Three-Point Shooting Analysis using Player and Ball Movement
Grad Sports
Nate Bailey, Karan Bhuwalka, Hin Lee, Tim Zhong

Introduction

In this study we use a rich dataset with NBA player movements from the first half of 2015-2016 to:
1. Analyze and predict shot quality
   - Identify features in movement data which correlate with successful shots
   - Use logistic regression to estimate probability of success for individual attempts
2. Evaluate players by types of shot attempts
   - Discover similarities between players based on movement data rather than position
   - Compare player 3-point shot percentage to predicted shot success accuracy

Motivation

Basketball is one of the fastest growing sports in the US. Over the past ten years, annual league revenue has increased almost 75% to 7 billion dollars. For such a large sport the use of data analytics is a useful tool to inform coaches and viewers of trends that may not be visible to the naked eye. We hope with this study will appeal to coaches, fans, and viewers of trends that may not be visible to the naked eye. We hope with this study will appeal to coaches, fans, and viewers of trends that may not be visible to the naked eye.

Data

Player movement data from the first half of the 2015-2016 season is publicly available. This data includes:
- (x,y) positions of all 10 players on the court
- Averages of 25 frames per second resolution.
- 64.5 GB of data from 636 games
- 17977 three point shots

When we speak of features, we refer to the coordinates:
- (x,y,z) positions of the ball
- (x,y) positions of all 10 players on the court

Features derived from the coordinates:
- Catch and Shot times from ball position
- We define features from the coordinates

Does Player Speed Influence 3-Point Shot Success?

Is there a significant difference in shooter speed for made vs. missed 3-point attempts?

- Faster movement can make defending a 3-point shot more difficult
- But it also requires more precision to shoot on the move

Hypothesis Test:

\[ H_0 : \ P_M(s[M=1]) = P_M(s[M=0]) \]
\[ H_1 : \ P_M(s|M=1) \neq P_M(s|M=0) \]

Use two-sample Kolmogorov-Smirnov Test on speed distributions of makes (M=1) and misses (M=0)

Results:

- Catch & Shoot (C&S) attempts: some measures of speed significantly differ for makes and misses
  - Average speed before catch for 0.5-2 seconds prior to the catch
  - Average speed between catching and shooting
- Off-dribble attempts: no speed measure was significantly different between makes and misses

Logistic Regression

Can we obtain features of three point shots that affect the probability of a shot being made?

Logistic Regression: Y = shot outcome (1 make / 0 miss) X = features of shot derived from movement data

- Significant correlations (p<0.05)
  - Shot Clock (-), Shot Distance (-), Defender Distance (+)
  - Time between Catch and Shoot (-)
  - Shooter Avg Velocity between Catch and Shoot (-)
  - Shot Distance Traveled Towards Basket (-)

Model Performance:

- 54% accuracy on predicting outcomes of individual shots
- Good at differentiating b/w better and worse attempts
- Make probability increases by 3.4% / ft

Potential Extensions

- Are there synergies between certain players or with certain lineups?
- How can these metrics influence offensive strategies and vice versa?
- Are different defensive strategies more successful against different player clusters?
- Are there other datasets that can be added that contain important information about shot quality?

Conclusions

- Shooter speed significantly differs between made and missed C&S attempts, but not off-dribble
- Identify 6 key features using movement data which allow for differentiation between good and bad 3-point attempts (of which 3 are new)
- Cluster players into 3 types based on movement characteristics of 3-point shooting tendencies
  - Better similarities than grouping by position
- Use movement data features to find differences in shot attempts between clusters and describe the shooting style of each cluster

Player Clustering

Can we characterize shooter types by the movements associated with their attempts?

- N~116 players who attempted >60 3-point shots in our dataset
- Use k-Means clustering on feature means for catch-and-shoot / off-the-dribble attempts

Movement-based clusters vs. position groups:

Cluster 0: Shooters who step up and shoot at top of the arc, with fewer catch-and-shoots.

Cluster 1: Shooters with greater sideways movement.

Cluster 2: Defenders who step up and shoot at top of the arc, with fewer catch-and-shoots.

Validation:

- MSE=0.003 (p = 0.07)
- R = 0.93
- R^2 = 0.87

Cluster Types:

- Cluster 0:射手在弧顶附近投篮，很少进行接球后投篮
- Cluster 1:射手具有较大的横向运动。
- Cluster 2:防守者在弧顶附近投篮，很少进行接球后投篮。