

NBA Statistical Model Project Proposal

Vivek Ramchandran

June 5, 2018

Problem

One of the longest standing questions in sports is whether or not you can accurately predict what is going to happen in a match; how to effectively forecast different matchups and factors that go into deciding who wins a game, all the while accounting for human error. An incredible amount of work has gone into the current landscape of NBA single game prediction models, but they are all flawed for a number of reasons. One of the most popular and advanced models, created by FiveThirtyEight, uses both team and individual player statistics to create a win percentage probability for each team, the standard result for these models. However, nowhere do these models take into account how different matchups between teams and players affect the result, or adjust during the game based on certain performances or developments. There is no explanation as to why these models either perform accurately or fail to do so, or even the factors that must be taken into consideration. The current models do not analyze the free flowing game of basketball at the levels that are required - both from a data collection standpoint as well as an in-depth matchup analysis between teams.

Note

This proposal is written as if the final goal is simply to make a single game win probability model. This is simply one potential use case that the creation of this model will allow us to analyze.

Solution

Overview

The goal is to create an accurate model for game predictions that gives legitimate matchup factors to back up the prediction. Respective five-man units for each team will be analyzed for how they will play against each other, resulting in

suggested lineups to play against specific opponents. Furthermore, this model will allow for detailed mid-game adjustments in the predictive model. In order to do this, each team will need to be analyzed with a number of variables that extends beyond basic statistics. Player Tracking Data will be needed in order to measure spacing, defensive transition prowess, screen setting ability, and a number of other factors. Individual players will also need to be modeled, to determine how effectively they operate in a number of different situations. The team and player analysis will be used together to predict how the game will be played, and why the game will go a certain way.

0.1 Metrics

This model relies upon quantifying different aspects of basketball that cannot be measured with common statistics. The Player Tracking Data provided by STATS LLC SportVU will be used in order to measure the following metrics:

Offensive: Off Ball Metrics

1. Offensive Rebounding (Boxing out)
This metric will grade how well a player puts himself in a position to grab an offensive rebound.
2. Screen Quality (Screener)
This metric will grade how much space is created for a teammate when a screen is set by this player.
3. Screen Attention
This metric quantifies how much attention this player receives when a screen is set for him.
4. Spacing
This metric determines how closely the primary defender guards a player based on how far away from the basket he is.
5. Movement
This metric is defined based on how well a player moves without the ball, and how often he cuts. This metric says as much about the style of team play as it does an individual player.

Offensive: On Ball Metrics

1. Spacing
This metric determines how closely the primary defender guards a player based on how far away from the basket he is.
2. Attacking Space
This metric grades how effectively a player is at scoring/creating a scoring opportunity for a teammate when he has space in front of him (includes Transition opportunities).
3. Screen Attention
This metric quantifies how much attention this player receives when a screen is set for him.
4. Attacking Switches
This metric grades how effectively a player creates a shot for himself

or a teammate when his primary defender switched onto him.

5. Doubled

This metric grades how much secondary attention a defense pays to him.

6. One-on-One Shot Creation

This metric grades how effectively a player can create a shot for himself or his teammates when taking on a defender one on one.

7. Tough Shots Taken

This metric determines how many tough contested shots a player takes.

8. Tough Shots Made

This metric determines how many tough contested shots a player makes.

9. Post up

This metric determines how effectively a player can create for himself or a teammate when in a post position.

10. Clutch Factor

This metric determines how a players game is effected late in games with the game potentially on the line and he is fatigued.

11. Heat Retention

This metric determines how a player plays once he makes a couple shots.

12. Mid-Range

This metric grades a player's comfort level with the mid range shot.

13. Finishing

This metric grades how often and how successfully a player attempts and makes shots in the paint.

14. Transition Passing

This metric determines how well a player sees an early non-half court pass and the regularity with which he completes it.

Defensive Metrics

1. Switching

This metric grades how often and how successfully a player switches on defense.

2. One-on-one Defense

This metric grades the level of shot an opposing team will get when the opponent creates a shot for him or a teammate by taking on this player one on one.

3. Help Defense

This metric grades how effectively a player helps out on defense when one of his teammates is beaten.

4. Boxing Out

This metric grades how well a player boxes out or puts himself in a position to grab a rebound.

5. 3-Point Defense

This metric determines how effectively a player prevents the player

he is guarding from shooting three point shots.

6. Pick and Roll Defense

This metric grades the opponents shot quality after being involved in guarding a pick and roll.

7. In-Space Defense

This metric determines how well this player guards the opposing team when in transition or in space.

Note that not all of these metrics are fully thought out in what exactly they will look like. Some of them might need to be tweaked, and others will be quite difficult to quantify from the player tracking data.

0.2 Creating Graphs

Once these metrics have been created, and each player has been graded in terms of these metrics, this project will have access to a whole new layer of information that cannot be emulated (as they will be defined here). For each team, every metric will be represented in a graph. Each player will be represented as a single node with a value (his score in the given metric), and the metric graph will be connected with weighted edges to rank each of these players in the metric. A player who scores highly in the metric will have a directed weighted edge to a player that is not as proficient in the metric. This graph does not need to be fully connected, as there is no need to waste an edge between two players who have no value in a specific metric.

Once the metric graphs have been created, individual graphs for all of the players will be made. As each player is represented as a node in each of the metric graphs, all of the nodes representing a specific player will be put into a single graph. Those nodes will contain the players' score in each of the metrics (each of the nodes in the individual graph now representing a metric). The edges between nodes (or relating the metrics for a specific player) will be determined by running regression on all the different metrics to try to find correlations.

Once the metric graphs and the individual player graphs have been created, creating a graph representing the entire team is the next step. It will be a combination of the player graphs (weighted based on minutes played) and the metric graphs for each team. It is likely that certain metrics can be collapsed for the team graph, so as to not overcomplicate things. These team graphs should hopefully be able to be analyzed against each other in order to gain basic understanding of how teams play, and different styles and strategies they will attempt to implement against a specific opponent.

0.3 What's Next

Once all of these graphs have been created, there will be enough information to do a number of things. They can be used together to create a single game win probability model that is more complex and detailed than any model that is available publicly. They could be used to help a team choose which players to pursue in the off-season to best find their team needs, essentially acting as

a general manager. It can be used to visualize individual players, and even to determine how players should be valued. There are a number of directions that the project can take once these graphs have been created.

Other Works

There are a number of other models on the NBA that use similar methods. Many use regression, a couple use graph theory, and an even more select few create their own metrics in order to make use of "additional" data than is just available with normal statistics. The works that will be mentioned here all make use of the same Player Tracking data.

These models are not necessarily doing the same thing. Two very interesting papers are models that are focused on analyzing shot selection and possession value. One, <https://hwchase17.github.io/sportvu/> uses logistic regression and the random forest model to predict the result of each shot, and analyze how shot quality and results change based on game score and time. These results can be used to analyze player shot tendencies, as well as which teammates result in better shots and higher percentages. A second, <https://arxiv.org/pdf/1408.0777.pdf> works to predict basketball possession outcomes, or the EPV (expected possession value) based on who has the ball and at what time. It makes use of the spatial occupancy of each player to determine tendencies, and then grades each player on a number of movement parameters that it uses to get more information about how a possession should be graded based on which players are playing. Another one of these models <https://arxiv.org/pdf/1703.07030.pdf> actually creates certain features or metrics (same general concept as the first step of this model) in order to determine which team models most lead to 3 point success. It can use the additional data that it scrapes to also analyze whether or not players are taking shots that a player of their caliber should take, or whether or not players should be shooting more or less.

There are a number of other models that use creative and very interesting methods to try to analyze different aspects of the game. Only three works are listed above, but many more have been read and analyzed.