Nothing but Neural Nets

Using Machine Learning to Predict Basketball Shot Outcomes

Sriraman Madhavan, Ezra Van Negri, and Lisa Ann Yu
SUnet IDs: sriraman, ezravn, layu

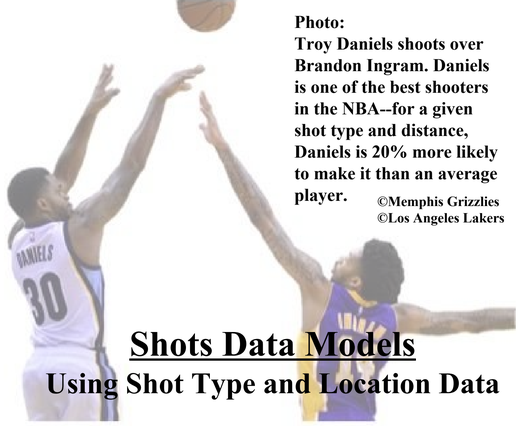


Photo:
Troy Daniels shoots over Brandon Ingram. Daniels is one of the best shooters in the NBA--for a given shot type and distance, Daniels is 20% more likely to make it than an average player.

©Memphis Grizzlies
©Los Angeles Lakers

Shots Data Models

Using Shot Type and Location Data

- Response: Shot made or missed
- Features:
  - Shot Distance
  - Shot Zone (3pt, Midrange, RA, etc.)
  - Shot Range (24+ ft., 8-16 ft., etc.)
  - Shot Area (Right Side, Left Side, etc.)
  - Shot Angle
  - Field Goal Type (3pt or 2pt)
  - Shot Type (Jumper, Layup, etc.): *Derived*
  - Shot Style (Fadeaway, etc.): *Derived*
  - Team *(in "Full" models only)*
  - Player *(in "Full" models only)*
- Methods:
  - Logistic Regression
    - With Lasso Regularization
    - With Elastic Net Regularization
  - Support Vector Machine (SVM)
  - Random Forest
  - Gradient Boosting Machine (GBM)
  - Neural Net

Data Description

- We use SportVU data from the 2015-16 NBA Season
  - SportVU data tracks players with cameras installed in the rafters of every NBA arena.
  - The data provide coordinates (x,y) for each player and the ball (x,y,z) 25 times per second throughout each game.
- From this raw data source, there are two datasets that we use:
  - 1) **Shots Data:** The shots dataset is pre-processed and has 24 variables (including shooter, distance from the basket, shot type, and shot made) and 205,538 observations (shots).
  - 2) **Coordinate Data:** From the full coordinate dataset, we extracted the ball's x, y, and z coordinates for all observations in our "Shots" dataset. These 20x3 coordinates represent the ball's path during approximately the first second of each shot.

Photo:
Corey Brewer, one of the worst shooters in the NBA according to our model. Relative to an average player, Brewer is 30% less likely to make a given shot.

©Houston Rockets

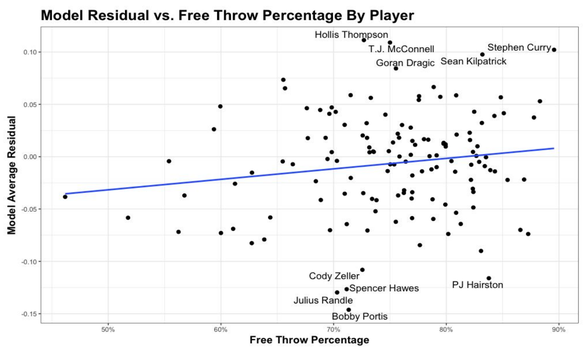
Coordinate Data Models

Using Basketball Trajectory Data

- Response: Shot made or missed
- Features (preprocessed):
  - 20 x, y, z coordinates relating to the ball's location during the first second of the shot attempt
- Methods:
  - Feedforward Neural Network

Best Coordinate Data Model	Train AUC	Test AUC
Feedforward Neural Network	0.852	0.712

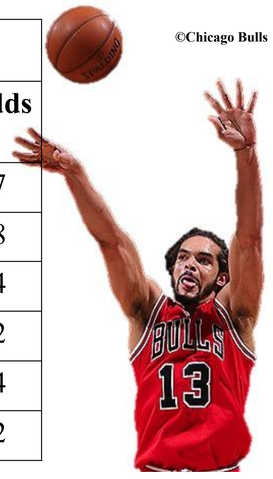
Best Shots Data Models	Train AUC	Test AUC
GBM (Partial)	0.666	0.657
Random Forest (Partial)	0.673	0.657
Logistic Regression (Lasso, Full)	0.656	0.654
Logistic Regression (Elastic Net, Partial)	0.654	0.653
SVM (Radial, Partial)	0.638	0.633
Neural Net (Full)	0.623	0.608



Shots Data Model Feature Importance	
Partial Model Feature	**Importance**
Shot Type (e.g. Jump Shot)	50.68
Shot Distance (e.g. 5 ft.)	34.03
Shot Style (e.g. Running)	11.24
Shot Angle (e.g. -15°)	1.96
Shot Court Zone (e.g. Midrange)	1.26
Shot Range (e.g. 8-16ft.)	0.65
Shot Side (e.g. Left Side)	0.17
Field Goal Type (e.g. 3 pointer)	0

©Golden State Warriors

Regularized Logistic Regression Player Coefficients			
Highest Player Coefficients	**Log Odds Impact**	**Lowest Player Coefficients**	**Log Odds Impact**
Carl Landry	1.453	Tony Wroten	0.527
Karl-Anthony Towns	1.369	Joakim Noah	0.608
Jared Dudley	1.335	Terry Rozier	0.624
Andre Miller	1.319	Festus Ezeli	0.702
JJ Redick	1.308	Corey Brewer	0.704
Stephen Curry	1.272	Alex Len	0.712



©Chicago Bulls

Discussion

- Basketball shots are hard to predict because they are random and are influenced by many factors
  - The Shots models performed moderately well
  - The Coordinate model greatly improved accuracy
- The Shots models were able to identify the league's best and worst shooters through player fixed effects
  - These mostly align with conventional wisdom
- Shot type and distance were by far the most influential determinants of shot success
  - After controlling for those factors things like shot angle and zone were not very important
  - Player and team effects had minimal predictive impact

Future Work

- Implement Recurrent Neural Network on "Coordinate" data to capture relational information in features
- Extract defender position from raw coordinate data to include defender distance measures in "Shots" data models
- Improve "partial" model accuracy sufficiently so model residuals by player capture true shooting skill