Effective Defensive Court Positioning in the NBA

NBA Basketball Analytics Hackathon 2016

Prime Suspects

Contents

1 Introduction

Defensive effectiveness is often difficult to quantify because many characteristics such as speed, movement, and proximity to offensive players can all contribute to the overall defensive skill of an individual player. Often, the result of an interaction between an offensive player and a defensive player is attributed to the skill of the offensive player. However, as all coaches know, defense is a key to success. We aim to classify between effective and ineffective defensive strategy. The primary focus of this paper is to answer the following question?

Where should a defender be while his team is on defense ?

We have trained a logistic model to classify the effectiveness of defense at possible positions of the court based on game circumstances. Then, we use this model to build two small visualization tools:

1. A tool that allows a user to drag an offensive player and the ball to locations on the court and then determines where the optimal position is based on historical data.
2. A tool for post-game analysis that shows the optimal defensive positions during real time game play. This can be used for coaches and players to recognize where and when they deviate significantly from the optimal position that we have determined based on historical data.

2 Restatement of the Problem

To simplify this difficult problem, we will first make the assumption that any defending player has a corresponding offensive player that he is guarding. This is a reasonable assumption because we know that most NBA teams in fact do not play zone defense and instead focus on man to man defense. Given this information, we can consider two parameters when evaluating how well a defender is doing at any given play: *the location of the defender with respect to the offensive player* and *the location of the defender with respect to the ball.*

Now, we will characterize a defensive player as "effective" if he is consistently within zones that put pressure on the opposing player that he is supposed to be guarding. Similarly, he is a "ineffective" defensive player if the player he is guarding is able to obtain the ball and/or shoot the ball. We will define what exactly we mean by effective and ineffective in Section 4.2.

Then our task is as follows: Use past data regarding the location of the ball, the location of the offensive player, the location of his corresponding defensive player, and the outcome of their interaction to classify defensive positions as effective or ineffective.

3 Initial Assumptions and Simplifications

1. **Assumption:** If a team is on defense, then all the players on that team have one player that they are guarding.

   Justification: As mentioned above, most NBA teams do not play zone defense and instead play man to man defense.

2. **Assumption:** We will ignore the quality of the offensive player is when evaluating how well our defender performs.

   Justification: Our model only focuses on whether or not a particular defender gives his corresponding offensive player the chance at making a shot. While the skill of the offensive player does influence if he is able to make a shot given the defender's positioning, we reason that over numerous plays and games, the average quality of the offensive player that a particular defender has to guard will be an average quality consistent across all defenders.

3. **Assumption:** The assigned defender to a particular offensive player is the closest player on the opposing team.

   Justification: Since most NBA teams play man to man defense, it is reasonable to assume that the players on the defending team will guard the closest offensive players to them.

4. **Assumption:** The success of a defensive interaction between offensive player A and defensive player B is determined by whether A attempts a shot during the interaction.

   Justification: If B is doing a good job on defense, A may not even gain possession of the ball in the first place. However, even if A does gain possession of the ball, we can gage B's defensive skill by how confident A feels in his ability to score. If B is a bad defender, A will likely attempt a shot. On the other hand, if B is a bad defender, A will likely pass the ball to a teammate instead.

5. **Assumption:** We are limiting defensive and offensive matchups to players on the same half of the court.

   Justification: During the moments when players are transitioning from one half of the court to the other half of the court, offense-defense matchups are in flux, so the use of proximity as a means of determining these matchups is inaccurate. For simplification purposes, we restrict analysis to those players who are in nearest proximity to each other on the same side of the court.

4 Methodology

4.1 Data Cleaning

We will extract player match-ups from the Hackathon-sv-raw-playoff-2016 file. First, we reduce the granularity of our data to 1s from 40ms. Note that our analysis applies to lower-granularity data, since a player's defensive rating is measured as a percentage of the time that he spends in the optimal zones. We reduce granularity in order to make our computation times faster. Note that even with this reduced granularity, for a player playing 30 min/game, assuming he spends half of his time on defense, we still get 900 data points per player per game. Thus, 1 sec will be defined as our time step, ΔT.

Now for each defender at each time step, we fix that defender and attempt to determine which player he is guarding. For every player in the defending team, we will say that the player he is guarding is the closest player to him in the offensive team. We will do this until all 5 pairs of defender/attackers are completed, while ensuring that there are no overlaps across pairs. This first approximation doesn't take into account screens, for example, but is still accurate enough for an initial model. See *this paper* for an example of a more advanced model.

4.2 Creating the Model

For each time step and for each defensive player P_{def} on the court, we have the following information (in polar coordinates):

1. (r_{def}, θ_{def}): Location of P_{def}
2. (r_{off}, θ_{off}): Location of corresponding P_{off} from the other team
3. (r_b, θ_b): Location of the ball

This gives us a tuple of six elements for each defender, offensive player pair: $(r_{def}, \theta_{def}, r_{off}, \theta_{off}, r_b, \theta_b)$ We will define this as the **Position Tuple**. This information is displayed in Figure 1.

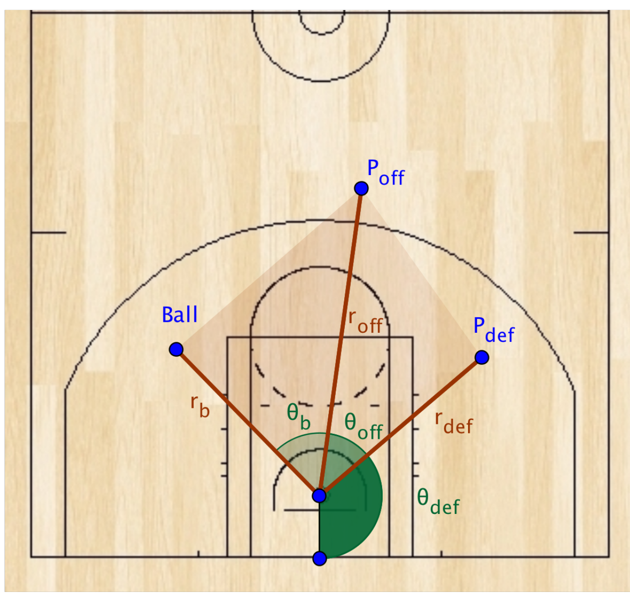

Figure 1: Locations of P_{off}, P_{def}, and the ball in polar coordinates.

Note that P_{off} may switch between time steps if P_{def} gets closer to a different member of the opposing team. However, as we are treating each time step as a different data point, this causes no problems.

In order to quantify whether the defensive strategy was effective or not, we will look at the result of that play. Using Assumption 4, if P_{off} attempts a shot at time T, and P_{def} was his defender immediately before this play, then we consider the the defensive positions of P_{def} of P_{def} prior to time T to be ineffective (denoted by 0). If P_{off} does not attempt a shot at time T, then we consider the defensive strategy of P_{dens} prior to time T to be effective (denoted by 1). Note: though is is likely that P_{def} was in good position at some point during his defense of P_{off}, we will consider all his positions as ineffective because they ultimately led to a attempted shot.

4.3 Logistic Regression

As stated in the above section, we can classify each tuple representing the locations of the defender, the offensive player, and the ball as either a 1 or a 0 with 1 representing that the defender was in an effective location and 0 otherwise.

Now our goal is to use the data given to us to help predict which locations are optimal for a defender in future games. Thus, our problem is reduced to a binary classification problem where we input in the location of the offensive player, the location of the ball, and the location of a particular defender and output whether or not the defender is in an effective position in the court. This task can be accomplished by running a logistic regression. Our training data will be the position tuple and the values (0 or 1) that they correspond to that we acquired from the playoff 2016 data as described in section 3.1. Running the logistic regression will output an equation which will give us the **probability** that a particular position is effective. Thus, we need to determine a threshold probability to label a particular position effective or not effective. Since we want to prioritize good defense, we will set the threshold to be $\frac{2}{3}$. Thus, only positions that have a high probability of being effective according to our model will be labeled as effective.

5 Solution

5.1 Logistic Equation

Running the logistic model gives us the following equation. For ease of notation, we have taken the trigonometric expressions in our formula and transformed them into x, y, z coordinates. Let the x, y, z vector representing the location of the offensive player, the location of the hoop, and the ball be v_{off}, v_{hoop}, and v_{ball} respectively.

$$\text{Probability that a position is effective} = \frac{1}{1 + \exp(-\frac{10.3 \cdot v_{off} + .97 \cdot v_{hoop} + 1.21 \cdot v_{ball}}{12.21})}. \tag{1}$$

Using this equation, we can classify whether or not a particular defender's position is optimal. More specifically, we can input it the parameters about the location of the ball, the location of the offensive player, and the location of the defensive player and our equation will output a probability of that position being effective. We then compare that output with our threshold value of $\frac{2}{3}$. If that outputted value is less than the threshold value, then we will label the position of the defender as ineffective and if the outputted value is greater than or equal to the threshold value, we will label the position of the defender as effective.

We can further use our results to calculate the **Optimal** location of the defender. The optimal location of the defender will be defined as the location which results in the highest value output in our function above.

5.2 Use Case of the Solution

5.2.1 Visual Tool

We created a Java application to better visualize our result. In the application, we are allowed to move the position of the ball and the offensive player. Our defender will then automatically move to his **optimal** location where optimal is defined above. In our application, the **orange** circle corresponds to the ball, the **red** circle corresponds to the offensive player and the **blue** circle corresponds to the defender. Here are a few screen shots from our application:

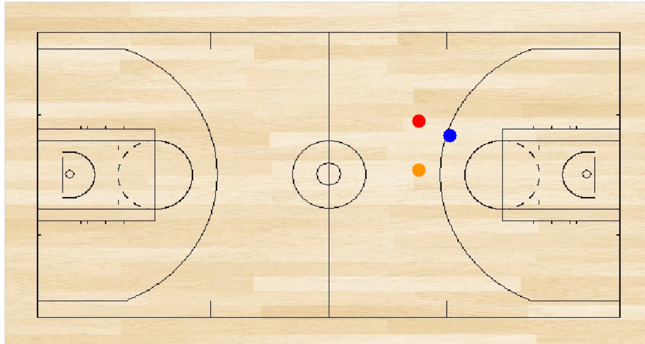

Figure 2: Example 1 of our application.

In figure 2, you can see that the defender is directly between the defender and the hoop. This supports conventional wisdom; it's unlikely that he would be able to stop such a short pass to the offensive player that he is guarding.Furthermore, moving up ahead begs for the offender to drive.

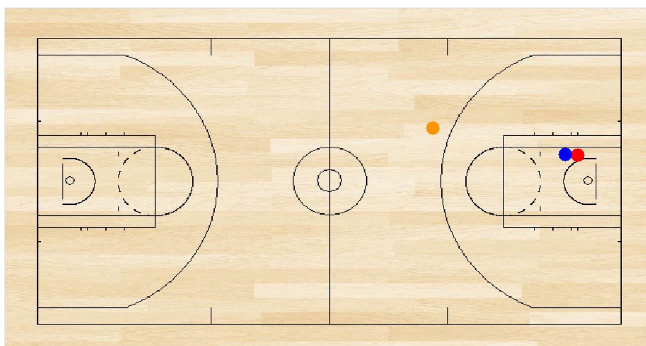

Figure 3: Example 2 of our application.

In the scenario given in figure 3, it doesn't make as much sense for the defender to be in between the offensive player and the hoop. Rather, he should try to prevent the pass as shown in the figure.

5.2.2 Tool for Post-Game Analysis

We have also developed a tool that can analyze a game after it happens, and suggests alternative defensive positions for the defense based on our model. We know that each defensive location has a effectiveness rating from our logistic regression model, so maximizing this function gives us the optimal defensive position for the defense player. Thus, coaches can use this tool after a game to evaluate the success of their players in maintaining good defensive positions throughout the game.

Here is a screen shot of this application:

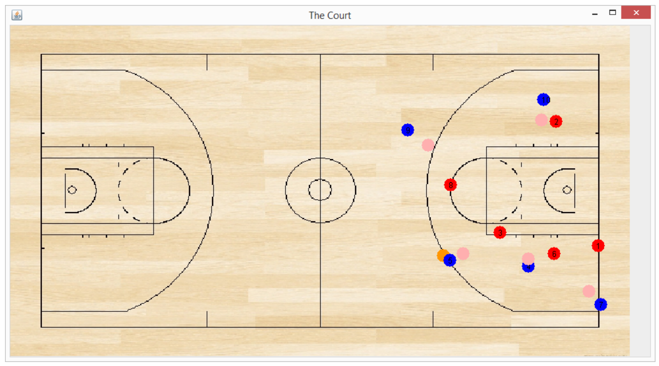

Figure 4: Example 2 of our application.

5.2.3 Ranking Players by Defensive Effectiveness

We can use our classification of effective and ineffective positions to form a metric that measures the effectiveness of the defensive play of an individual player. More formally, we can calculate the fraction of a player's effective positions to his total number of positions in the game. Let M be our metric.

$$M(\text{Player X}) = \frac{\#\{\text{effective positions for X}\}}{\#\{\text{total positions for X}\}}$$

This metric allows us to rank players based on their defensive effectiveness.

6 Strengths, Weaknesses, and Further Research

6.1 Strengths

1. Because our defensive positions depend only on the location of the ball and the offensive player, there are few components for a player to take into consideration when trying to implement our strategy. For this reason, the defensive strategy that results from our model is intuitive and will be simple to teach to professional basketball players.

2. Our visualization application makes it easy for both players and coaches to understand where the good defensive positions are based on certain game situations.

3. We have taken into consideration all player locations from the 2016 data set. In this way, we have included all unique defensive styles in order to make robust and universal recommendations.

4. Our threshold for a good defense is high. More specifically, our model tells us that a particular position is good if and only if our logistic equation outputs a value greater than $\frac{2}{3}$ for a particular position of a defender. While a choice of another threshold such as $\frac{1}{2}$ would also have been sensible, our choice gives us more confidence in our classifier method.

6.2 Weaknesses and Further Research

1. We have simplified the problem so that our defenders strategy is independent of the identity of the offensive player. Though this yields a reasonable default strategy, it is likely that a defensive player would slightly modify this default strategy depending on the unique characteristics of the person he is defending. We could study these unique players as an extension of our project.

2. We have assumed man to man defensive strategy for the data we have used. An interesting extension to our work would use other alternatives to determining which offensive player each player is covering, including using the inner product of their velocity vectors. Another

extension to our work could involve studying good positions for players when using the strategy of zone defense.

3. Our model simply takes in a position and outputs a binary results which states that the position of the defender is effective or ineffective. We can easily expand our model as follows. We can assign the position of a defender to a number which represents the expected value of the points that will be scored by the offensive player corresponding to the defender. Thus, a position corresponding to a lower number will represent a better position since less points will be awarded to the opposite team in expectation. This will allow us to classify how good a position is in much more finer detail than a binary classification.

7 Conclusion

Our paper presents a model to answer the following question:

Where should a defender be while his team is on defense ?

More specifically, our goal was to identify differences within defensive strategy using spatial data and use this data to categorize effective and ineffective defensive positions. Some key assumptions that we made while creating this model was that each player on a defending team is assigned to one unique offensive player. Our justification for this assumption was that most NBA teams play man to man defense and not zone defense. Using the historical data provided from the NBA, we have identified two parameters that we believe can be used as the basis of a simple but effective defensive strategy. These parameters are the location of the ball and the location of the offensive player at a given moment in time.

We have used a machine learning algorithm, logistic regression, to classify potential defensive positions as effective or ineffective. In addition, we have trained this model based on the success rates of the past for defense player in those same situations. Furthermore, we have created a Java application to display the optimal position of a defender given the ball location and the location of the offensive player. Moreover, we can also use our model to rank current players based on their defensive ability by looking at the proportions of the time they are in an effective positions according to our model. Some strengths of our model include its simplicity and our visualization tool. Lastly, we can extend our model in future iterations by increasing the granularity of our classifications by calculating the expected value of points given up in a particular defensive position. This would be an improvement over the binary classification that our model currently outputs.