NBA Player Type Clustering by Position Reflecting the Modern Basketball Trend

Namtaek Kwon

Finite Mixture Models

SungKyunKwan University

Nov 08, 2022

Outline

- Introduction
- Data
- Analysis
- Future Works

Introduction of Final Project

Traditional basketball has as few as three or as many as five positions.

- Guard (1.Point Guard, 2.Shooting Guard)
- Forward (3.Small Forward, 4.Power Forward)
- Center (5.Center)

Positional restrictions are gradually fading in modern basketball, and an increasing number of players are filling multiple roles.

- Dual Guard (1+2), Swing Man (2+3), Stretch Forward (3+4), etc...

We try to subdivide player types for each position using clustering methods, rather than simply limiting to existing roles.

Modern Basketball Trends

- Small lineup
  - The Twin Towers strategy is unpopular.
  - They play fast basketball with an undersized big man in the center position.
- A versatile player
  - A three-point shot has become a skill that not only guards but also forwards and centers should possess.
  - The center is responsible for not only scoring at the rim, but also for overall game coordination, assists, and jump shots.

To reflect the above trend, clustering of player types for each position is conducted.

Small Lineup

(a) S.O'Neal - 216cm

(b) Bam Adebayo - 206cm

(c) D.Green - 198cm

Figure: Centers become small

Versatile Players

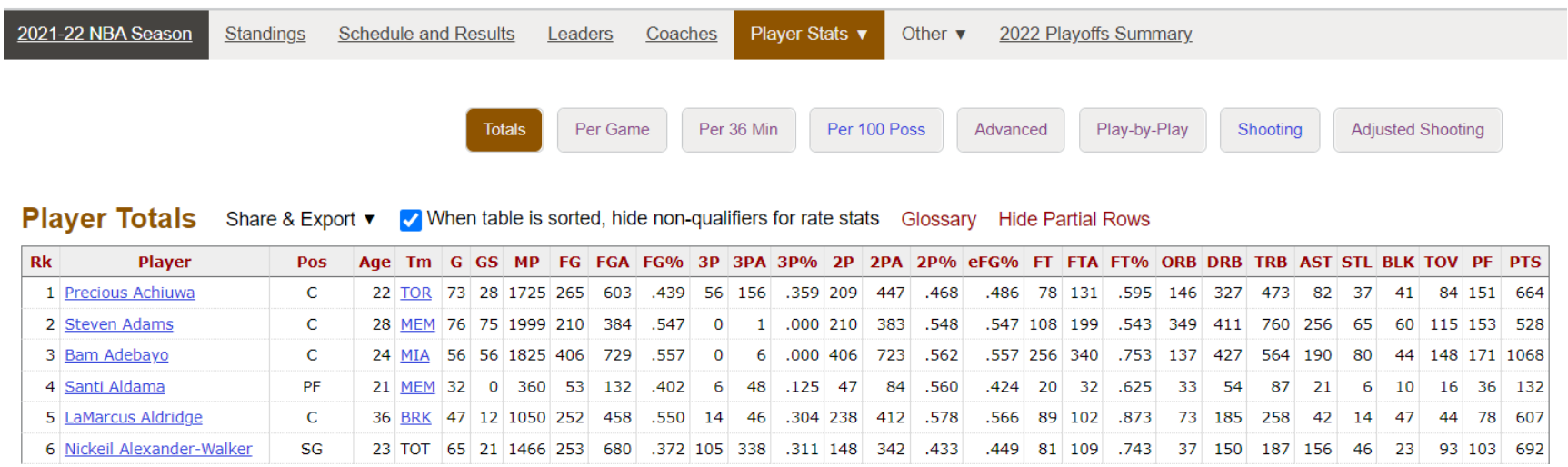
(a) Stephen Curry - main ball handler & the best 3 point shooter

(b) Nikola Jokic - center with high BQ & back-to-back season MVP

DATA

2021-2022 season NBA player stats

- Basketball-Reference.com
  - Player stats for totals, per game, also advanced indices (ex. VORP).
  - There are so many indices, we should select the variables or reduce the dimensions.



2021-22 NBA Season | Standings | Schedule and Results | Leaders | Coaches | Player Stats ▾ | Other ▾ | 2022 Playoffs Summary

Totals | Per Game | Per 36 Min | Per 100 Poss | Advanced | Play-by-Play | Shooting | Adjusted Shooting

Player Totals Share & Export ▾ ☑ When table is sorted, hide non-qualifiers for rate stats Glossary Hide Partial Rows

Rk	Player	Pos	Age	Tm	G	GS	MP	FG	FGA	FG%	3P	3PA	3P%	2P	2PA	2P%	eFG%	FT	FTA	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS
1	Precious Achiuwa	C	22	TOR	73	28	1725	265	603	.439	56	156	.359	209	447	.468	.486	78	131	.595	146	327	473	82	37	41	84	151	664
2	Steven Adams	C	28	MEM	76	75	1999	210	384	.547	0	1	.000	210	383	.548	.547	108	199	.543	349	411	760	256	65	60	115	153	528
3	Bam Adebayo	C	24	MIA	56	56	1825	406	729	.557	0	6	.000	406	723	.562	.557	256	340	.753	137	427	564	190	80	44	148	171	1068
4	Santi Aldama	PF	21	MEM	32	0	360	53	132	.402	6	48	.125	47	84	.560	.424	20	32	.625	33	54	87	21	6	10	16	36	132
5	LaMarcus Aldridge	C	36	BRK	47	12	1050	252	458	.550	14	46	.304	238	412	.578	.566	89	102	.873	73	185	258	42	14	47	44	78	607
6	Nickeil Alexander-Walker	SG	23	TOT	65	21	1466	253	680	.372	105	338	.311	148	342	.433	.449	81	109	.743	37	150	187	156	46	23	93	103	692

VARIABLES

table

Table: Description of Variables used in this Study

Variables	Descriptions
POS	Position
Age	AGE
Tm	Team
Tm	Team
G	Games
MP	Minutes Playes
FG,FGA,FG%	Field Goals, Attempts, Percentages
3P,3PA,3P%	3-Points Field Goals, Attempts, Percentages
2P,2PA,2P%	2-Points Field Goals, Attempts, Percentages
eFG%	Effective Field Goal Percentage - This statistic adjusts for the fact that a 3-point field goal is worth one more point than a 2-point field goal.

VARIABLES

table

Table: Description of Variables used in this Study

Variables	Descriptions
FT,FTA,FT%	Free Throw Goals, Attempts, Percentages
ORB, DRB, TRB	Offensive, Deffensive, Total Rebounds
AST, STL, BLK	Assists, Steals, Blocks
TOV, PF	Turnovers, Personal Fouls
PER	Player Efficiency Rating - A measure of per-minute production standardized such that the league average is 15.
TS%	True Shooting Percentage - A measure of shooting efficiency that takes into account 2-point field goals, 3-point field goals, and free throws.

VARIABLES

table

Table: Description of Variables used in this Study

Variables	Descriptions
WS	Win Shares - An estimate of the number of wins contributed by a player.
BPM	Box Plus/Minus - A box score estimate of the points per 100 possessions a player contributed above a league-average player, translated to an average team.
VORP	Value over Replacement Player - A box score estimate of the points per 100 TEAM possessions that a player contributed above a replacement-level (-2.0) player, translated to an average team and prorated to an 82-game season.

ANALYSIS

Implementing various model-based commonly used clustering methods.

- For each position groups (G/F/C)
- GMM with various covariance structure
- K-means, K-medoids and hierarchical clustering methods.
- comparing with clusetering evaluation metrics such as dunn index, silhouette coefficient.

Expected Results

For Center,

- 1) Typical Center, 2) Undersized Big Man Style, 3) Control Tower

For Forward,

- 1) Two-way Player, 2) Stretch Big Man, 3) Rim Attacker, 4) Superstar, 5) Swing Man

For Guard,

- 1) Main Ball Handler, 2) Catch-and-Shooter, 3) Role-Player

DIFFICULTIES

1. Some players have few playing time - Criteria for dropping obs.
2. Many Variables - How to select variabels or reduce dimension?

EM Algorithm

Incomplete likelihood is defined with summation of complete likelihood.

$$\log p(\boldsymbol{X}|\boldsymbol{\theta}) = \log \Big[\sum_{\boldsymbol{\gamma}} p(\boldsymbol{X}, \boldsymbol{\gamma}|\boldsymbol{\theta}) \Big]$$

- Summation in the log makes optimization difficult.
  - For E-step, compute $p(\boldsymbol{\gamma}|\boldsymbol{X}, \boldsymbol{\theta'})$ and take the expectation of log-likelihood of complete data.

$$Q(\boldsymbol{\theta}; \boldsymbol{\theta'}) = E_{\boldsymbol{\theta'}}\{\log p(\boldsymbol{X}, \boldsymbol{\gamma}|\boldsymbol{\theta})|\boldsymbol{X}\} = \sum_{\boldsymbol{\gamma}} p(\boldsymbol{\gamma}|\boldsymbol{X}, \boldsymbol{\theta'}) \log p(\boldsymbol{X}, \boldsymbol{\gamma}|\boldsymbol{\theta})$$

  - For M-step, find θ which maximize Q-function

$$\boldsymbol{\theta}^{new} = argmax_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta'})$$