

Player Compatibility in the NBA

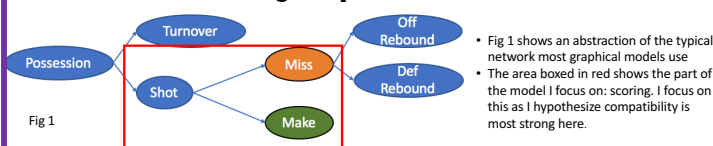
By Harrison Chase (Kensho Technologies)

All code for analysis and data
fetching/preprocessing:
<https://hwchase17.github.io/sports/>

Motivation

- Single number metrics (Plus-Minus variants) do not account for player compatibility
 - General ridge regression set up does not allow for easy estimation of interaction terms between players due to the sheer number of interaction terms (and corresponding increase in degrees of freedom) that would have to be introduced.
- Graphical methods rarely attempt to account for a player's influence on teammates
 - Closest: Joseph Kuehn, Sloan Sports Conference 2016, *Accounting for Complementary Skill Sets When Evaluating NBA Players' Values to a Specific Team*
 - Even the above player does not estimate player specific effects.

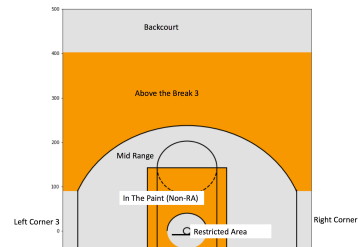
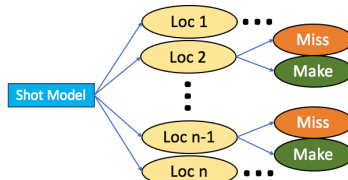
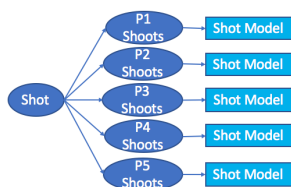
Existing Graphical Models



$$I \text{ model } P(\theta_i | c_i) = \sum P(\theta_i | l_n, c_i) P(l_n | c_i)$$

Terminology

- θ_i = Points scored on possession i
- s_i = Shooter on possession i
- $o_{i1}, o_{i2}, o_{i3}, o_{i4}$ = Offensive Teammates on possession i
- $d_{i1}, d_{i2}, d_{i3}, d_{i4}, d_{i5}$ = Defensive opponents on possession i
- Let $c_i = s_i, o_{i1}, o_{i2}, o_{i3}, o_{i4}, d_{i1}, d_{i2}, d_{i3}, d_{i4}, d_{i5}$
- Want to model $P(\theta_i | c_i)$
- Divide court into n locations, l_1, \dots, l_n
- Can also add position dummies for players on court
 - Guard, Forward, Center
- Caveat: do not model fouls at all

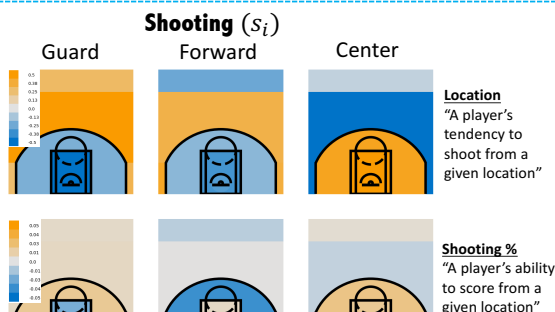


Estimate:

- Player's {tendency to shoot, ability to score} from court locations
- Player's influence on teammates' {tendency to shoot, ability to score} from court locations
- Player's influence on opponents' {tendency to shoot, ability to score} from court locations
- Position averages for all the above (See visualizations to right)

Novel Features

- Allows for independent estimates of a player's tendency to shoot from a location & their ability to score from there
- More granular court locations than usual
- Inclusion of position dummies, and estimation of their coefficients



Metrics

- Parameters:
- Season: 2017-18
 - Train set is first 1100 games (189K shots)
 - Test set is last 130 games (23K shots)

- For the purposes of equal comparison, all models are compared using multiclass logloss.
- The classes which the loss is calculated on are the most abstract outcomes of a shot:
 - Zero points scored
 - Two points scored
 - Three points scored
- When the modeling more granular outcomes than just these three, those are aggregated up to these three classes

Future Work

- Utilize x,y coordinates of shot
- Combine location tendencies/shooting ability into one visualization
- Model who will shoot
- Combine in larger graphical model

Model Features	Model Target/s	0-PT logloss	2-PT logloss	3-PT logloss	AVG logloss
None (Baseline)	Miss				
	Made 2 pointer				
For estimate use train set means	Made 3 pointer	0.6896	0.6393	0.3683	0.5657

Conditional on knowing the shooter:

Model Features	Model Target/s	0-PT logloss	2-PT logloss	3-PT logloss	AVG logloss
Shooter dummy	Location (multinomial, n=7) For each location fit separate make/miss classifier	0.6856	0.6144	0.3427	0.5475
Offensive teammate dummies					
Defensive opponents dummies					
Positional dummies					
Shooter dummy					
Offensive teammate dummies					
Defensive opponents dummies					
Positional dummies	Multinomial - Miss, Made 2pt, Made 3pt	0.6931	0.6201	0.3489	0.554
Shooter dummy	Location (multinomial, n=7) For each location fit separate make/miss classifier	0.6859	0.6155	0.3435	0.5483
Offensive teammate dummies					
Defensive opponents dummies					

"Shot Model" - the focus of this poster

Can you just model Miss/Made 2 pts/ Made 3 pts directly?

Do player position dummies help?

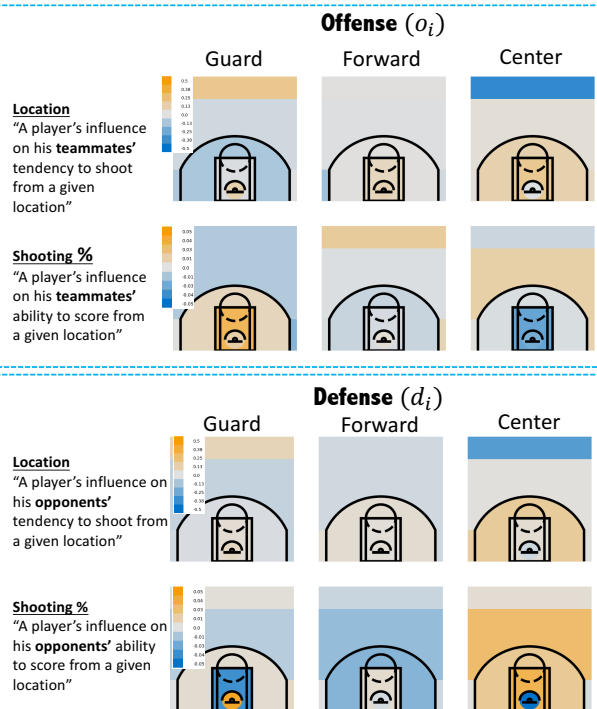
Not conditional on knowing the shooter:

Model Features	Model Target/s	0-PT logloss	2-PT logloss	3-PT logloss	AVG logloss
Shooter dummy, offensive teammate dummies, defensive opponents dummies, positional dummies	Location (multinomial, n=7) For each location fit separate make/miss classifier	0.6893	0.6386	0.367	0.565
Offensive player dummies, defensive opponents dummies, positional dummies	Multinomial - Miss, Made 2pt, Made 3pt	0.692	0.6395	0.3682	0.5666

To estimate P of each player shooting uses naive 20%, from there uses shot model

Models outcome directly (worse than baseline!!!!)

Inference



Case Study: LeBron James

Which players would help LeBron score the most efficiently (controlling for position)?

Top Players:

- DeMarcus Cousins
- Kevin Love
- Marc Gasol
- Dirk Nowitzki
- Draymond Green

Shooting Location

Shooting %

Bottom Players:

- Joel Bolomboy
- Andre Roberson
- DeAndre Jordan
- Salah Mejiri
- Rudy Gobert

Big men who can space the floor

Players who can't space the floor

Which players would LeBron help score the most efficiently (controlling for position)?

Top Players:

- DeAndre Liggins
- Wesley Johnson
- Arron Afflalo
- Rodney McGruder
- Jason Terry

Offensive Location

Offensive Shooting %

Bottom Players:

- Tyson Chandler
- Andre Roberson
- DeAndre Jordan
- Salah Mejiri
- Rudy Gobert

Corner 3 specialists

Can't shoot