Operations Research Perspectives 5 (2018) 105–112

Contents lists available at ScienceDirect

Operations Research Perspectives

journal homepage: www.elsevier.com/locate/orp

A computationally intensive ranking system for paired comparison data

David Beaudoin[a,1], Tim Swartz[b,1,*]

[a] Département Opérations et Systèmes de Décision, Faculté des Sciences de l'Administration, Université Laval, Pavillon Palasis-Prince, Bureau 2439, Québec (Québec) G1V0A6, Canada

[b] Department of Statistics and Actuarial Science, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A1S6, Canada

ARTICLE INFO

Article history:
Received 26 December 2017
Revised 7 March 2018
Accepted 24 March 2018
Available online 27 March 2018

Keywords:
Nonparametric methods
NCAA basketball
Ranking
Simulated annealing
Statistical computing

ABSTRACT

In this paper, we introduce a new ranking system where the data are preferences resulting from paired comparisons. When direct preferences are missing or unclear, then preferences are determined through indirect comparisons. Given that a ranking of n subjects implies $\binom{n}{2}$ paired preferences, the resultant computational problem is the determination of an optimal ranking where the agreement between the implied preferences via the ranking and the data preferences is maximized. Comparisons are carried out via simulation studies where the proposed rankings outperform Bradley–Terry in a particular predictive comparison.

© 2018 The Authors. Published by Elsevier Ltd.
This is an open access article under the CC BY-NC-ND license.
(http://creativecommons.org/licenses/by-nc-nd/4.0/)

1. Introduction

The problem of ranking can be simply stated and has an extensive literature in the statistical sciences. Given data on n subjects, the objective is to determine a permutation (ranking) $R = (i_1, \ldots, i_n)$ where the interpretation is that subject i_j is preferable to subject i_k whenever $i_j < i_k$. The term "preferable" depends on the application and the methods used to determine the ranking depend on aspects of the data structure.

In sport, ranking is an important problem. For example, in National Collegiate Athletic Association (NCAA) basketball, there are over 300 teams competing in Division I where a typical team plays only a subset (~ 25) of the other teams during a season. At the end of the season, the NCAA Selection Committee is set with the task of creating a tournament structure known as "March Madness" involving 68 of these teams. In determining the invitees, team rankings (in terms of team quality) form part of the decision making process.

Similarly, in NCAA football, various team rankings are regularly reported during the regular season (e.g. Associated Press, FCS Coaches' Poll, Sagarin, etc.). Although such rankings are no longer used for determining Bowl bids (i.e. identifying pairs of teams that compete in prestigious holiday matches), the rankings receive considerable media attention and are available to the selection committee. Part of the intrigue involving the determination of the rankings is that there are not many crossover matches involving teams from different conferences.

Ranking also occurs in non-sporting contexts. For example, universities rank students, employers rank job candidates, there are rankings corresponding to the quality of journals, and so on. Clearly, the type of data used to inform the rankings varies greatly on the application.

In this paper, we focus on the ranking problem associated with NCAA basketball. More specifically, we consider the ranking of n Division I teams ($n = 351$ in 2015/2016). The data used to inform our ranking are the result of paired comparisons. Sometimes a comparison is explicit (e.g. based on the result of one team playing another team). In other instances, the comparison between two teams is determined by considering the results of matches involving common opponents. Our approach searches for an optimal ranking $R = (i_1, \ldots, i_n)$ which has maximal agreement between the $\binom{n}{2}$ implied paired preferences via the ranking and the observed data preferences. The approach is appealing in its simplicity and its lack of assumptions. It may be regarded as nonparametric in the sense that there is no underlying probability model. However, the

* Corresponding author.
E-mail address: tswartz@sfu.ca (T. Swartz).

[1] Both authors have been partially supported by the Natural Sciences and Engineering Research Council of Canada. This research was enabled in part by support provided by Calcul Québec (http://www.calculquebec.ca/en/) and Compute Canada (www.computecanada.ca). The authors thank three reviewers who carefully read the paper and whose comments have improved the manuscript.

https://doi.org/10.1016/j.orp.2018.03.002
2214-7160/© 2018 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license. (http://creativecommons.org/licenses/by-nc-nd/4.0/)

approach provides computational challenges. For example, a simple search amongst rankings is not possible since there are $n! \approx 10^{742}$ potential rankings.

Ranking methods based on paired comparison data originate from the work of Thurstone [26] and Bradley and Terry [4]. The approach suggested by Park and Newman [19] is most closely related to our approach in the sense that it is also nonparametric and extends comparisons to indirect matchups between teams. Park and Newman [19] rank teams according to team wins w minus team losses l in both direct and discounted indirect matches. The statistics w and l correspond to a matrix-based network centrality measure involving adjacency matrices.

From the seminal work by Thurstone [26] and Bradley and Terry [4], the statistical literature on methods for paired comparison data has flourished. For example, many extensions to the original models have been considered such as the provision for ties in paired comparisons [9], multiple comparisons [20], Bayesian implementations [6,8,17] and dynamic ranking models [12,13] which have been used in chess. The treatment of the margin of victory in paired comparison settings has also led to various models and methods. For example, Harville [14], Harville [15] considers linear models where truncations are imposed on large margins of victory. A central idea is that teams should not have an incentive for running up the score. Mease [18] considers a model based on normal likelihoods and penalty terms that attempts to correspond to human judgments. A general review of the literature related to paired comparison methods is given by Cattelan [7]. Rotou et al. [22] review methods that are primarily concerned with dynamic rankings where data are frequently generated such as in the gaming industry.

In Section 2, we describe our approach which is intuitive and simple to describe. However, the method gives rise to challenging computational hurdles for which we propose a stochastic search algorithm. For example, we demonstrate how time savings can be achieved in the calculation of our metric which measures the agreement between the implied ranking preferences and the data preferences. The algorithm implements a simulated annealing procedure which optimizes over the $n!$ candidate rankings. Section 3 assesses the proposed ranking procedure by forecasting matches based on established ranks. We first investigate our procedure in the context of real data from previous NCAA basketball seasons. We compare our rankings with rankings obtained by other popular procedures. Our second example is based on simulated NCAA basketball data where the underlying strengths of the teams are specified. This allows us to compare forecasts against the truth. The final forecasting example is based on data from the 2016/2017 English Premier League season. This is a substantially different dataset in that we have a much smaller number of teams ($n = 20$). In Section 4, we consider various nuances related to our approach. In particular, we compare our procedure to the Bradley–Terry approach where we observe the proposed method places more importance on individual matchups than Bradley-Terry. We conclude with a short discussion in Section 5.

2. Approach

Our approach is based on data arising from paired comparisons. In basketball, this represents a data reduction since each team scores a specific number of points in a game. However, sometimes the actual number of points scored can be misleading. For example, in "blowouts", teams often "empty their benches" near the end of a game, meaning that regular players are replaced by players who do not typically play in competitive matches. In such cases, margins of victory may not be representative of true quality. Interestingly, a requirement of the computer rankings used in the former BCS (Bowl Championship Series) for NCAA football was that the computer rankings should not take into account margin of victory [10].

With respect to paired comparison data, it is straightforward to determine the preference when one team plays another team in a single game. We let h denote the number of points corresponding to the home court advantage in NCAA basketball. If the home team defeats the road team by more than h points, then the home team is the preferred team in the paired comparison. Otherwise, the road team is the preferred team.

Our ranking procedure requires the specification of the home team advantage h. And since our approach does not contain a statistical model, the estimation of h must be done outside of the procedure. In NCAA basketball, Bessire [3] provided an average home team advantage of $h = 3.7$ points. The value $h = 3.7$ roughly agrees with Gandar et al. [11] who estimated a home court advantage of 4.0 points in the National Basketball Association (NBA). Since an NBA game is 48 min in duration and a college game is only 40 min in duration, the mapping from $h = 4.0$ in the NBA to college is $4.0(40/48) = 3.3$. In an independent calculation, we studied pairs of NCAA basketball teams during the 2006/2007 through 2015/2016 seasons. Suppose team A played at home and defeated team B by h_A points (h_A is negative for a loss). And similarly, suppose team B then played at home and defeated team A by h_B points (h_B is negative for a loss). In this matchup, home advantage is estimated by $(h_A + h_B)/2$, and h is obtained by averaging these terms over all matchups. In the 26,206 matchups where pairs of teams played more than once with both home and away matches, we estimated the home court advantage as $h = 3.4$ points. Therefore, although there is a range of estimates for the home team advantage h in NCAA basketball, it appears safe to assume that $h \in (3.0, 4.0)$. The reason why this is important is that the determination of binary preferences in paired comparisons (the statistic used in our ranking procedure) is insensitive to $h \in (3.0, 4.0)$. Therefore, we arbitrarily set the home team advantage at $h = 3.5$ points. When a game is played at a neutral site, then $h = 0$. Whereas there is frequent discussion of differential home team advantages for individual teams (as opposed to an overall home court advantage h for all teams), we are inclined to believe that differential advantages are primarily the manifestation of multiple comparison issues [24] and unbalanced schedules.

More generally, suppose that two teams have played each other more than once. Let p_{Ai} and p_{Bi} be the points scored by Team A and Team B respectively in the ith game. Then from Team A's perspective, define the differential

$$d_i = \begin{cases} p_{Ai} - p_{Bi} - h & \text{if Team A is the home team} \\ p_{Ai} - p_{Bi} + h & \text{if Team B is the home team} \end{cases} \tag{1}$$

In this case, Team A is the preferred team in the particular paired comparison if the average of its d_i values is positive.

When two teams have played each other directly, then we use (1) to determine the preference, and we refer to this as a level L_1 preference. With $n = 351$ NCAA basketball teams, there are $\binom{n}{2} = 61,425$ potential paired comparisons. Based on the 5948 matches that took place in 2015/2016, 3918 level L_1 preferences were observed. The level L_1 preferences represent only 6.38% of the potential $\binom{n}{2}$ paired comparisons.

We now consider cases where Team A and Team B have not directly played against each other. Our approach for determining preferences in these situations borrows on ideas from the RPI (Ratings Performance Index) where strength of schedule is considered; see [2] for a definition of RPI. Specifically, suppose that Team A and Team B have a common opponent Team C. Then (1) can be used to obtain an average differential $\bar{d}_{AC}$ from the point of view of Team A versus Team C. Similarly, (1) can be used to obtain an average differential $\bar{d}_{BC}$ from the point of view of Team B versus Team

Download English Version:

https://daneshyari.com/en/article/6899182

Download Persian Version:

https://daneshyari.com/article/6899182

Daneshyari.com