Winning Team Prediction using Weighted PageRank

Abhinav Agarwalla
Indian Institute of Technology Kharagpur
abhinavagarwalla@iitkgp.ac.in

Madhav Mantri
Indian Institute of Technology Kharagpur
madhavmantri@iitkgp.ac.in

ABSTRACT

We analyze the relative performance and rankings of international ODI cricket teams using various graph-based methods. Primarily, ranking methods on graphs are employed such as weighted PageRank. We compare the performance results derived with different weighing functions serving as edge weights of the team graph. Match summaries and team rankings scraped from open data platforms are utilized for analysing team performance. Different features like number of matches played, average runs scored per match, average number of wickets taken per match, average margin of victory, etc. are taken into consideration in deriving the weighting functions. The appropriate choice of weighing function results in improvements over naive PageRank implementations with respect to prediction accuracy.

Keywords

Cricket Team Rankings, Player Ranking, PageRank, Weighted PageRank, Weighted Graphs

1. INTRODUCTION

Cricket is one of the most followed sports in the world with a strong presence in commonwealth countries. It is played in various formats like slow moving 5-day Tests, 1-day 50-over format, and fast 20-over T-20 format, and at various levels in more than 100 countries. Cricket is a team sport in which a team of 11 players play against an opponent team. The team with greater number of runs is declared as the winner. As with any sport being played for more than 30 years, cricket has a vast collection of historical match data which are openly available online. It generates a large amount of data comprising of match outcomes, batting and bowling performances, fielding strategy, etc. However, rigorous analysis of the factors affecting match outcomes and their correlation with team performance are very scarce. There are many benefits of such an analysis to player selection committee, and betting communities. Such a study can aid them in identifying players' strengths and weaknesses against a particular team, and help selection committees select a high performing balanced team. Gamblers can use match prediction results for making an informed decision, thus mitigating risk. Efficient ranking methods can help International Cricket Council (ICC) introduce better ranking schemes.

We address the problem of ranking teams, and match outcome prediction in One Day International (ODI) format of the game. Ranking problems can be solved by 1. point-wise approach using algorithms like ordinal regression and classification algorithms, 2. pairwise approach using bipartite ranking algorithms like binary classifiers and 3. list-wise approach using algorithms like SVM map and BayesRank and link analysis mathematical algorithms like PageRank. We follow the list-wise approach because our aim is not only to find a preference function, or a binary response prediction but to get an entire ranking of all teams. In this paper we are also comparing various weighting methods for ranking cricket teams using match statistics. Therefore, the change in ranking of teams after a game is not only affected by the rankings but also on the match summary statistics of the teams playing.

Recent approaches suggested to account for these factors. Daud and Muhammad[1] devise T-index which is based on h-index. T-index is motivated by the need to account for margin of victory. They also employ graphical methods such as PageRank[3] along with its weighted and unified variants. They use 60-20-20 for weighing final result, runs and wickets without any proper justification for the numbers. Mukherjee in [2] explores PageRank further to identify best captains in Test as well as ODI formats. We extend their approach through a search for the best weighing function which results in the best prediction accuracy.

2. PAGERANK AND ITS VARIANTS

PageRank[3] was initially introduced to rank web pages, but has since then been used in recommendation engines, neuroscience, bio-informatics etc. PageRank assigns a score to each web page using a set of outgoing and incoming edges. PageRank works by counting the number and quality of links to a page to determine a rough estimate of how important the website is. It gives high scores to pages that have incoming links from pages with high link incorporating trust, and legitimacy in web. We extend the algorithm to our domain by modelling teams as pages, and directional links with winning team at the arrow head. In cricket, not only winning but winning against strong teams matters the most which

S.No.	Weighting Function
1.	$W_1 = \frac{w_{ij}}{G}$
2.	$W_2 = \frac{r_{ij}}{G}$
3.	$W_3 = \frac{m1_{ij}}{G}$
4.	$W_4 = (\frac{m1_{ij}}{G} * \frac{w1_{ij}}{G}) + (\frac{m2_{ij}}{G} * \frac{w2_{ij}}{G})$
5.	$W_5 = \alpha_1 * W_1 + \alpha_2 * W_2 + \alpha_3 * W_3 + \alpha_4 * W_4$

Table 1: Definitions of weighting functions used

is what PageRank tries to capture. If a team wins against a strong team, the score is increased heavily as compared to winning against a weaker team. Similarly, losing a weak opponent is worse than losing against a very strong one.

$$t_i = \frac{q}{N} + (1-q)\left\{\frac{t_1}{s_1} + \frac{t_2}{s_2} ... + \frac{t_N}{s_N}\right\} \quad (1)$$

where t_i is the PageRank of team i, $t_1..t_n$ are the PageRanks of teams which links to team i, s_i is the number of outbound links for the team i and $q \in [0,1]$ is the damping factor which is set to 0.15.

In a Weighted PageRank[5] setting, every link from one team A to team B has an assigned weight which is fraction of matches team A lost to team B, out of the total matches played in between them. These weights act as a proxy for the probability of losing of team A against team B.

$$t_i = (1-q)\sum_j t_j \frac{w_{ji}}{s_j} + \frac{q}{N} + \frac{1-q}{N}\sum_j t_j \delta(s_j) \quad (2)$$

where w_{ij} denotes weight assigned to the link joining team i to team j with arrow head at team j.

The weight of the link joining team i and j is determined by a weighting function that involves one or more matrices such as number of games played between a teams i and j, the number of matches won, lost and drawn games, the total runs scores or wickets taken, or the margin of victory. The various weighting functions we have tested are given in Table 1.

Within the functions we use the following notation:

- $w_{i,j}$: number of games won by team i against team j;
- $w1_{i,j}$: number of games won while chasing, out of $w_{i,j}$;
- $w2_{i,j}$: number of games won while defending, out of $w_{i,j}$;
- $r_{i,j}$: number of runs scored by team i against team j;
- $m1_{i,j}$: mean margin of victory (by runs) of team i against team j;

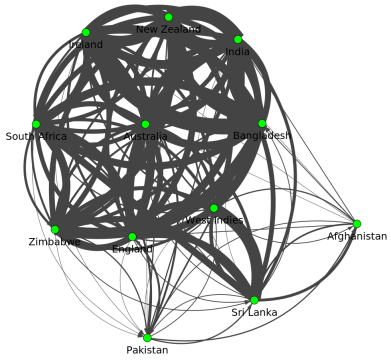

Figure 1: A weighted graph with green nodes as teams.

- $m2_{i,j}$: mean margin of victory (by wickets) of team i against team j;
- G : number of games played between the two teams.

3. RESULTS AND DISCUSSIONS

We crawl the final results for international ODI cricket matches from http://cricinfo.com. The match statistics includes: 1. venue of the match 2. the date and time of the game 3. teams playing the match 4. toss results 5. the playing 11 players of both teams 6. match results including the margin of victory.

S.No.	Weighting Function	Accuracy
1.	W_1	0.6346
2.	W_2	0.6993
3.	W_3	0.6346
4.	W_4	0.6130
5.	W_5	0.6873

Table 2: Test set accuracies of weighting function used

Directed graphs are constructed using weighted adjacency matrices after every successive match and the page ranks of all teams were calculated using power method and saved.Table 2 reports the test set accuracies achieved with different weighing functions. Parameters $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ are optimised using Bayesian optimisation[4] which led to the best accuracy of 0.6873. Fitting arbitrary weighing functions or a machine learning model which can better approximate the probability of winning is left to future work.

4. REFERENCES

[1] A. Daud and F. Muhammad. Ranking cricket teams through runs and wickets. In *International Conference on Active Media Technology*, pages 156–165. Springer, 2013.

[2] S. Mukherjee. Identifying the greatest team and captainâĂŤa complex network approach to cricket matches. *Physica A: Statistical Mechanics and its Applications*, 391(23):6066–6076, 2012.

[3] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: bringing order to the web. 1999.

[4] J. Snoek, H. Larochelle, and R. P. Adams. Practical bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems*, pages 2951–2959, 2012.

[5] W. Xing and A. Ghorbani. Weighted pagerank algorithm. In *Communication Networks and Services Research, 2004. Proceedings. Second Annual Conference on*, pages 305–314. IEEE, 2004.