OR APPLICATIONS IN SPORTS SCHEDULING AND MANAGEMENT

Celso C. Ribeiro and Sebastián Urrutia

Catholic University of Rio de Janeiro, Department of Computer Science
Rua Marquês de São Vicente 225, Rio de Janeiro 22453-900, Brazil
E-mail: {celso, useba}@inf.puc-rio.br

October 2003

1. Introduction

Sports management is a very attractive area for applications of Operations Research. Sports competitions involve many economic and logistic issues. Athlete evaluation, team evaluation, tournament planning, clubs management, economic estimation, marketing politics, security issues and designing fair rules are some of the issues involved in sports management.

OR methods have been successfully applied in the evaluation of team performance. Team managers, players, fans, and journalists are often eager to know how close a team is of classification. Probabilistic and exact methods are used to tackle this question. Timetabling is the major OR application in sports. Sport leagues and teams need schedules satisfying different types of constraints and optimizing, for example, the fairness or the distance traveled. Timetable optimization in sport competitions is a hard problem for which several OR techniques have been applied. Also, sport applications motivate the students and therefore can also be successfully used in OR courses and classrooms.

2. Deciding playoff elimination

Major League Baseball (MLB) in the United States and the soccer national championship in Brazil are two major competitions. Both tournaments are divided in two stages. In the first stage, qualification for the playoffs is decided. Baseball fans in United States are much like soccer fans in Brazil. Their love for statistics is worldwide known. Both tournaments are followed by millions of people, who attend the games in the stadiums, follow radio and TV transmissions, and check newspapers, radio, TV, and, more recently, the Internet in search of information about the performance and the chances of their favorite teams.

The US press was used to base their predictions about chances of qualification to the playoffs on statistics called "games by", i.e. the number of games a team has to win to reach the leading team in case the latter were not to win again. If the number of games to

1

be played by the trailing team is smaller than this value, it can be said eliminated. This method does not take into account the remaining schedule of games to be played. It is therefore conservative, in the sense that it gives only a weak sufficient condition for elimination. Adler et al. [1] developed integer programming models to calculate the number of games a team has to win to reach a playoff spot. These models can determine earlier than the previous method used by the press if a particular team is eliminated. Besides the models themselves, their work is focused on detecting elimination and on making this information easily reachable and available through the Internet at their web site located at http://riot.ieor.berkeley.edu/~baseball.

Several characteristics of the MLB make this problem easier to solve than in the context of the Brazilian soccer championship. First, as games do not end tied in baseball, a game can only have two final outcomes, instead of the three possible outcomes of a soccer game. While baseball follows the simple (1,0) rule, soccer follows the more complex (3,0)-(1,1) rule: each team gets three points for a win and one point for a tie. Moreover, teams in MLB are divided in two leagues and in six divisions of at most six teams each, which gives rise to small subproblems where a few teams dispute a single playoff spot (the wildcard spot is disputed between all teams in each of the two leagues, 14 teams in the American League and 16 in the National League). On the contrary, in the case of the 2002 edition of the Brazilian championship, all 26 teams competed for eight playoff spots. Bernholt et al. [2] showed that deciding playoff elimination in soccer under the (3,0)-(1,1) rule is NP-complete, while Schwartz [7] showed that it can be solved by a maximum flow algorithm in the case of the MLB.

Soccer is a major economic activity in Brazil. Thousands of jobs in teams, leagues, newspapers, radio and TV stations, hotels, stadiums, and air companies depend directly from the activity of the teams. Teams which are not qualified to the playoffs loose a lot of money and are often forced to dismantle their structure and to sell their best players to be able to cover the maintenance expenses during the almost two months of the playoffs, when they do not get any income from tickets and TV rights. The major goal of each team is to be classified in one of the eight first positions in the standing table at the end of the qualification stage, so as to be qualified to the playoffs. Playoff games are those with more attendance and higher TV rights. Teams in the last four positions at the end of the qualification phase are degraded to a lower division and do not play the main league in the next year. Teams degraded to the second division often go even into bankruptcy. The press and fans of every team speculate and follow the standing table every day, attempting to anticipate the chances of qualification and elimination.

Ribeiro and Urrutia [5] proposed two integer programming models which are able to detect in advance when a team is already qualified to or eliminated from the playoffs. These models are based on the computation of the *Guaranteed Qualification Score*

(GQS), which is the minimum number of points a team has to obtain to be sure it will be qualified, regardless of any other results, and the *Possible Qualification Score* (PQS), which consists in the minimum number of points a team has to win to have any chance to be qualified. A team is *mathematically qualified* to the playoffs if and only if its number of points won is greater than or equal to its GQS. Only at this point its qualification to the playoffs can be announced without any risk of misinformation. A team is *mathematically eliminated* from the playoffs if and only if its current number of points plus the number of points it is still able to win in the remaining games to be played is smaller than its PQS. An additional nice feature of these models is that they can be easily extended to accommodate some of the most usual tie breaking rules.

Results from these models can be used not only to guide teams and fans, but are also very useful to identify and correct wrong statements made by the press and team administrators. Such statements are often based on obscure probabilistic estimations of the chances of qualification. Early in the qualification stage, it is said that any team will be qualified if it reaches a certain threshold in the number of points won. The estimation of this threshold is an approximation based on rough tournament statistics. When the first team reaches this threshold, the media says that its chance of qualification is 100% and that it is "mathematically qualified". Very often, this information is not correct.

These models were applied to the 2002 edition of the Brazilian national soccer championship, played by the 26 major teams from August 10 to December 15. The games were distributed over 29 rounds and not every team played at every round. The *guaranteed qualification* and *possible qualification* problems were considered in two contexts. In the first, we were interested in observing the quest for the first eight positions in the standing table, which define the teams qualified to the playoffs. In the second case, we were interested in observing the teams in the four last positions of the standing table, since these teams would not play in the first division in the next year. The two integer programming models were solved for every team immediately after the games in each round were completed. Since there were 26 teams in the tournament and we were interested in computing the guaranteed qualification score and the possible qualification score for each of them in the two above contexts (qualification to the playoffs in the first eight positions and not finishing in the last four positions), 104 integer programming problems had to be solved immediately after every round was completed. CPLEX 5.0 was used as the integer programming solver. GQS and PQS scores for each team, together with other statistics, were made available immediately at the web site of the **FutMax** project, now hosted at http://futmax.inf.puc-rio.br.

TV and radio interviews, together with articles in daily newspapers, generated huge debates about the quality and the use by the press of the GQS and PQS scores and other statistics. The two models and the web site of the **FutMax** project showed their usefulness several times during the championship. After the 11th round, on September 15, the coach of Vasco da Gama (one of the major teams) and supervisor of the Brazilian national soccer team, Antonio Lopes, said to the media in a press interview that his team would be qualified to the playoffs if it could win ten out of the remaining 14 matches to be played. The GQS and PQS scores showed that his statement was not true and there was indeed a set of possible results for the remaining games that led Vasco da Gama outside the qualification zone even if it had won all of its remaining games. São Paulo

reached 43 points at the 24th round on October 31. Using rough probabilistic estimates, the press loudly announced that it was the first team mathematically qualified to the playoffs. Once again, the GQS and PQS scores showed this was not true, showing the existence of a combination of results of the remaining games that led São Paulo out of the eight first positions.

Figure 1 displays a typical plot illustrating the evolution of the GQS and PQS scores round after round for Fluminense, a major team from Rio de Janeiro. Besides these two scores, it also plots the number of points P obtained after each round and the maximum number of points MNP that could be obtained if the team won all remaining games (three additional points for each of them). PQS and P are non-decreasing, while GQS and MNP are non-increasing along the championship. A team is mathematically qualified to the playoffs as soon as P gets larger than or equal to GQS. Analogously, a team is mathematically eliminated from the playoffs when MNP gets smaller than PQS. A not yet qualified team depends only on its own to be qualified to the playoffs if MNP is larger than or equal to GQS. Similarly, a not yet qualified team depends from results of the other teams to be qualified if MNP is smaller than GQS.
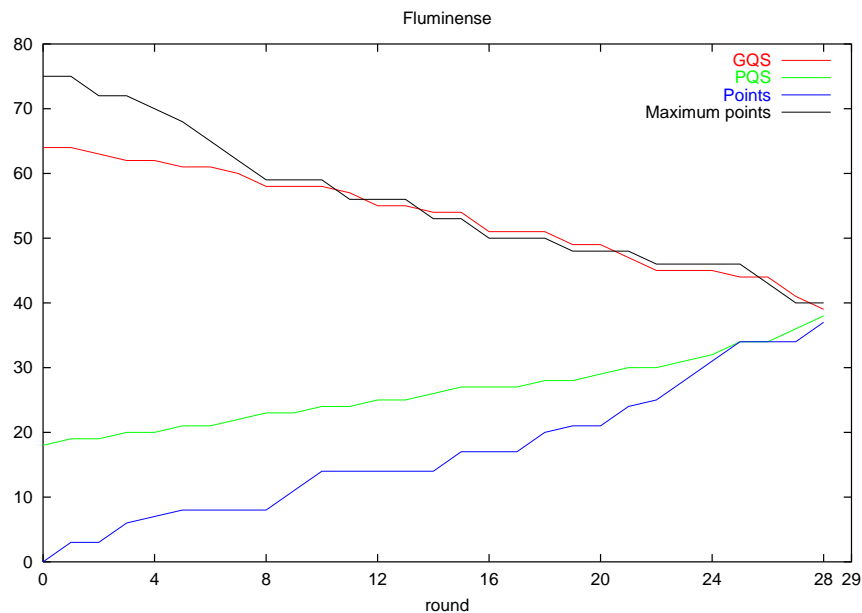

Figure 1: GQS and PQS scores for Fluminense along the 2002 Brazilian national soccer tournament.

This figure illustrates that MNP got smaller than GQS for Fluminense at the 11th round, showing that at this time it was not on its own: even if it obtained the maximum number

4

of possible points by wining all its remaining games, there existed at least one set of results leading Fluminense outside of the qualification zone. At the 25th round, its number of points reached its PQS, meaning that it had a chance to qualify even if it lost all remaining games. Indeed, due to a sequence of favorable results, Fluminense did qualified at the 8th and last position only at the final round.
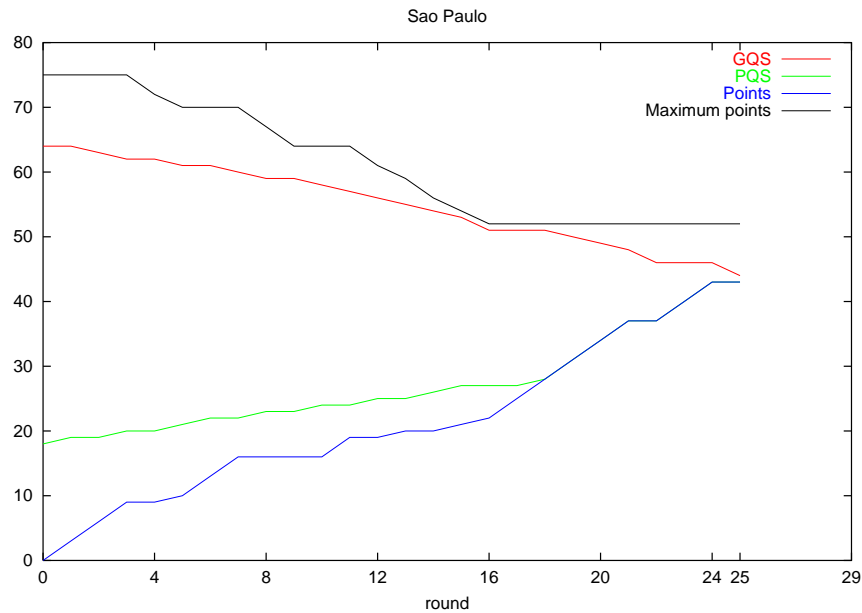

Figure 2: GQS and PQS scores for São Paulo along the 2002 Brazilian national soccer tournament.

Figure 2 illustrates the plots of the GQS and PQS scores for São Paulo, a major team from the city of São Paulo, which ended up the qualification phase in the first position of the standing table. This team was the first to be mathematically qualified for the playoffs, when its number of points reached its GQS score at the 26th round.

3. Sports timetabling

Professional sports leagues are a major economic activity around the world. Television costs in sports such as baseball, soccer, hockey, and basketball amount to hundreds of millions of dollars in some competitions. Recent transfers of major soccer players such as Ronaldo and Beckham to Real Madrid amounted to more than 50 millions dollars each. The sponsorship of national teams by sporting goods makers such as Adidas, Nike, and Reebok involve huge contracts. As an example, the contract of the Brazilian Soccer Confederation (CBF) with Nike amounts to almost 400 millions dollars. Investments in marketing and advertisement are also enormous.

Teams and leagues do not want to waste their investments in players and structure as a consequence of poor schedules of games, involving e.g. unattractive teams playing on prime dates or several important games played at the same time with losses in television rights. National and international competitions played in parallel require strong coordination of travel and game schedules. In this context, efficient schedules are of major interest for teams, leagues, sponsors, fans, and the media [3,6]. This problem is further complicated due to the traveling distances involved. In the case of the Brazilian soccer national championship, a single trip from Porto Alegre to Belém takes almost a full journey and many stops, due to the absence of direct flights to cover a distance of approximately 4000 kilometers. The total distance traveled becomes an important variable to be minimized, so as to reduce traveling costs and to give more time to the players for resting and training along a season that lasts for approximately six months. In the case of the Brazilian national soccer championship, fairness rules stipulate that no team is supposed to play more than three home or away games consecutively. Other additional constraints vary from tournament to tournament.

A computer system whose core is a heuristic mostly based on the GRASP metaheuristic [4] was developed and applied to the scheduling of the First Division of the 2003 edition of the Brazilian national soccer championship, in which 24 major teams took part. Feasible schedules were obtained in a few seconds of processing times. The results are impressive: the total distance traveled amounts to only 554,378 kilometers, while the game schedule obtained by hand by the personnel of CBF corresponded to approximately 1,100,000 kilometers. Total savings in airfares amount to 2.7 million dollars. Savings on traveling distances and, consequently, in airfares, amount to 50% of the current total cost! Hotel costs during the road trips can also be taken into account.

Besides the quality of the solutions found, the main advantages of the new heuristic and the associated computational system are their ease of use and the construction of several alternative schedules, making it possible to the decision maker planning the competition to compare and select the most attractive schedule from among different alternatives which can contemplate other secondary goals and constraints, such as scheduling to weekends or major holidays the classical regional games characterized by the rivalry of teams and fans; scheduling only one game at each round in small cities or in cities with only one stadium; avoiding two rival visiting teams in the same city at the same day to avoid clashes between their fans; complementary schedules for teams sharing the same stadium; or ensuring that at least one game is played at every round in large cities.

4. Concluding remarks

OR methods have a large potential of applications in sports and are a useful strategy to motivate students in classrooms in introductory courses of optimization and simulation. Better game schedules in long competitions can significantly reduce traveling costs. There are many applications of optimization techniques to timetabling in different sports such as soccer, baseball, football, basketball, hockey, and rugby, leading to reductions in traveling costs and to more fairness.

Models for deciding playoff elimination are very useful tools for team administrators, the press, and the fans. They are quite effective to correct common misleading statements

made by the press and team managers. The **FutMax** project currently follows the 2003 edition of the Brazilian national soccer championship and the Latin America qualification round for the XVIII World Cup to be played in 2006, and is a useful tool to make several statistics available to the fans. The press is closely following the results posted in the web page of the **FutMax** project as the end of the 2003 Brazilian soccer championship approaches, since for the first time in history several among the major and traditional teams are threatened of major league elimination from next year's competition. Results obtained with analysis of the GQS and PQS scores were also used to correct wrong information given by other Internet sites that give probabilistic chances of qualification based on simulation models.

References

[1] I. Adler, A.L. Erera, D.S. Hochbaum, and E.V. Olinick, "Baseball, optimization, and the World Wide Web", *Interfaces* 32 (2002), 12-22.

[2] T. Bernholt, A. Gullich, T. Hofmeister, and N. Schmitt, "Football elimination is hard to decide under the 3-point rule", *Lecture Notes in Computer Science* 1672 (1970), 410-418.

[3] G.L. Nemhauser and M.A. Trick, "Scheduling a major college basketball conference", *Operations Research* 46 (198), 1-8.

[4] M.G.C. Resende and C.C. Ribeiro, "Greedy randomized adaptive search procedures", *Handbook of Metaheuristics* (F. Glover and G. Kochenberger, eds.), 219-249, Kluwer, 2002.

[5] C.C. Ribeiro and S. Urrutia, "An application of integer programming to playoff elimination in football championships", submitted, 2003.

[6] R.A. Russell and J.M. Leung, "Devising a cost-effective schedule for a baseball league", *Operations Research* 42 (1994), 614-625.

[7] B. Schwartz, "Possible winners in partially completed tournaments", *SIAM Review* 8 (1966), 302-308.