Predicting the final seeds of National Basketball Association teams, an Elo based approach

Thomas Abraham
210639757
Hazar Emre Tez
MSc Computer Science, QMUL

***Abstract*—The introduction of statistical analysis into the National Basketball Association has radically changed the way basketball shot selection and plays are thought about. The Elo algorithm is a popular ranking algorithm used to rank players in a competitive setting. The Elo algorithm is often associated with chess as it is used extensively by FIDE (the governing body of international chess) to determine world rankings of chess players. For the model being considered here, the Elo rating will depend on the players +/- for each game. The model will include both regular season and playoff games. Teams will gain points after winning matches and lose points post defeat. More points are awarded for blowout wins and upsets. Finally, teams will be seeded from one to fifteen in each conference based on their Elo scores. The generated seeds will be validated by creating the model for a previous season and cross-checking its accuracy with real world results.**

Keywords—plays, Elo algorithm, playoffs, home-court advantage, blowout, seeded.

I. INTRODUCTION

The National Basketball Association (NBA) is a men's professional basketball league in North America, composed of thirty teams divided into western and eastern conferences. Each NBA team has a maximum of fifteen players, out of which thirteen are allowed to be active in each game. Players on a basketball court position themselves in five locations as shown in Figure 1.1. Each of these positions require distinct abilities and physical attributes. These positions are divided into three main categories.[1]

Centre. Centres are usually the tallest players on the team, and they position themselves near the basket. On the offensive side, the centres goal is to make high percentage shots close to the basket. On the defensive end, the centre is responsible for shot blocking and cutting off internal passing lanes. They are also expected to attempt to retrieve defensive and offensive rebounds.

Forward. Forwards are usually the next tallest players on the court and are responsible for passing to cutting players, taking shots, retrieving rebounds in certain instances, and driving to the basket.

Guard. These are likely your shortest players and are responsible for bringing the ball up the court and setting up offensive plays, driving to the basket and long-range shooting.

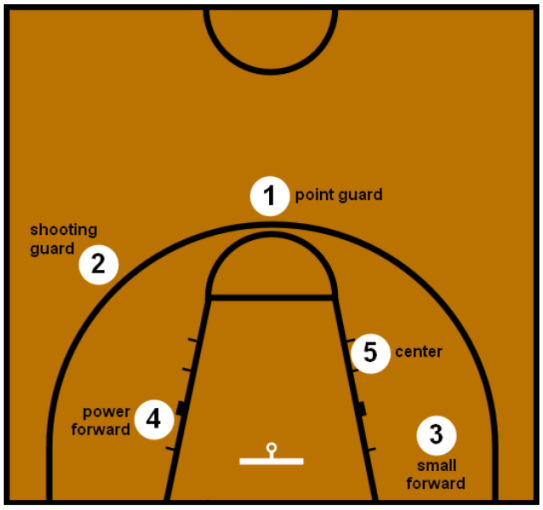

Figure 1.1. Traditional positions in basketball

At the highest level of professional basketball, every play is executed by a team is carefully devised by a team of coaches. Basic plays such as a switch or a give-and-go are widely used and can be situational. However, particular plays such as those centred around off ball movement in order to ensure that a particular player loses their defender are drawn up after taking into consideration the player as well as the defenders ability. The ultimate goal of the offense is to shoot the ball, the strategy lies in devising plays to create good shot opportunities. Defensive plays on the other hand are often drawn up to restrict the movement and ability of one particular player or a set of players. These types of plays are often drawn up after watching a particular player over multiple games and analysing their offensive and defensive capabilities.

General managers such as Darryl Morey set out to prove that data driven decisions would result in a competitive edge. Previously implemented conventional methods used to simulate gameplay and deduce plays have ignored that in a sport such as basketball the dynamics of movement and cohesiveness are unique from line-up to line-up and do not depend solely on individual offensive and defensive ability.

The Elo rating system is popular and widely used, this is mainly because it is elegant yet simple in its execution. A rating system analyses the outcomes of matches and assigns a value to the strength of a team relative to others. This information then allows teams and coaches to make beneficial long term decisions regarding team formation and drafts [4].

[1] A raw version of the data mentioned is available at https://www.myactivesg.com/Sports/Basketball/How-To-Play/Basketball-Rules/Basketball-Positions-and-Roles

II. Literature Review

Accurate predictions of a team's seed greatly influences a team's chances of winning prior to the playoffs. The ability to discern a potential round one matchup at the least will give coaching staff the opportunity to condition players and design game plans specifically suited to their opponents.

Building on previous work [4], this project sets out to create a model that can accurately predict each of the fifteen seeds in both conferences for a season. An Elo based approach is employed to obtain individual player ratings. The individual player ratings are then combined to obtain a team rating. Team ratings are then compared pairwise to obtain the probability of a win by each of the teams during the season. The probable wins are then cumulated, and a seed is calculated for each team. The rating system is validated by running them over real-life data from previous NBA seasons.

III. Methodology

The original purpose of the Elo algorithm was to develop a viable rating system for chess players. As the popularity of the algorithm increased, analysts and statisticians began modifying the algorithm so it could be applied to various other sports. In the simulations being considered here, players alone are not given individual attention when it comes to win prediction. Instead, the team is considered as a single entity.

The true value of a player is not specifically quantifiable and therefore, cannot be measured and analysed. Hence, we depend on the observable metrics of the sport such as points scored, rebounds, assists and so on. The primary statistic being considered in the algorithm is the plus-minus score. The algorithm as a whole is designed to track the performance of individual basketball players and combine their ratings in order to obtain a team score which can then be used in the simulations.

A. Plus-Minus Score (+/-)

The Plus-Minus score reflects how the team did while a particular player was on the court.

B. Elo Approach

1) Player Strength

2) Estimate

In the algorithm, it is assumed that each team's actual strength is derived from a normally distributed random variable, with the team's actual strength being represented by the mean. A team that maintains the same lineup every game should perform at the same strength. Due to this reason a normal distribution is chosen. The rating of a team is updated continuously based on observed wins and losses. If $Team_i$ plays, $Team_j$ then the rating is updated as:

$$R_{i_{new}} = R_{i_{old}} + K(S_{ij} - x_{ij}) \qquad [3]$$

Where *R* refers to the rating, *K* refers to the K factor, *S* refers to the actual score and *x* refers to the expected score.

3) Actual Score (S)

The definition of S_{ij} is depicted as:

$$S_{ij} = \begin{cases} 1, if\ Team_i\ beats\ Team_j \\ 0, if\ Team_j\ beats\ Team_i \end{cases} \qquad [3]$$

4) Expected Score (x)

Variable x_{ij} is used to denote the expected outcome of a match between $Team_i$ and $Team_j$.

5) K Factor (K)

The K-factor determines how quickly the rating reacts to new game results [9]. A high K value allows the estimate to adapt quickly, however if K is set to high it will result in the large variations in the estimate. On the other hand, if the K value is set too low then the estimate will take too long to recognize important changes. The K factor being selected here depends on the total number of minutes being played by each individual player.

6) F(x)

7) Match Outcome

8) Seed Outcome

C. Algorithms

1) Algorithm 1: Update Team Rating

2) Algorithm 2: Update Player Ratings

3) Algorithm 3: Predict Match Winner

D. Datasets

1) Physical Data:

The required datasets are obtained by scraping data from the NBA website. The data is obtained from the 2017-2018, 2018-2019 and 2020-2021 regular seasons. Both the player box score and the team box score are recorded.

2) Data Scraping

Data or web scraping refers to the process of importing information from a web page, typically written in HTML or XHTML, into a locally saved spreadsheet. A Python program was written to extract the required statistics from online tables and is stored locally as a CSV file.

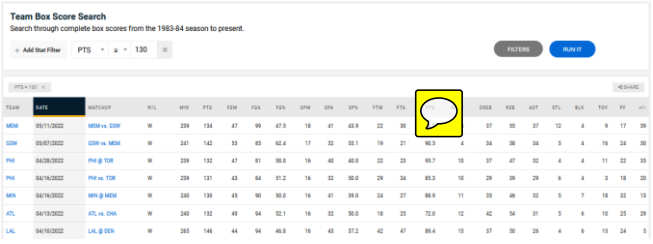

Fig 3.1 Website containing NBA statistics, reprint from [8]

TEAM	DATE	MATCHUP	W/L	MIN	PTS	+/-
MEM	05/11/2022	MEM vs. GSW	W	239	134	39
GSW	05/07/2022	GSW vs. MEM	W	241	142	30
PHI	04/28/2022	PHI @ TOR	W	239	132	35
PHI	04/16/2022	PHI vs. TOR	W	239	131	20
MIN	04/16/2022	MIN @ MEM	W	240	130	13
ATL	04/13/2022	ATL vs. CHA	W	240	132	29
LAL	04/10/2022	LAL @ DEN	W	265	146	5
CLE	04/10/2022	CLE vs. MIL	W	240	133	18

Fig 3.2 Scraped data stored in a CSV file

3) Synthetic Data

4) Initialization

5) Metric Calculation

IV. Results

A. Testing on the NBA 2018-19 Season

Algorithms 1 and 3, as depicted above, are used to predict the outcome of a match for the base Elo algorithm and the modified algorithm, respectively. During the testing phase the ratings of both players and teams are updated repeatedly even as results are predicted.

V. Conclusion

The Elo algorithm is a widely used rating system due to its simplicity and the fact that it offers relatively high prediction accuracy. The main reason it was selected is because the algorithm considers the whole team as a fundamental unit. In the models being discussed above, a modified version of the Elo algorithm is used where a team's performance is modelled using the +/- metric of individual players. Individual player ratings are combined to obtain a team rating which is then used to predict the outcome of matches. The chosen algorithm is compared with the base Elo algorithm by analyzing the implementation of the algorithms over real-world and synthetic data. From the above simulations it is observed that the base algorithm is more accurate than its modified version. This is largely due to the fact that the addition of individual player ratings increases the complexity of the algorithm. The model does however offer additional insight into player strengths which further enable the analysis of injuries and possible transfers. This particular factor should allow teams to better quantify a players worth and overall trade value.

VI. Future Work

Analysis of test results indicate that further modifications to the algorithm may result in further understanding of individual player strengths and susceptibility to injuries. This may be able to aid front office management when it comes to player transfers and contract extensions.

References

[1] Franks, Alexander M., D'Amour, Alexander, Cervone, Daniel and Bornn, Luke. "Meta-analytics: tools for understanding the statistical properties of sports metrics" Journal of Quantitative Analysis in Sports, vol. 12, no. 4, 2016, pp. 151-165. DOI: https://doi.org/10.1515/jqas-2016-0098.

[2] Elo, Arpad E. "The Rating of Chessplayers, Past and Present. New York: Arco Pub, " 1978. Print.

[3] Hao Ji, Erich O'Saben, Adam Boudion, and Yaohang Li. "March Madness Prediction : A Matrix Completion Approach" 2015.

[4] Marveldoss, Richard Einstein Doss. "An Elo-Based Approach to Model Team Players and Predict the Outcome of Games, " August 2018 unpublished. URI: https://hdl.handle.net/1969.1/173956.

[5] Song Yan, Siyuan Meng, Qiwei Liu, Jing Li. "Design and Implementation of NBA Playoff Prediction Method Based on Elo Algorithm and Graph Database, " November 2019. ISSN 23275219. DOI: https://doi.org/10.4236/jcc.2019.711004.

[6] Raquel YS Aoki, Renato M Assuncao, Pedro OS Vaz de MElo. "Luck is Hard to Beat: The Difficulty of Sports Prediction" June 2017. DOI: https://doi.org/10.48550/arXiv.1706.02447.

[7] Keshri, Suraj Kumar. "Essays in Basketball Analytics, " September 2019, unpublished. DOI: https://doi.org/10.7916/d8-1ghx-zy51.

[8] National Basketball Association (2022).Player Box Score Search [online]. Availabile at: https://www.nba.com/stats/search/player-game/?CF=PTS*gt*40 [Accessed June 2022].

[9] Five Thirty Eight (2015). How We Calculate NBA Elo Ratings [online]. Available at: https://fivethirtyeight.com/features/how-we-calculate-nba-elo-ratings/ [Accessed June 2022].

[10]

[11]