall of magnets to schedule 2430 games!

Goal

In order to boost attendance and public interest in Chilean Football League soccer, the ANFP sought to:

1. Schedule the soccer season to make the match calendars more attractive to fans.
2. Schedule the soccer season to make the match calendars more attractive to the teams.
3. Make the soccer season fair and balanced for all the teams in sporting and economic terms.

Note

In order to simplify my project, I am going to focus on optimizing the regular season schedule in the First Division. The Chilean Professional Soccer Association's (ANFP) work since 2004 has focused on each level of Chilean professional soccer, not just the First Division.

Step 1

First, the schedule needs to be optimized to reduce travel time for all teams. Ideally, the travel time is about the same for each team in the league. Data is not hard to collect since we can simply measure distances between cities and analyze travel times via airplane, bus, train, etc.

> **Given** distances between cities, flight durations between cities, train/bus ride durations between cities
> **Use** constraint programming and integer programming
> **To** minimize the travel time between games for all the teams.

The Chilean soccer season is defined as a double round robin tournament (DRRT). That is because each team plays each other twice, once at home and once away. For the purposes of the Chilean Soccer League our DRRT is split across a spring and fall section of the season.

We can use an integer program and a constraint program to set up the schedule. A constraint program applies higher level constraints that apply to integer values. An integer program has decision variables that are constrained to be integer values.

- Input: A set of *n* teams where *n* is an even number, a symmetric *n* by *n* integer distance matrix with elements dij, *u* integer parameters, *l* <= *u*
- Optimization Program
    - Variables: A set of columns composed of optimal tours for each team, a vector representing the distances associated with the tours, the set of teams in the league, the set of the slots, the indices of tours for each team, a binary decision variable that represents the tours
    - Constraints:
        - Select exactly one tour for each team.
        - Each team plays exactly once in each slot.
        - Each team will play the same number of home and away games.
    - Objective Function: Minimize the distance traveled by each team.
- Output: A DRRT tournament on the teams in T

At this point, we can compare the integer program and the constraint program to see which one came up with the better schedule. The better schedule is the one that minimizes the travel time for teams and satisfies the constraints.

Some potential pitfalls with this approach are that integer programming and constraint programming can take a long time to achieve an optimal solution, if one is even obtained. A heuristic could be used if an optimal solution cannot be reached.

Step 2

Next, the schedule needs to be set up to make the home/away splits fair for teams by giving them an even number of home games, the same number of home games on TV, etc.

> **Given** the schedule from step 1, stadium capacities, team revenues, average home attendance
> **Use** integer programming and constraint programming
> **To** optimize the schedule for each team related to home games.

We'll set up both an integer program and a constraint program just like in step 1.

For this, the program is the same, but we have some new constraints.

1. The length of each home stand and road trip must not exceed three games.
2. Each team is at home for at least 45% (50% if possible) of the weekend games.
3. Schedules where a team plays home-away-home will be minimized to avoid excessive travel.

I would expect the integer program to perform better at generating a schedule, but it's worth checking both approaches to see which one is the most effective.

The same problems would exist with this step, just like in step 2 because the same models are being used. This step makes the model more complex, so that could make an optimum solution more difficult to obtain. An option would be using a heuristic that would give a good solution that is not necessarily optimal.

Step 3

Finally, I will seek to update the second half of the schedule based on the results of the first half of the schedule.

> **Given** goal differential, points, TSR, %TSoTt, PDO obtained from the teams
> **Use** linear regression
> **To** adjust the second half of the schedule (fall) to make it more competitive.

Since there is a break between the first half of the season (spring) and the second half of the season (fall), that will provide the perfect time to make an optimal schedule for the second half (fall) of the season. The schedule will be released during the summer and it will be adjusted to ensure that the most important games are played on the weekends and during the end of the

season. This sort of analysis can be added to our integer programming and constraint programming models to make a schedule that ensures premium matchups at the end of the season.

What is missing is an analysis of who the best teams really are. It is easy to go simply by points (3 points for a win, 1 point for a draw, 0 points for a loss) to see who is the best team, and that would provide a good start. However, teams often are not as good as their record suggests, or vice versa. For example, a team with an above league average point total and a goal differential close to zero would be over performing and prime for regression in the second half of the season. If teams that are over and underperforming could be identified, we could select the very best teams to focus on in order to set up the best games for the end of the season.

Team strength could be measured with several factors.

1. Goal differential
    a. (Goals scored - Goals allowed)
2. Points
    a. 3 points for a win, 1 point for a draw, 0 points for a loss
    b. (Points from wins + Points from draws)
3. TSR
    a. Total Shots Ratio
    b. (Total Shots for / (Total shots for + Total shots against))
4. %TSoTt
    a. The propensity for a team to take shots on target and force its opponent to take shots that are off target
    b. (Shots on target for/Total shots for) + (Shots off target against/Total shots against)
5. PDO
    a. The propensity of a team to score shots on target and save shots on target by its opponent
    b. 1000 * (goals for/shots on target for + (shots on target against = goals against)/shots on target against)

These factors could be fed into a linear regression model to predict the relative strength of the teams, or determine a team rating. The teams with the higher team ratings would have “priority schedules”. A priority schedule would have more games against teams with high team ratings and prime time (evening and weekend) games during the last quarter of the season. This would increase fan interest at the end of the season.

The teams with lower team ratings could play each other at the end of the season. This is because the games where the loser of the game would face relegation could also be intriguing for fans.

The drawbacks with this approach might be that there are too few factors considered when determining the team rating or relative strength of the team. Another option would be to use Five Thirty Eight’s Global Club Soccer Rankings, should they ever include Chilean clubs.

Step 4 (Maintenance)

Each year, the models would need to be re-run to set up a new schedule. Calendars change, and another important factor to consider is that each year two teams are moved up from the Second Division to the First Division, and two teams are relegated from the First Division to the Second Division. This creates different distances to travel and a new set of variables to optimize.

Summary

These are some of the methods that could have been used to improve the scheduling of Chilean soccer since 2004. The topic of sports scheduling is evolving quickly and is using cutting edge technology to come up with even better solutions. Clearly, exciting and groundbreaking changes are happening in the world of sports scheduling that can enhance the team, fan, and player experience.

Bibliography

Dinitz, Jeffrey H. “Designing Schedules for Leagues and Tournaments” University of Vermont-Burlington, 13 Nov. 2004, http://www.emba.uvm.edu/~jdinitz/preprints/design_tourney_talk.pdf

Easton, Kelly, et al. “The Traveling Tournament Problem Description and Benchmarks.” *Principles and Practice of Constraint Programming — CP 2001 Lecture Notes in Computer Science*, 2001, pp. 580–584., doi:10.1007/3-540-45578-7_43. https://mat.tepper.cmu.edu/trick/ttp.pdf

Grayson, James W., “Methodology and Validation of the Team Ratings” James’ Blog, 11 Aug. 2014, https://jameswgrayson.wordpress.com/2014/08/11/methodology-and-validation-of-the-team-ratings/

Informs. “Leveraging O.R. to Achieve the Competitive Edge.” *INFORMS*, 12 Apr. 2016, www.informs.org/Impact/O.R.-Analytics-Success-Stories/Leveraging-O.R.-to-Achieve-the-Competitive-Edge.

Trick, Michael A. “Integer and Constraint Programming Approaches for Round-Robin Tournament Scheduling.” *Practice and Theory of Automated Timetabling IV Lecture Notes in Computer Science*, 2003, pp. 63–77., doi:10.1007/978-3-540-45157-0_4. https://mat.tepper.cmu.edu/trick/tourn_final.pdf