2Bucketz NBA Hackathon 2017

Calvin Lui, Michael You, Wilson Yu, Richard Zhao

September 23rd 2017

Abstract

In order to give basketball enthusiasts a better idea of what is really going on in a game, we developed a new way of presenting box scores and created a new statistic called the Effective Playmaking Metric. From a visualization standpoint, we shifted away from the summary-based table system of the box score towards a graphic-based accumulating metrics plot. Using play-by-play data, we are able to divide up the game into its smallest significant events to show not only what occurs in the aggregate, but when these occurrences take place. To supplement this accumulating metrics plot, we also graphed the ball and player positions on the court, frame by frame, with SportVU data. Taking these frames, we animated smooth movements on the court and updated the accumulating metrics plot in real time. Along with our visual overhaul of the traditional box score, we developed the Effective Playmaking Metric (EPM), which was inspired by the apparent lack of a relevant metric for determining true offensive prowess. The EPM considers the following: dribbles per possession, expected value of shots modeled by a neural network, and assist to touch ratio. We hope this new metric will help differentiate between true team players and stat-padders, as those who make decisions with the team in mind will achieve higher EPM ratings. To compare this new metric of playmaking and offensive effectiveness with existing ones, we included a radar graph on our visual dashboard with points, assists, plus-minus, FG%, and 3P% on the axes and EPM represented by the intensity of color. This radar is updated along with the accumulating metrics plot and the player-ball positions, so we can see how all of these variables change with each other. The engaging visuals of our dashboard, along with the Effective Playmaking Metric, represent our vision for capturing the action and complexity of the game in a simple and accessible interface.

1 Introduction

1.1 Problem Statement

Although the abundance of advanced basketball analytics has allowed fans and experts to understand the game better now more than ever, two problems still exist with the present box score system that prevent us from understanding what is really going on in the game. The first problem lies in the presentation of game data, while the second problem involves the lack of a metric for playmaking ability.

1.1.1 Revisualizing the Game

When most fans want to know what happened in a game, they tend turn to box scores, a table that lists the total number of points, assists, rebounds, and other statistics that each player achieved in the game. However, two drawbacks to this approach of presenting data is that 1) a table is not as visually engaging as a graphic and 2) a box score's summarizing nature does not allow us to see what happens at what times.

1.1.2 A Metric for Playmaking

Despite how important it is as a skill, playmaking ability is often overlooked in basketball analytics because it is not an easy thing to measure. Metrics such as assists can give an idea of how well a player can run an offense, but this statistic can be inflated in several ways - such as handling the ball very often and by passing as the shot clock expires and forcing teammates to take (bad) shots that count towards the passer's assists. Thus, a high number of assists does not necessarily represent a player's ability to make good decisions that will help the team's offense.

Unfortunately, identifying and ranking good floor generals often comes down to the eye test and other qualitative methods, current advanced metrics only examine whether a player chooses to do something and not whether that player's choice was good or bad. Thus, we attempt to devise a new metric that will consider whether a player is making the right choices or not, depending on the situation.

1. Compute a numerical value for a player's ability to make the correct play and to run an offense.
2. Visualize a simulation of a team's offense in a single possession.

1.2 Assumptions

We focus on measuring a player's playmaking ability under low-stakes circumstances, as late game situations often rely on defined plays that do not necessarily reflect the playmaker's own choice. However, during typical circumstances, it is the player that chooses what to do with the ball.

1. **Assumption:** The designated defender to a player is the opposing player closest to him.

   Justification: Most defenses in the NBA are man to man.

2. **Assumption:** We ignore an individual player's ability to convert shots at a certain location.

   Justification: The goal of our study is to evaluate a player's ability to make plays that lead to smart shots and not to evaluate ability to convert difficult shots or to identify a player's personal hot spot. Situations where a player is wide open at a high-conversion spot but personally have low conversion rates are not common enough to heavily skew results.

3. **Assumption:** We omit players who play less than 15 minutes per game.

   Justification: Players who play less than 15 minutes are often not put in a playmaker position and see a substantial part of their minutes in garbage time. They likely do not run the bench unit, nor are they likely to be a starter.

4. **Assumption:** Long periods of time spent dribbling the ball with little movement implies poor ball movement and little playmaking.

Justification: We do not evaluate a player's scoring ability but rather his ability to involve teammates and to create good shots. Dribbling the ball for a long period of time often indicates either an isolation play or a stagnant offense.

5. **Assumption:** Traditional big men will be more penalized in the study than other positions.

   Justification: Big men usually don't run an offense and are never completely open in the paint. The study is aimed moreso for wings and guards and does not reflect on a big man's ability to shoulder an offensive load.

6. **Assumption:** A poor shot is defined to be a shot that is tightly contested and from a low percent make location.

   Justification: Open shots are obviously better to take than highly contested ones and shots taken from certain regions are obviously better than shots taken from other regions.

1.3 Model

The things that we consider as good playmaking under typical circumstances include:

1. **Fewer dribbles:** Players who tend to dribble for a long time before passing the ball tend to do so only because they have to. Dribbling less means more ball movement and less isolation.
2. **Taking good shots:** A good playmaker does not necessarily have to always look to pass. If the player is in a good position, he should take the shot. However, if a player is taking contested shots, he should be penalized as a playmaker - unless he had just received the ball and had to take the shot with time running down.
3. **Ability and willingness to pass:** A good playmaker should be willing to share the ball, but also make strategic passes that lead to points.

Weighing these three factors, we calculated a new metric we denote Effective Playmaking (EPM).

2 Methods

2.1 Data Processing

We use several sources for our data visualizations and playmaking metrics. For visualizing the players' position on the court, we plotted SportVU data from several games from the week of 10/25/16. Then, we cross-referenced plays from certain times to play-by-play data in order to live update the statistics. During this time, we also calculated our new metric, Effective Playmaking (EPM) and incorporated EPM into a radar graph that included points, assists, plus-minus, field goal percentage, and three-point field goal percentage on five axes, with EPM represented by intensity of color.

2.2 Implementation of UI

We used d3.js to create a webthingy to visualize our new box score. We update a court depicting the ball and each player's location in real time and have a radar chart depicting selected players' raw numbers and EPM. We also implement a chart that allows for live comparison between teams' starters and bench units with regards to traditional statistics and EPM.

2.3 Crafting of EPM

To come up with EPM, we took in three inputs:

1. **Dribbles:** We wanted to penalize players who dribbled more than necessary, though after a certain number of dribbles, we figured there was no point in further penalty. Thus, we decided on a flipped logistic function. We estimated that players who dribbled for four or more times were not effectively utilizing their ability to make a move or pass to someone in a more strategic position, so we made four dribbles the point at which EPM would fall fastest.

2. **Expected Shot Value:** We decided to model shot selection as a factor for the EPM, since playmaking and offensive prowess is heavily dependent on shooting behavior. To do so, we employed a deep neural network, which is superior to traditional linear techniques in regressing over complex trends.

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left[-y^{(i)}\log(h_\theta(x^{(i)})) - (1-y^{(i)})\log(1-h_\theta(x^{(i)}))\right].$$

(a) Minimizing the Cost Function J with features X and output Y, Shot Value.

$$\frac{\partial J}{\partial \theta_j} = \frac{1}{m}\sum_{i=1}^{m}\left((h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)}\right).$$

(b) Gradient Descent Equation for Optimization.

For our inputs, we chose distance from basket, touch time during possession, distance with the closest defender, and the attempted point value (2 or 3 points for field goals). These features combined to generate a fairly accurate model when trained over SportVU shot data from 2015 to 2017. The error for a randomly sampled test set (which was not included in the training set), was $7.08 \times 10^{-}5$ with our model of three hidden layers, one unit per layer. The layer-unit configuration of the model was determined by a simple grid search to minimize cost.

3. **Assist/Touch Ratio:** We chose Assist/Touch ratio as the best metric for a players' ability and willingness to pass the ball. A player could easily inflate their assist/pass ratio by not passing unless they see a clearly open player, and a player could just as easily inflate their pass/touch ratio by passing every time they get the ball. However, it is much harder for a player to artificially inflate their assist/touch ratio.

We then adjusted the weights for these values based on what we believed was most important. Dribbling was worth about 10%, expected shot value was worth about 50%, and assist/touch ratio was worth about 40%. After computing a raw score, we then scaled the score from 0 to 10.

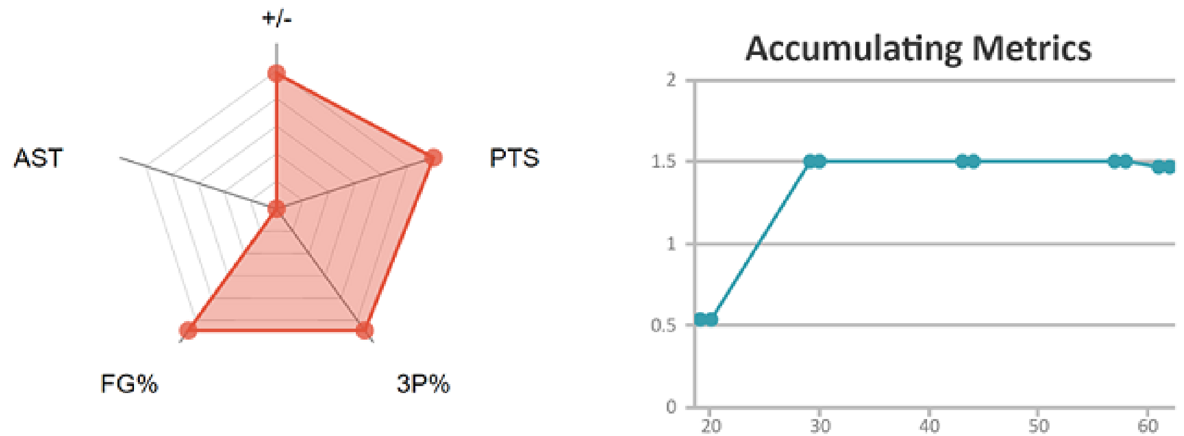
$$EPM = \frac{1}{1+e^{Dribbles-4}} + ESV + \frac{Ast}{Touch} \tag{1}$$

3 Results

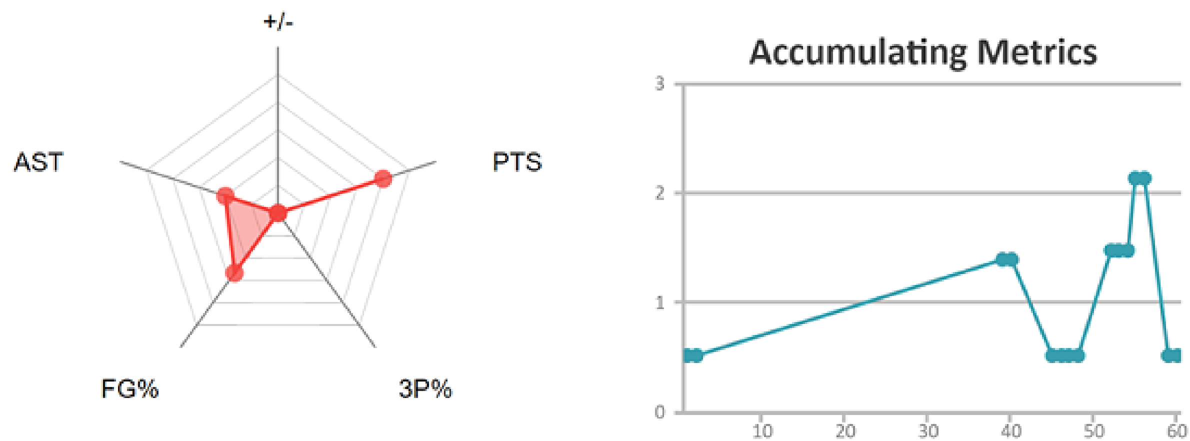
3.1 Cavaliers versus Knicks

3.1.1 Case Study: Carmelo Anthony

Carmelo is an example of a player that looks like a highly proficient offensive player from box stats alone. Upon closer inspection, we see that he indeed is able to shoulder a substantial offensive scoring load, but this does not indicate that he is able to shoulder the burden of running a team's offense. We consider his traditional statistics and compare it with his EPM. We note that his predisposition to involving the whole team on his offense is very slight, and positions where he is in control often leads to a stagnant offense. That is to say, Carmelo is not a good floor general, though this does not reflect on his overall offensive prowess. As a player who likes to iso and who likes to take jump shots, our model reflects poorly on this behavior and assesses his ability to be a playmaker as "poor." His rating fluctuated frequently due to his tendency to take poor shots and yet convert on those at a high rate. Even when he would go on runs, his EPM would not skyrocket as expected since we evaluate the goodness of a shot and not his scoring ability. We see that as a result, his EPM decreases considerably when he misses a shot, and doesn't go up as much when he does make a shot.



(c) Kyrie Irving's statistics and EPM after 2 quarters



(d) Carmelo Anthony's statistics and EPM after 2 quarters

Figure 1: $p_d(t), p_{dt}(t)$ for Positive and Negative Benefit Cases

3.1.2 Case Study: Kyrie Irving

Kyrie Irving is a premier isolation scorer and point guard, yet has the reputation of falling short as a playmaker point guard. Our statistical model and new EPM metric reflect this aspect as well, as Kyrie is able to score plenty of points while taking low percentage plays. We note that while he is able to carry the Cavaliers as far as scoring goes, the red coloring in his graph indicate his low EPM and poor ability to make plays for the good of his entire team. This is a more accurate representation of Kyrie and his playstyle and offers more insight than a simple box score.

4 Discussion

4.1 Impact on Stakeholders

This new box score will allow for better understanding of the game that would go beyond showing post game numbers. Spectator wise, it would enable the viewer to see how a game and player evolved during the game, as well as the player's general style. It would reveal ability to lead a comeback or start a run, disposition to crumbling under an opponent run or pressure, evolution in on court behavior, willingness to pass and how it can vary depending on how close the game, and whether or not high numbers are a product of high usage rate. Post game wise, this new box score and visualization would allow for a more accurate and clear judgment on a player's fit, true worth, and impact on a team. High numbers that were attained at the cost of the team's success would become more apparent, which would help coaches, front offices, and spectators alike to identify a player's success as a product of a selfish and inefficient play-style, or as a product of genuine skill and basketball IQ. Front offices would also be able to see a quantified version of a player's playstyle, which could be a key factor in deciding which players to sign and seeing whether or not a player could fit into their offensive system. This new visualization of a box score combined with a metric to measure offensive IQ, offers a deeper insight in what a player brings to the court, as numbers do not necessarily reflect production.

5 Conclusion

5.1 Future Work

1. We have simplified defensive schemes to reflect simple man to man defense; to more accurately assess offensive IQ, we would have to model how the opposing defense adapts. That is, we would have to account for set plays, pick and rolls, switches, player matchups, defender velocity, help defense, hybrid man zone defenses (i.e box-and-one defense), and entirely new defenses implemented to counter specific teams.

2. We do not account for many niche offensive situations. That is, the free flowing nature of offense as well as a team's personal preference for a certain offensive scheme make it difficult to evaluate what is an objectively good shot, as the definition of a good shot can differ from team to team. We can look to include in our metrics a team's more individual characteristics when evaluating for "proper decision making." For example, Russell Westbrook has a reputation of being a low efficiency player with sometimes questionable decision making; we could look to not penalize him as heavily due to the fact that the Thunder's offense is meant to revolve around him initiating.

3. We also conceptualized a new statistic for the box score, but ultimately did not implement it since the statistic was not developed. This statistic, called the Action Rating, would express a player's excitability to fans based on the density and type of his activity on the court. This statistic can be used to connect box scores to highlight videos, and provide fans with a quick snapshot of who to watch on the court.

Appendix

Data Usage

Main

1. `SportVU Tracking Data`
2. `NBAPlayerTrackingData_2014-17`
3. `Play_by_Play`
4. `Player_Boxscores`

Auxiliary

1. `Team_Map`
2. `Player_Map`

Technologies

1. **Python**
   (a) Numpy
   (b) Pandas
   (c) Jupyter Notebook
   (d) Sklearn
   (e) Matplotlib
2. **Web Application**
   (a) Flask Backend Server
   (b) Webpack
   (c) Node.js
   (d) d3.js
   (e) SASS