A PROPOSED MODEL FOR ANALYZING BASKETBALL LINEUP MATCHUPS

ZOHAIB SYED

Abstract. Current methods of predicting the outcome of NBA games rely on but are not limited to factors such as win/loss rates, point differentials, and home/away status. This paper briefly looks into an alternative method of predicting the outcome of games. The greater focus of this paper however is to propose an alternative method of optimizing player rotation based on some analyses.

The purpose of this paper is to propose a hypothetical model for optimizing NBA lineups based on the opposing team (for all purposes, this model will remain broken until a reliable source of certain data, later described, is published.) The idea behind this model is analyzing the effect of individual players and their opponents.

Contents

1. The Data Required 1
2. Remarks about the Required Data 2
3. Calculating Counter Efficiency Between Players for an Individual Game 2
4. The Counter Efficiency Matrix 4
5. Who is this all really relevant to? 5
6. Limitations 6
7. Conclusion 6

1. The Data Required

Suppose we wish to optimize player rotation for a particular basketball game. The data required for our model consists of the following:

1) Average Points per Game Data of Both Teams
2) The Counter Efficiency Matrix of Both Teams
   - $\star$) Counter Efficiency Data of Players on either Team (with respect to one another)
     - a) Average Points per Minute Data of the Players on Both Teams
       - i) Average Points per Game Data of the Players on Both Teams
       - ii) Average Minutes Played per Game Data of the Players on Both Teams
     - b) Play by Play Data of Every Game Played between Each Pair of Players
     - c) Records Outlining When Each Player was Substituted on Both Teams

Date: May 21st, 2018.

of Every Game between Each Player

The first three are widely available while the fourth may be calculated through the use of i) and ii).

2. Remarks about the Required Data

The Counter Efficiency Matrix in 2) may be found using $\star$. We also have that a) may be obtained from i) and ii). Also note that neither b) nor c) are currently reliably available from any source (that I have managed to find.)

b) Play by Play data generally contains enough omissions to imply absurdities to the reader (i.e. by strictly following the Play by Play data, we may erroneously find times where more than 5 people from the same team are on the court. Other times, shots that are made are not reported.) Lately, the Play by Plays seem to contain less of these kind of errors but they do still exist. It should also be noted, Play by Play data seems to be more accurately posted when it pertains to starter players than it does for the bench players.

For the sake of this paper, in the later example to show counter efficiency, we will use Play by Play data from ESPN.com.

Note that if Play by Play data is one day fully reliable, we can use it to obtain c) as well.

c) Comprehensive data on substitutions isn't publicly released outside what is shown in the Play by Play (see above). Basketball-Reference.com seems to have it, however, instead of posting specific times when substitutions were made, they seem to post an image of the block of time played for each player.

For the sake of this paper, in the later example to show counter efficiency, we will use the graphics from Basketball-Reference.com to estimate when players were substituted in and out.

3. Calculating Counter Efficiency Between Players for an Individual Game

For this example, we will measure the counter efficiency between Kyle Lowry (Toronto Raptors) and Giannis Antetokounmpo (Milwaukee Bucks) for a game played on February 23rd, 2018.

We start by calculating item a). There is no surprise here. We simply take the values from item i) and divide by the values from item ii).
In this case we would see (from Basketball-Reference.com) that Giannis averaged 26.9 points per game in the 2018 season (normally we would use however many points per game they were averaging at that specific point in the season, or if the season was still fresh we could use previous season data.) Also from Basketball-Reference.com, we see that he was averaging 36.7 minutes played per game (again we would be using current/previous season data depending on how deep into season it is).

Then naturally we estimate Giannis averaging $\frac{26.9}{36.7} \approx 0.733$ points per minute of each game.

For Kyle Lowry we have (from Basketball-Reference.com) that he averaged 16.2 points per game in the 2018 season and 32.2 minutes per game.

Then naturally we estimate Kyle averaging $\frac{16.2}{32.2} \approx 0.503$ points per game.

Next we examine item c). We check to see at what points in the game Kyle Lowry and Giannis Antetokounmpo were simultaneously on the court.
For this part we examined the graphics representing both players being substituted in and out from www.Basketball-Reference.com.

Since no points were given for the graphic, the methodology used to estimate here was right clicking on the the different parts and selecting inspect element to measure the individual pixel widths. Then by taking the sum of the pixel width we find the pixel width of the entire game and can work for there to estimate substitution times.

Peculiarly, the graph for Lowry and the graph for Giannis had different pixel lengths (only by ≈ 0.0013 however.) To account for this, the times for Lowry were multiplied by a small adjustment constant.

The intersecting playtime then came out with the following interval (in minutes):

$$[0, 6.49262] \cup [17.5666, 17.6079] \cup [18.2785, 29.8876] \cup [34.369, 35.9261] \cup [42.3573, 53]$$

We shall denote it for simplicity as

$$I_1 \cup I_2 \cup I_3 \cup I_4 \cup I_5$$

(Note that this game went into overtime and two of the intervals occur during more than one quarter.)

Now we check item b) (the Play by Play data) to see how many points were scored in each of the intervals by Kyle Lowry and Giannis Antetokounmpo.

Player	I_1	I_2	I_3	I_4	I_5
Kyle	5	0	4	0	3
Giannis	2	0	10	2	10

Then the total number of points scored by Kyle this game amounts to 12 and the total number of points scored by Giannis amounts to 24.

The total amount of intersecting time they played is 31.2 minutes.

Then for this game we have that Giannis Antetokounmpo averaged $\frac{24}{31.2} \approx 0.769$ points per minute.

We also have for this game that Kyle Lowry averaged $\frac{12}{31.2} \approx 0.385$ points per minute.

We calculate that Giannis Antetokounmpo has a counter efficiency rating (with respect to scoring) against Kyle Lowry in this game of $\frac{0.769}{0.733} \approx 1.049$.

Conversely, we calculate that Kyle Lowry has a counter efficiency rating (with respect to scoring) against Giannis Antetokounmpo in this game of $\frac{0.385}{0.503} \approx 0.765$.

Now suppose that we had the same two counter efficiency ratings via averaging it out over many games.

Essentially, this would tell us that assuming Giannis Antetokounmpo and Kyle Lowry are on the court with four other random players on either side, we would expect Giannis to perform roughly as well as he normally does and Kyle would perform slightly worse than he normally does (a counter efficiency rating of 1 would imply they would perform exactly as well as they normally do.)

The significance of this is that we can use the same ideology to not just calculate the proficiency of the two players based on each other, but based on each others complete lineups on the court.

4. The Counter Efficiency Matrix

Here is where the analysis gets interesting. Suppose we have the counter efficiency data calculated for the players on the court from both teams. We may use this to predict the outcome of the lineup.

Suppose we are given the following counterefficiency ratings for the starting lineup of team α versus team β.

$$C = \begin{bmatrix} 0.853 & 0.938 & 1.182 & 0.788 & 1.211 \\ 0.893 & 1.102 & 0.784 & 1.209 & 0.982 \\ 0.698 & 0.873 & 0.872 & 0.891 & 1.102 \\ 0.987 & 1.231 & 0.892 & 0.798 & 0.894 \\ 1.107 & 0.987 & 0.878 & 0.875 & 0.989 \end{bmatrix}$$

C_{11} represents player 1 from team α's counterefficiency against player 1 from team β.

C_{12} represents player 1 from team α's counterefficiency against player 2 from team β.

As you can see, each row represents a member of team α's starting lineup and each column represents a member of team β's starting lineup.

What we do next is simply take the average of each row and multiply it by that player's points per minute multiplied by the number of minutes before the first substitution.

For the purposes of our example, suppose the first substitution occurs after exactly 5 minutes of game time and the average points per minute of team α is given by the following:

Player 1 points per minute $= 0.583$.
Player 2 points per minute $= 0.812$.
Player 3 points per minute $= 0.763$.
Player 4 points per minute $= 0.371$.
Player 5 points per minute $= 0.614$.

Refer back to our counterefficiency matrix C.

The average counter efficiency of player 1 versus team β's starters is simply given by $\frac{0.853+0.938+1.182+0.788+1.211}{5} \approx 0.994$.

Using the same idea, we find that for player 2 we have 0.994, for player 3 we have 0.887, for player 4 we have 0.960, and for player 5 we have 0.967.

Then we have the following expectations (in terms of points) for team α's starting lineup.

Player 1 is expected to score $0.994 * 0.583 * 5 \approx 2.898$.
Player 2 is expected to score $0.994 * 0.812 * 5 \approx 4.036$.
Player 3 is expected to score $0.887 * 0.763 * 5 \approx 3.384$.
Player 4 is expected to score $0.960 * 0.371 * 5 \approx 1.781$.
Player 5 is expected to score $0.967 * 0.614 * 5 \approx 2.969$.

We then expect the starting lineup to score 15.068 points before the first substitution (which we previously assumed at 5 minutes.) After this substitution, we would obtain a new counterefficiency matrix based on the new matched lineup.

Therefore in order to predict the outcome of the game, we must also predict the substitution pattern of the two teams.

For the purposes of time and keeping this paper short, we end the discussion about the computation here and will move on to talk about the general use.

5. Who is this all really relevant to?

We may combine these techniques with a predictive model involving substitutions to predict the final score of different games.

There are however, other ways to do this as well that are much simpler and perhaps more reliable.

Who this might really appeal to however is the coaching staff. Counterefficiency analysis can provide a coach insights involving the most efficient ways for lineups to get played. It can show interesting results about who is best suited to be played on the court to contain/overwhelm the opposing team.

The biggest power to the coach however comes from analyzing different aspects of the game using the counterefficiency. For the sake of this paper, we have used points as the variable we are concerned about. This can be easily switched to almost any other variable including blocks, steals, and more.

There are often scenarios in NBA games where the score is very close late in the game and lineup optimization becomes of the utmost priority. Imagine 30 seconds on the clock, team α being behind 2 points, and team β having possession of the ball. Conventional NBA wisdom suggests team α look for a foul to stop the clock. With a tool such as this, we can think of alternatives. With the right lineup, we can drastically increase our chances of forcing a turnover without giving team β their freethrows (using counterefficiency data with respect to steals for example).

The opportunities for a coach to gain an asymmetrical advantage are truly endless using counterefficiency analysis.

6. Limitations

The first major limitation is one we already covered earlier in the paper regarding the elusiveness of substitution data.

The second major limitation is the very size of our data. We are essentially looking at every player in the NBA (who is still active) versus every other player in the NBA they have ever faced. This can get huge but luckily in this day and age, there are ways around it (parallel computing, Hadoop, etc.) Also, once we have counterefficiency data it could be easily stored for the future. This would mean that if we could get caught up with the data, we would simply need to update it after games which would be very easily manageable.

The last major limitation is again the size of our data. There will always be some matchups which are severely lacking on data (imagine two deep bench players who rarely see any minutes or a player who has been prone to many injuries.) There is also the matter of players who are new to the league who lack data altogether. Among well known NBA veterans we can probably expect more than enough matchups however (ex. Lebron James and Carmelo Anthony.)

7. Conclusion

This paper isn't meant to explain a model which is already in use but to rather suggest a model for further consideration. There are both advantages and limitations to this form of analysis but it seems worth looking into for the sake of research if nothing else.

I believe however, this form of analysis can be very powerful in the hands of a coach. With both coaches having access to this type of analysis, I hypothesize that lineup generation may even be reduced to an insightful simultaneous game in itself.

E-mail address, Zohaib Syed: zohaibrsyed@gmail.com