
A Game Theoretic Approach to Modelling Basketball

Darshan Chakrabarti and Rishav Dutta
(darshanc and rishavd)

1 Overview

Each of the 30 teams in the NBA play 82 games a season (and then top teams participate in the playoffs which progress towards the championship). Typically, coaches and other team personnel analyze tapes of past games and statistics to try to understand deficiencies that a team has and understand vulnerable points of the other team. We want to make suggestions for strategies for teams to adopt that can be backed quantitatively, so that teams can try to achieve the optimal outcome.

2 Problem Statement

Given the current game state, what is the optimal strategy for the offensive team and the defensive team to adopt. For example, consider a matchup between the Houston Rockets and the Golden State Warriors with their starting line-ups on the court: James Harden, Chris Paul, Nene, P.J. Tucker, Clint Capela for the Rockets and Stephen Curry, Klay Thompson, Kevin Durant, Draymond Green, and DeMarcus Cousins for the Warriors, and James Harden has the ball. What is the optimal strategy for the Rockets? Potential options could be to pass to Chris Paul for a three pointer or Harden could just shoot himself. In order to minimize the potential outcome, the Warriors may choose to assign their best defender to the player most likely to shoot from the Rockets (Klay Thompson on Chris Paul) or they could play aggressive defense on James Harden to prevent him from passing or shooting. Finding the optimal strategy may allow teams to perform better during games. Furthermore, we choose to find the optimal strategy (we use Nash equilibrium as a starting point) that is reasonable and realistic for players to remember i.e. there aren't too many options for either team (at most 5 options). In this way, our model could provide teams with a real world plan for games.

3 Related Work

Basketball has been modelled as zero-sum game in previous work to analyze the most efficient way to foul, and whether a 2 point shot or 3 point shot should be taken in late game situations requiring clutch (consistent, timely, and well-executed) plays [10].

Other work done in modelling of zero-sum games includes association football in which team formations are used to determine probabilities of conceding and scoring a goal and then modelling tactical changes as moves in the zero-sum game [7].

Some of the recent work that has been done in basketball analytics has been exploring better ways to evaluate defensive movement [6]. Through the use of hidden Markov models and a prior on the position of the defenders relative to the ball and the offensive player, Franks et al. determined an efficient way to identify switches and screens on defense. Using this work, they were able to determine if switching or staying was the better defensive strategy for certain players. Some work done on the offensive side is viewing offensive movement as a network and demonstrating that taking high percentage shots aren't synonymous with having the most efficient offensive game [11]. In this paper, the authors consider the empirical fact that shooting percentage generally goes down for a player when they take a larger fraction of shots for their team. They use the argument that if a play is run too often then the defensive team can easily adjust their strategy to shut down subsequent plays

(which might explain this empirical dropoff in efficiency). These ideas could be incorporated into our model in future work.

4 Data

The data that we work with is taken from the already available NBA stats on the official NBA website. We only consider data from the 2017-18 NBA season. We access this data through a Python API called *nba_api* that allows one to access various endpoints on the NBA website. These endpoints are locations of specific player stats such as individual shooting stats or individual defensive stats.

In this paper we use the individual player data in order to construct the expected points (entry in the game matrix). We first collect individual player shooting percentages from each of the 7 zones as defined by the NBA: Left Corner 3, Right Corner 3, Midrange, Above the Break 3, Restricted Area, In the Paint (non-RA), and Backcourt. We then collect each players defensive field goal percentage from each of the same zones. A defensive field goal attempt for player a is a shot taken by player b with player a as the primary defender.

We also gather player passing data. We use this data in order to determine difficulties of passing within certain zones.

The final piece of NBA data that we use is the player vs. player metric. The player vs. player metric compares the performance of two player against one another. Given a player a and a player b the player vs. player metric reveals player a 's offensive performance from each of the 7 zones above against player b 's defense. Thus, we get the field goal percentage that a has against b in a certain zone in addition to getting the defensive field goal percentage of player b against player a .

Through each of these pieces of data taken from the collected NBA stats, we are able to model mean player performance for our game from real player performance.

5 Game Theory Model

5.1 Basketball Model

We model basketball as a zero-sum (one-shot) game with a set of offensive actions and a set of defensive actions, each potentially performed by each team. The game state consists of an offensive team (of five players), defensive team (of five players), and a ball handler.

We define the possible offensive actions as some sort of pass and shoot or dribble and shoot. The ball handler is the only player allowed to dribble the ball, any player who is passed to must shoot from the zone they are in. Each offensive possession starts with the ball handler in the zone called Above the Break which is at the top of the 3 point arc straight on from the basket. From this point, the ball handler can chose to shoot from this zone at no cost or dribble to another zone and shoot. Dribbling to each zone incurs a cost for the ball handler, and dribbling to certain zones is more expensive than others (for example dribbling into the paint is rather difficult, but dribbling around the 3 point arc is not). If the ball handler chooses not to dribble, he must pass to another player in a different zone. It costs nothing for a player without the ball to relocate to different zones, but passing to certain zones is more expensive (passing into the paint is more difficult than passing along the 3 point arc much of the time). Since it is not possible for the offensive to dribble or pass into the backcourt, we choose to assign a utility of 0 to this zone. Hence, with the offensive model that we have described, there are 7 zones for the ball handler to dribble into and shoot from, and 4 other team mates to pass to in 7 zones for a total of 35 possible moves for the offense to make.

We define the possible defensive actions as a choice of combination of a matchup and aggression level. The defense can decide which of each guards which offensive player. We chose to allow only guards to defend other guards and frontcourt players to defend other frontcourt players, i.e. it is not possible for the defense to pick a match up in which a guard is defending a forward as this is generally optimal. Furthermore, for each match up, the defense assigns an aggression level for each defender. In our model, this aggression level is binary; either the defender is aggressive or not. In the case that the defender is aggressive, it is easier for the offensive player to dribble past them, but harder to shoot and pass, but a less aggressive defender would allow an easier shot. Furthermore, passing to different areas of the court becomes harder based on the number of players that are aggressive on

the defense. If there are no aggressive players then passing into the restricted area becomes harder because there are more players inside the arc, but if there are only aggressive players then passing into the restricted area becomes easier as most of the defenders are on the outside. Thus our chosen model of the defense has to decide which players to be aggressive on and how many players need to be aggressive for the optimal strategy. With this model, there are 2^5 possible aggression levels for each matchup. Since only guards can defend other guards there are only 2 ways to assign that match up and, and 6 ways to assign the front court. Thus there are $12 \cdot 32 = 384$ total possible defensive strategies.

5.2 Game Matrix

Now that we have considered each of the moves for both players in the game, we create each entry in the game matrix by looking at the offensive move O and the defensive move D . The offensive move has 3 cases:

1. The ball handler shoots from above the break without passing or dribbling
2. The ball handler dribbles to another zone and shoots from that zone
3. The ball handler passes to another player and that player shoots from their zone

The value in this entry is calculated in the following way:

First we calculate the following p value. This is our heuristic designed to capture how the defensive matchup affects the probability of success (we encapsulate offensive player general shooting percentage, offensive player shooting percentage while defensive player on court, and defensive player shooting percentage).

$$p(s) = 0.5 \left(\frac{(\text{Overall FG Made}) + (\text{FGMDC} \cdot \text{FGADC})}{(\text{Overall FG Attempted}) + (\text{FGADC})^2} \right) + \text{DFGRatio}$$

Where FGMDC: field goals made with defender on the court, FGADC: field goals attempted with Defender on the court. Both are taken from the player vs. player metric, and DFGRatio: defensive field goal percentage for defender.

Then we determine the expected number of points from the zone the shot is from depending on the number of players aggressive on the defense:

If a 2 point zone (not Mid-Range) :

$$E[\text{points from zone}] = 1.8 + 0.1 \cdot (\text{number of aggressive players})$$

If the zone is Mid-Range:

$$E[\text{points from zone}] = 2$$

If the zone is a three point zone:

$$E[\text{points from zone}] = 3.225 - 0.075 \cdot (\text{number of aggressive players})$$

Next, we determine the cost to perform the action. If the action is to dribble and shoot the cost is:

$$\text{cost of action} = \begin{cases} 0.45 & \text{Restricted Area} \\ 0.5 & \text{In the Paint (non-RA)} \\ 0.6 & \text{Mid-Range} \\ 0.7 & \text{Left Corner 3} \\ 0.7 & \text{Right Corner 3} \\ 0.7 & \text{Above the Break 3} \end{cases}$$

If the action is to pass and shoot from the zone, the cost is:

$$\text{cost of action} = \begin{cases} 0.55 & \text{Restricted Area} \\ 0.6 & \text{In the Paint (non-RA)} \\ 0.7 & \text{Mid-Range} \\ 0.8 & \text{Left Corner 3} \\ 0.8 & \text{Right Corner 3} \\ 0.8 & \text{Above the Break 3} \end{cases}$$

We then fill the entry of the game matrix with the value (if the corresponding defender is aggressive):

$$v = \frac{p(s)}{1.5} \cdot E[\text{points from zone}] \cdot (\text{cost of action})$$

If the corresponding defender is not aggressive:

$$v = p(s) \cdot E[\text{points from zone}] \cdot (\text{cost of action})$$

The values we chose for the costs reflect the difficulty of passing into that zone relative to others and also the variation in this difficulty based on the number of aggressive defenders. We can calculate this value by looking at the passing success rates from zones. We also choose the specific calculation of $p(s)$ in order to weight a specific defenders ability against a shooter and the shooters own ability to make a shot from that zone.

5.3 Determining Moves

Given an offensive team, a defensive team, and a ball handler, we construct the matrix as described as above with the relative expected values as entries. Then we add Gaussian noise to the matrix to account for the fact some entries in the matrix are very close to one another, and it doesn't make sense to dominantly pick one strategy over another if there is only a negligible in expected points. In basketball, variation in strategy better encapsulates features of the sport as a whole that are hard to model (player shooting slumps, stadium atmosphere, etc). Thus by adding noise, we make sure there is a variety in the strategy without too much loss in utility. We can then calculate the Nash Equilibrium of the zero sum game using a LP where x is the output of the mixed strategy (we can do this twice to find strategies for both the offensive and defensive teams).

$$\begin{aligned} \max v \\ ve - Ax &\leq 0 \\ e^T x &= 1 \\ x &\geq 0 \end{aligned}$$

We perform average the strategies from the Nash equilibria over a 100 iterations of generating noisy payoff matrices to calculate the optimal strategy. Then, we take the top 4 strategies and then re-normalize the probabilities. We do this to make sure there is a realistic plan for every play rather than potentially having too many for players to remember.

6 Results

We have been able to use the data to construct a matrix of payoffs for a given set of 5 offensive players and 5 defensive players. Based on this matrix, we solve for optimal strategies, accounting for noise (as described above).

We tested the model on the example of the Golden State Warriors vs. the Houston Rockets and the outcome gave the offense the following suggestion:

1. Stephen Curry Above the Break 3
2. Pass to Kevin Durant in the Paint (non-RA)
3. Pass to Kevin Durant Mid-range
4. Stephen Curry Mid-range

Looking at the suggestion, it is interesting to note that Stephen Curry is the ball handler, and the plays that are suggested are a shot himself or a pass to Kevin Durant. The 4 options that the algorithm outputs actually describes a 'Pick and Roll' a common basketball play in which the ball handler either shoots or passes to the player setting the pick. In the 2017-18 season, this play was rarely run by the two players, however, was considered to be the most lethal play combination in the NBA [1, 5]. Our model predicts exactly that.

We further tested the model on an example with the Houston Rockets against the Golden State Warriors. The outcome gave the offense the following suggestion:

1. James Harden Above the Break 3
2. Pass to Chris Paul Right Corner 3
3. Pass to Clint Capela Restricted Area
4. Pass to Trevor Ariza Left Corner 3

The outcome gave the defense the best suggestion:

- Klay Thompson Guarding James Harden extremely aggressively
- Stephen Curry Guarding Chris Paul extremely aggressively
- Kevin Durant Guarding Trevor Ariza less aggressively
- Draymond Green Guarding Clint Capela less aggressively
- Andre Iguodala Guarding PJ Tucker less aggressively

Looking at the suggestion for the offense, we see that James Harden is recommended to shoot from Above the Break or a pass to Chris Paul or Trevor Ariza for a three pointer. The Rockets are a volume three point shooting team and a majority of their points comes from behind the arc [3, 4]. Furthermore, we see that since this is the case, the defensive team is suggested to guard relatively aggressively, and the defensive teams best defensive guard (Klay Thompson) is on the best player on the offense (James Harden) [8]. Furthermore, since Draymond Green and Kevin Durant (both are good defenders) are guarding less aggressively (which they are known for doing well), it would make driving into the lane (one of Harden’s most common actions) difficult, and thus a pass to Clint Capela for an easy dunk in the restricted area is also suggested [12, 9]. Hence the suggestions for both the offense and the defense make sense for these two teams.

Finally, we took a look at the Cleveland Cavaliers vs. the Golden State Warriors. In this example the suggestions for the offense are:

1. LeBron James in the Paint (non-RA)
2. LeBron James Mid-Range
3. LeBron James Above the Break 3
4. Kevin Love Left Corner 3

Looking at the offensive strategies in this final case, we see that the strategy for the Cavaliers seems to be to give LeBron James the ball as often as possible, which makes sense as he is an all time great player, and the rest of his team is below average [2].

In general, it is difficult to validate game theoretic models without implementation in the real-world of our strategies. Note that our Nash equilibrium generated strategy will on average be the best response to any other strategy employed by the team (even accounting for noise).

7 Future Work

One of the key issues with the analysis done in our model is that we are considering expected point changes in the point differential. It is important to remember that it doesn’t matter if games are won by 1 point or by 30 so often times the best strategy is not run plays that give you the best outcome for the current play (this is too myopic). To this end, if we were able to formulate this as a repeated game, or in a large game tree which represented the entire game instead of a single play, we might have a better model. Obviously, the problem with a more complicated model is a lack of scalability; our model is quite scalable and allows for quick and efficient computation.

Another key issue is the lack of sophistication of our model of a play. During a play, basketball players typically pass multiple times, and defensive matchups are dynamic as well. The difficulty with modelling this as an extensive form game where the teams take turns making decisions is that the problem quickly becomes intractable. Although, we could try use algorithms such as MTCS or CFR minimization. However, we speculate that using a more advanced model wouldn’t necessarily lead to more insight because there are so many variables that evolve over the course of the game (and it would be nearly impossible to capture all of these in a sensible manner).

References

- [1] Apricot. Explain one play: the fabled curry-durant pick and roll, May 2017.
- [2] Les Carpenter. One-man army lebron james is playing a different game than the rest of the nba, May 2018.
- [3] Tim Cato. The rockets are shooting more 3s than should be humanly possible, and it's working, Nov 2017.
- [4] Kelly Cohen. The houston rockets love shooting threes in the playoffs, Apr 2019.
- [5] Scott Davis. Steve kerr explained why the warriors didn't go to one of their greatest weapons until they needed it most, Jun 2017.
- [6] Alexander Franks, Andrew Miller, Luke Bornn, and Kirk Goldsberry. Characterizing the spatial structure of defensive skill in professional basketball. *The Annals of Applied Statistics*, 9(1):94–121, May 2015.
- [7] Nobuyoshi Hirotsu and Mike Wright. Modeling tactical changes of formation in association football as a zero-sum game. *Journal of Quantitative Analysis in Sports*, 2:4–4, 02 2007.
- [8] Scott Howard-Cooper. 'splash brother' klay thompson can win games with his defense, Jun 2017.
- [9] Mike Prada. Why the nba's premier defender is at his best when guarding nobody, Apr 2017.
- [10] Shawn Ruminski. Game theory applied to basketball by shawn ruminski.
- [11] Brian Skinner. The price of anarchy in basketball. *Journal of Quantitative Analysis in Sports*, 6(1), 2010.
- [12] Jonathan Tjarks. Kevin durant's defensive player of the year candidacy isn't perfect, but it's real, Feb 2018.