Assist Quality: Measuring the True Value of Basketball Assists

Alexander Lee
University of Toronto

Introduction

Of the factors that influence a shot, one relatively untouched set of factors are those associated with the pass that lead to the shot, particularly the physical nature of the pass. Current pass statistics consist of binary variables with arbitrary definitions that are left to the non-uniform judgement of thirty different home court statisticians. We used SportsVU tracking data to examine the physical nature of passes through the following metrics:

- Pass Speed
- Pass Distance
- Path Congestion (discussed below)

After collecting the data for all passes that led to scores, we introduce metrics which allow us to measure the impact that a pass has on the resulting shot. By analyzing these metrics as applied to specific players' passes, we are able to see which players create the most value through their passes.

Data

This study looks at the shots taken by the Los Angeles Clippers in the 2014-2015 regular season. Using the optical tracking data provided by NBA stats, it is easy to capture two important measurements: pass time and distance, and consequently pass speed. However, these characteristics alone do an inadequate job of describing a pass.

Exceptional passers are said to have great vision and accuracy. While it is impossible, for the foreseeable future, to know exactly where a player is looking, we know that part of having great vision is being able to see openings in a defense, and having the precision to pass through tight windows. This is where our metric, path congestion, comes into play.

Let the points of the $defender_i$, ball and shooter at time t be $D_{t,i}$, B_t and S_t respectively and $d_{t,i}$ be the distance between $defender_i$ and the path of the pass at time t.

$$\theta_{t,i} = \begin{cases} 1 & \text{if } \angle B_t D_{t,i} S_t \leq \frac{\pi}{2} \text{ and } \angle D_{t,i} B_t S_t \leq \frac{\pi}{2} \\ 0 & \text{otherwise} \end{cases}$$

$$\gamma_{t,i} = \begin{cases} \frac{(15-d_{t,i})^2}{225} & \text{if } d_{t,i} \leq 15 \\ 0 & \text{otherwise} \end{cases}$$

$$C_{total} = \sum_{t=0}^{t_{pass}} \sum_{i=1}^{5} C_{t,i} = \sum_{t=0}^{t_{pass}} \sum_{i=1}^{5} .2 * \gamma_{t,i} * \theta_{t,i}$$

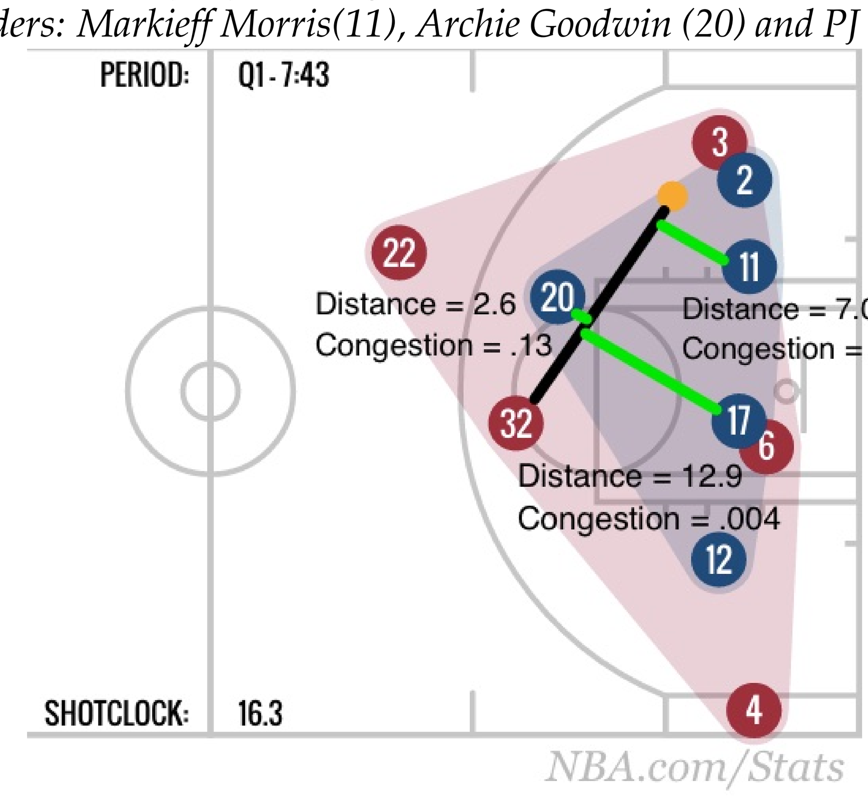
The first aspect to consider is whether $defender_i$ is actually between the ball and shooter. This is accounted for in the theta term which is an indicator variable based upon angles calculated using basic coordinate geometry. The second aspect is taking into account how close the defenders are to the pass, which raises a non-trivial question: how does one define being "close" to a pass?

A hypothetical line was placed between the current position of the ball and the shooter, a projected path. Then the distances between the defender points and this line were recorded.

A quadratic depreciating effect was implemented for the distance because we believe that a defender's effect decreases quickly as distance increases. If the defender is greater than 15 feet away from the path, his congestion contribution is set to 0. Path congestion is a unit-less metric and meaningless out of context, but gives us a relative scale to compare how crowded the pass' path is, and how difficult it was for the passer to see the passing lane.

Below is a snapshot from a shot taken during the last regular season game between the Los Angeles Clippers and Phoenix Suns. The play is a made Blake Griffin(32) midrange shot assisted by Chris Paul(3) and the image is the mid-pass tracking data.

Figure 1: *Sample path congestion calculation with three contributing defenders: Markieff Morris(11), Archie Goodwin (20) and PJ Tucker (17)*



The SportsVU tracking system records new coordinates every .04 seconds. So once we have calculated the range of time that the pass takes place, we iterate over each frame, calculate the path congestion for that specific set of coordinates, and average the total congestion over the length of the pass. This leaves us with a value between 0 and 1 for each pass: 0 if there were absolutely no defenders contesting the pass, and 1 if all five defenders were standing along the pass path.

Analysis

Intuitively, the value of a pass should decrease the greater the touch time of the shooter, so we separated our training data in to separate bins of touch times: [0-2], [2-4], [4-6], [6-8], [8-10] and [10+]. We utilized a logistic regression for each of the separate touch-time bins.

The model was trained using the new passing data, pass distance, pass time and path congestion. However, we also needed to consider what other determinants to include.

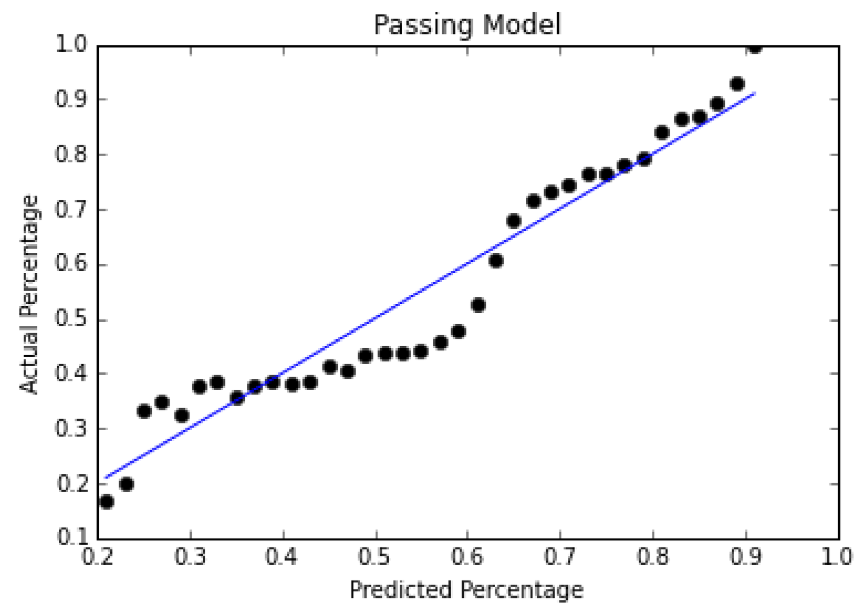
We looked towards previous modelling entries. There has already been considerable discussion regarding the quality of shot and the impact of certain factors on making or missing a shot. Several papers have focused on one variable or set of variables: spatial locations through shot charts (Goldsberry 2013), defender distance (Chang 2014), and difference in defender height and touch time before shot (Narsu 2015). We made a selection among the above factors and used the Bayesian Information Criterion to make our final selections for each individual bin.

Results

Passing parameters were only statistically significant for the touch times between 0 and 2 seconds. This makes empirical sense - the longer a shooter holds the ball, the more the shot creation is dependent upon the shooter's ability as opposed to the passer's.

With the model that we have fitted we can separate the terms that deal with the pass versus those associated with the shot or defender. To look at the model's accuracy we grouped the expected p-hats for the test data into individual 2% intervals.

Figure 2: *Plot of model predictions vs actual shooting percentages*



For each shot we can look at the value of the combination of all the passing terms. We call the combined value of these terms *assist quality*(AQ). Furthermore, we can multiply this value by the type of made shot, accounting for the added value of a 3pt assist. We call this metric *assist points*(AP).

If we look at the top five passers:

Name	Total AQ	Average AQ	Total AP	Average AP
Chris Paul	322	.716	747	1.66
Blake Griffin	129	.713	326	1.80
Matt Barnes	53.8	.728	126	1.70
Jamal Crawford	49.4	.716	116	1.68
JJ Redick	48.8	.717	114	1.68

One would anticipate that a lot of offense is generated through Paul and Griffin's passing, but Paul's total stats are still staggering. He accounts for more assist points than the other four combined.

Looking at average assist quality, we see that Barnes seems to slightly separate from the rest of the group. This matches what we empirically see during the game: Barnes is an underrated, excellent passer who can whip the ball across the court. His assist quality is also boosted by some of his longer, faster inbound passes.

One possible explanation for Griffin and Barnes's higher assist points is that they pass to 3pt shooters more often. Griffin can pass to Paul, Crawford or Redick, but when Redick has the ball, his only 3pt passing options are Crawford and Paul. 3pt shooters lose opportunities for valuable 3pt assists because they cannot pass to themselves.

Conclusion

From our results, we have shown that for touch times under 2 seconds, the characteristics of the previous pass have a statistically significant effect on whether the shot is made or missed. With rich locational data, it is possible to extract new information. This new data allows us to propose a method to quantify the effect of the pass, which is a more detailed metric than the current binary metric of assist or not assist.

The field of shot probability models is filled with brilliant work done by great statisticians who have already considered a cornucopia of factors. We have appended just a small set of potential parameters, which we hope future regressions will consider including with their determinants.

This paper has just scratched the surface of potential passing research. One of the next steps is to add all of the teams and samples to the models. The process of web scraping and sifting through the locational data for shots is very computationally intensive and requires more powerful resources than were available to us. Another major advancement that should be considered is refining the congestion metric. We only used a very basic distance based measurement. One improvement would be to create an expected deflection probability factoring in a specific defender's skills, who may be more or less adept at disrupting passes.

References

[1] Yu-Han Chang, Rajiv Maheswaran, Jeff Su, Sheldon Kwok, Tal Levy, Adam Wexler and Kevin Squire. (2014, February) *Quantifying Shot Quality in the NBA*. Paper Presented at the 2014 MIT Sloan Sports Analytics Conference, Massachusetts Institute of Technology, Cambridge.

[2] Kirk Goldsberry, Ph.D. (2013, February) *CourtVision: New Visual and Spatial Analytics for the NBA*. Paper Presented at the 2013 MIT Sloan Sports Analytics Conference, Massachusetts Institute of Technology, Cambridge.

[3] Narsu K.(2015). *Introducing Shot Difficulty: Comparing Game-Winning Shots in the Playoffs*. Retrieved from Nylon Calculus website: http://nyloncalculus.com/2015/05/13/introducing-shot-difficulty-comparing-game-winning-shots-in-the-playoffs/

Acknowledgements

We thank K. Narsu for invaluable season long data sets that are not found online such as difference in defender height and dribble type, Jefferson Lee for improvements in the data scraping and mining code and extensive proofreading, and Mac Jones and Ryan Hutchings for brainstorming and refinement of the assist quality concept.