Take 5 (to 14): Re-Defining the NBA Positional System

Daniel Brockett

Department of Sport Analytics, Falk College of Sport and Human Dynamics, Syracuse University

Syracuse University
Falk College of Sport & Human Dynamics

ABSTRACT

The aim of this study was to determine whether the current NBA Positional System properly fits player's on-court actions. Machine learning models utilize advanced tracking statistics to predict a player's position. The current NBA positional system has been around since before the creation of the 3-point line, and the game of basketball has evolved significantly, so these positions no longer accurately describe all of a player's on-court actions. New positions were created using K-Means clustering and analyzed for statistical differences to create a more accurate, functional NBA position system. Then a linear model was used to analyze whether age had a statistically significant impact on the new positions, asking the question how much, if at all, age impacts style of play.

INTRODUCTION

In any sport, a position is effectively used as a shorthand to describe what a player's responsibilities are on the playing field or court. In the sport of basketball, the positions tend to be slightly more complicated than other sports, as they each have several responsibilities. The point guard is responsible for ball handling, playmaking, and distributing. The shooting guard is typically the best shooter and scorer on the court. The small forward is often well-rounded with a balanced skill set. The power forward is a player capable of rebounding, playing solid defense, and scoring in the mid-range and close to the basket. The center is responsible for blocking shots and scoring around the rim.

These positions have been clearly defined throughout the history of basketball, and players have flawlessly filled these roles. Wilt Chamberlain was a center who dominated the game inside the paint, while Kobe Bryant was a textbook shooting guard. In the modern NBA, these same traditions are on their way out. Some of the most popular players in the league are examples of these outdated positions, like LeBron James, Nikola Jokic, Kevin Durant, and Draymond Green. These mold-breaking players could just be considered an exception, but there's a new exception every week in the NBA.

Several other papers have shown inadequacies with the NBA's current position system. To analyze whether these players are outliers, several unsupervised machine learning models were given on-court tracking data.

Often age impacts how a player operates on the court, but due to medical and training advancements, Father Time seems to have less of an impact than ever. Given that the new positions much more accurately represent a player's playing style, an analysis of the impact of age on a player's position was also done using a linear model.

METHOD

The tracking data was collected from NBA.com, as tracking statistics are available for every game dating back to the start of the 2013-14 season. The tracking data was scraped from NBA.com using a custom web-scraper written for the purpose of this research. Each player's listed position for each season was collected using the nbastatr package in R. Observations were limited to players that played at least 25% of all possible minutes in each season.

There was a relative balance among the 5 positions in terms of the number of observations that qualified, as all of the positions had roughly 600 observations.

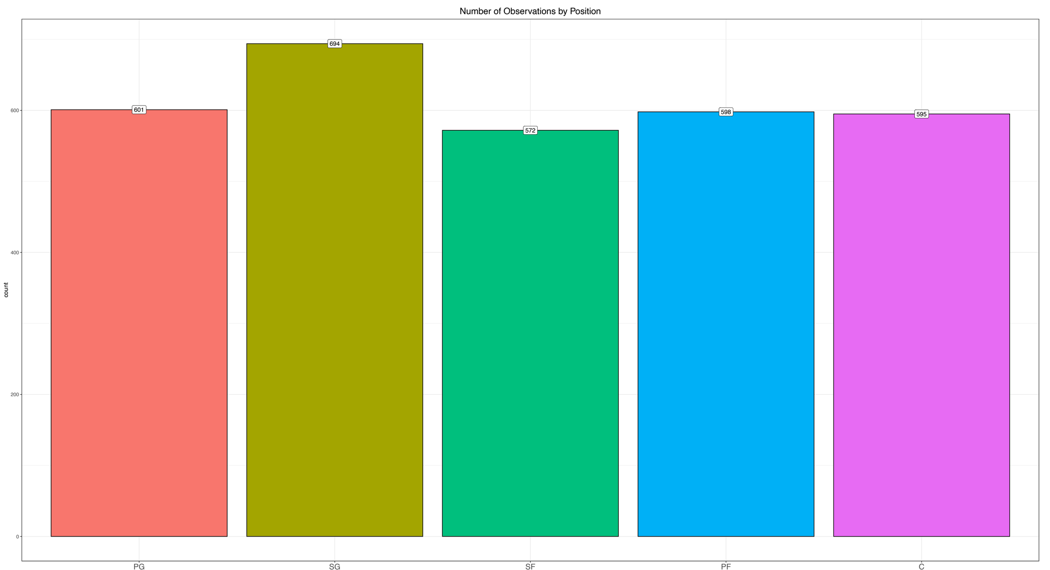

Figure 1: A bar graph showing the number of observations listed at each position.

A similar analysis was done across the seasons to see whether each season had a similar number of qualified players. There was significantly more observations in 2020-21 than any other season, with 420 observations, but that is likely due to the high number of players utilized due to the COVID-19 restrictions that forced an unusual number of players to sit out games, despite being technically healthy. Most seasons had roughly between 300-350 observations otherwise.

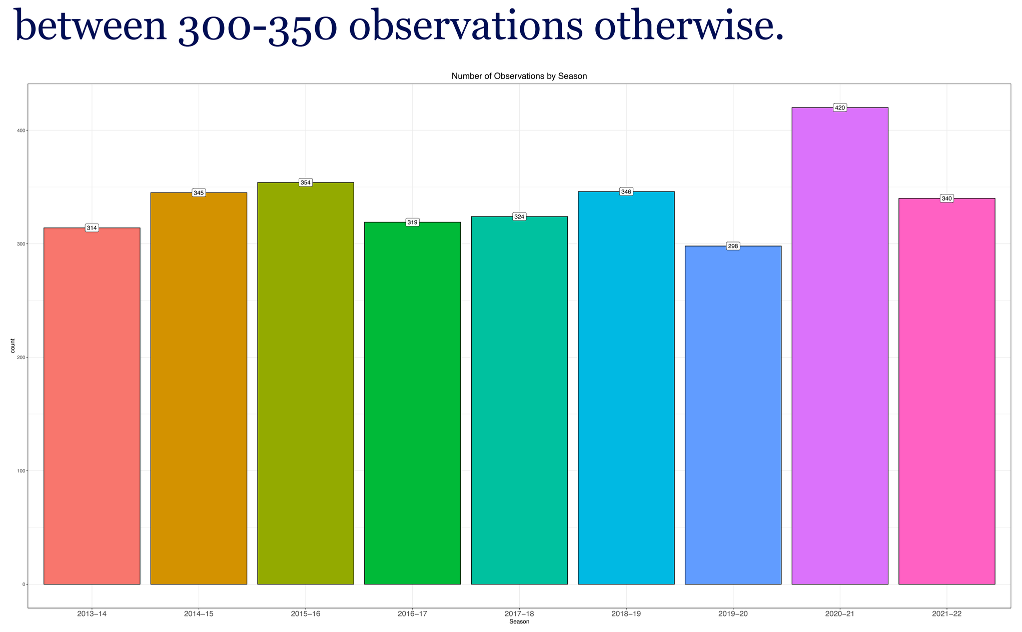

Figure 2: A bar graph showing the number of observations listed for each season.

Many of the variables in the initial data frame were repetitive or contained different variations of information, so it was sorted into 4 different data frames containing different versions of the redundant information to see what helped predict a player's position most accurately. The two most accurate data frames were the two that utilized shot totals as opposed to percentages, and these were the two used for all the models. One of the two, ***df2***, had shot totals for makes and attempts along with offense and defense splits for information like speed, distance covered, touches, and rebounding. The other, ***df4***, contained shot totals along with the averages of the offense and defense splits.

RESULTS

Of the three different machine learning techniques used, the Support Vector Machine (SVM) models posted the best accuracy for both data frames. The results of the SVM model for ***df4*** are shown below in Table 1, as this tuned SVM model produced the highest accuracy of any of the models. The most accurate model after tuning and variable importance analysis still resulted in approximately ¼ of the league's players being misclassified as another position.

Confusion Matrix:	SVM (df4)						
			Reference				
Best: Cost = 0.2			PG	SG	SF	PF	C
	Prediction	PG	178	25	0	0	0
		SG	20	162	41	1	0
		SF	0	42	110	26	1
		PF	0	0	36	123	25
		C	0	0	1	47	170
Overall Statistics:							
Accuracy	0.7371						
95% CI	(0.7088, 0.764)						
No Information Rate	0.2272						
P-Value (Acc > NIR)	2.20E-16						
Kappa	0.671						
Mcnemar's Test P-Value	NA						
Statistics By Class:							
		Class					
		PG	SG	SF	PF	C	
	Sensitivity	0.899	0.7074	0.5851	0.6244	0.8673	
	Specificity	0.9691	0.9204	0.9159	0.9248	0.9409	
	PosPredValue	0.8768	0.7232	0.6145	0.6685	0.7798	
	NegPredValue	0.9752	0.9145	0.9059	0.9102	0.9671	
	Prevalence	0.1964	0.2272	0.1865	0.1954	0.1944	
	Detection Rate	0.1766	0.1607	0.1091	0.122	0.1687	
	Detection Prevalence	0.2014	0.2222	0.1776	0.1825	0.2163	
	Balanced Accuracy	0.9341	0.8139	0.7505	0.7746	0.9041	

*Table 1: This table shows the resulting confusion matrix from the SVM model that utilized **df4**. Point Guards and Centers were the most accurately predicted positions likely due to the defining traits of each.*

The most accurate Random Forest models posted accuracies of about 72%, while the best K-Nearest Neighbor (K-NN) models only posted accuracies between 68-70%, which makes that the least accurate of the three models. Important information was provided during the K-NN models that indicates that the ideal number of groups should either be 11, 14, or 15. In fact, the K-NN models never resulted in fewer than 10 positions being the most accurate model. This along with the struggles of the models all signified that the NBA's positional system is not adequately describing what a player does on the court. The most accurate K-NN models returned 11 and 14 positions as the highest accuracies, respectively, so those were the two K's of interest when it came time to use K-Means Clustering to create the positional groups based on the tracking data.

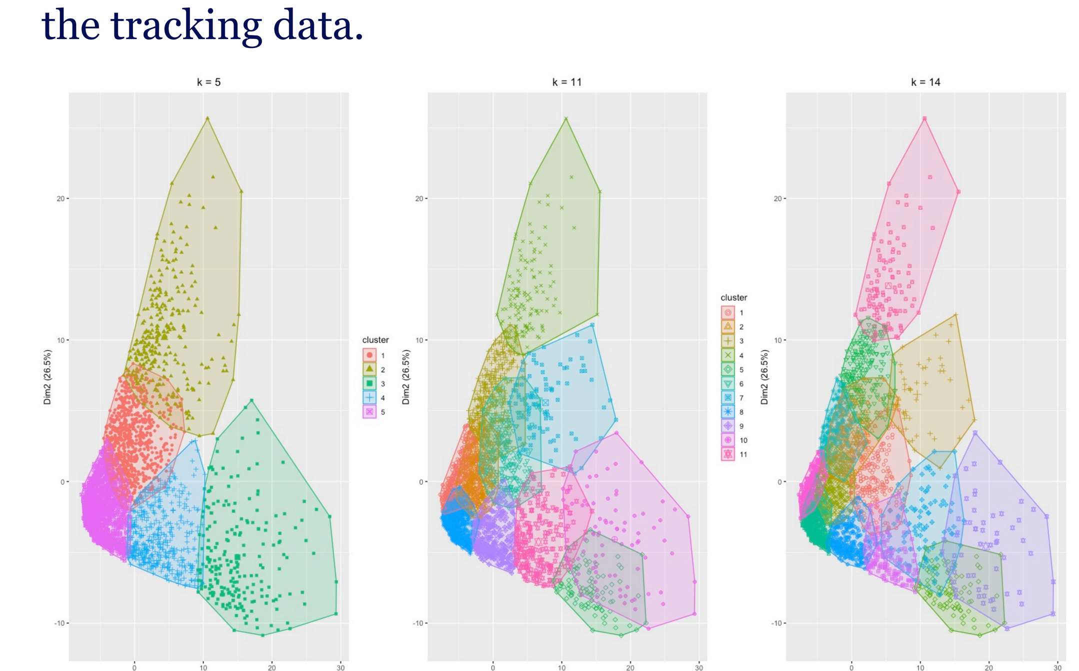

Figure 3: Cluster Visualizations depicting the results from the K-Means Clustering using dimensionality reduction.

Due to breaks at key points of congestion, and the fact that the model with k=14 was the K-NN model with the highest accuracy, the new number of NBA positions was set at 14. Then a linear model was run to determine if age along with box-score stats had any impact on differentiating a player's position.

CONCLUSIONS

If the role of a positional label is to describe how a player acts and behaves on the court, then these machine learning techniques and tracking data indicate that the NBA's positional system needs revamping. These models also indicated that the ideal number of positions is greater than 10, whether it be 11, 14, or some other number. Regardless of the new number, the current positional system only allows for the correct classification less than 75% of the time. As the game of basketball continues to evolve, these new positions will have to be updated, but these should prove to be serviceable until the game experiences a groundbreaking change, like the addition of the 3-point line.

Age had an impact on only one of the linear models. This indicated that age may not impact an athlete's style of play as much as previously thought. As training, rehabilitation, and injury prevention methods continue to develop, player's will have greater longevity, meaning age has less of an impact on their style of play. If positions accurately describe a player's style of play, then age will have less impact upon a player's position.

While many of the new positions do similar things on the court, there are important distinctions between them, and acknowledging those differences could allow for optimized lineups and roster construction.

REFERENCES

Abdelkrim, B. (2010, May). *Positional Role and Competitive-Level Differences in. . . : The Journal of Strength & Conditioning Research*. LWW. Retrieved December 4, 2022, from https://journals.lww.com/nsca-jscr/Fulltext/2010/05000/Positional_Role_and_Competitive_Level_Differences.27.aspx

Alagappan, M. (2012). *From 5 to 13*. Retrieved November 16, 2022, from https://web.math.utk.edu/~fernando/Students/GregClark/pdf/Alagappan-Muthu-EOSMarch2012PPT.pdf

An, M., Liang, E., & Zhang, M. (2016). Predicting the Trajectory of an NBA Player's Career. *Stanford University*. http://cs229.stanford.edu/proj2016/report/AnLiangZhang-PredictingTheTrajectoryOfAnNBAPlayer%27sCareer-report.pdf

Bianchi, F., Facchinetti, T., & Zuccolotto, P. (2017). *Role revolution: towards a new meaning of positions in basketball | Bianchi | Electronic Journal of Applied Statistical Analysis*. Electronic Journal of Applied Statistical Analysis. Retrieved November 16, 2022, from http://siba-ese.unisalento.it/index.php/ejasa/article/view/16464

Bosch, J., & Kalman, S. (2020, March 6). *NBA Lineup Analysis on Clustered Player Tendencies: A new approach to the positions of basketball & modeling lineup efficiency of soft lineup aggregates*. MIT Sloan Sports Analytics Conference. Retrieved November 16, 2022, from https://global-uploads.webflow.com/5f1af76ed86d6771ad48324b/5f6a65517f9440891b8e35d0_Kalman_NBA_Line_up_Analysis.pdf

Bruce, S. (2016). A scalable framework for NBA player and team comparisons using player tracking data. *Journal of Sports Analytics*, *2*(2), 107–119. https://doi.org/10.3233/jsa-160022

Chang, Y., Maheswaran, R., Su, J., Kwok, S., Levy, T., Wexler, A., & Squire, K. (2014, February 28). *Quantifying Shot Quality in the NBA - PDF Free Download*. Retrieved December 4, 2022, from https://docplayer.net/52631681-Quantifying-shot-quality-in-the-nba.html

Cheema, A. (2020, June 4). *Using Machine Learning to Classify NBA Players, Part I*. The Spax. https://www.thespax.com/nba/using-machine-learning-to-classify-nba-players-part-i/

Cheng, A. (2018, May 23). *Using Machine Learning to Find the 8 Types of Players in the NBA*. Medium. https://medium.com/fastbreak-data/classifying-the-modern-nba-player-with-machine-learning-539da03bb824

Cui, Y., Liu, F., Bao, D., Liu, H., & Zhang, S. (2019, October 22). *Key Anthropometric and Physical Determinants for Different Playing Positions During National Basketball Association Draft Combine Test*. NCBI. Retrieved November 16, 2022, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6820507/

Diambra, N. J. (2018). *Using Topological Clustering to Identify Emerging Positions and Strategies in NCAA Men's Basketball*. TRACE: Tennessee Research and Creative Exchange. Retrieved November 16, 2022, from https://trace.tennessee.edu/utk_gradthes/5084/

Duman, E., Sennaroglu, B., & Tuzkaya, G. (2021, December 30). *A cluster analysis of basketball players for each of the five traditionally defined positions*. SAGE Journals. Retrieved November 16, 2022, from https://journals.sagepub.com/doi/abs/10.1177/17543371211062064?casa_token=z3ZfL-bn81QAAAAA%3A_OCy2vN_bVk1mZxzPYjDXF4XnMQL7F0U-19atqMWuVNh8IcLGvb24cRhRWyjPCHU5in00F5eRe-7&journalCode=pipa

Jyad, A. (2020, November 16). *https://towardsdatascience.com/redefining-nba-player-classifications-using-clustering-36a348fa54a8*. Redefining NBA Player Classifications Using Clustering. Retrieved November 16, 2022, from https://towardsdatascience.com/redefining-nba-player-classifications-using-clustering-36a348fa54a8

Kalén, A., Pérez-Ferreirós, A., Costa, P. B., & Rey, E. (2020). Effects of age on physical and technical performance in National Basketball Association (NBA) players. *Research in Sports Medicine*, *29*(3), 277–288. https://doi.org/10.1080/15438627.2020.1809411

Li, B. (2021, July 8). *Application of Artificial Intelligence in Basketball Sport | Journal of Education, Health and Sport*. https://apcz.umk.pl/JEHS/article/view/JEHS.2021.11.07.005

Metulini, R. (2017, July 4). *Space-Time Analysis of Movements in Basketball using Sensor Data*. arXiv.org. https://arxiv.org/abs/1707.00883

Mills, J. (2015, August 13). *Decision-Making in the NBA: The Interaction of Advanced Analytics and Traditional Evaluation Methods*. https://scholarsbank.uoregon.edu/xmlui/handle/1794/19127

Narayan, S. (2019, June). *Applications of Machine Learning: Basketball Strategy*. MIT.edu. Retrieved November 16, 2022, from https://dspace.mit.edu/bitstream/handle/1721.1/123043/1127911338-MIT.pdf?sequence=1&isAllowed=y

Ostojic, S., Mazic, S., & Dikic, N. (2006, November). *Profiling in basketball: physical and physiological characteristics of elite players*. NCBI. Retrieved November 16, 2022, from https://pubmed.ncbi.nlm.nih.gov/17149984/

Pion, J., Segers, V., & Bourgois, J. (2018, March 29). *Position-specific performance profiles, using predictive classification models in senior basketball*. SAGE Journals. Retrieved November 16, 2022, from https://journals.sagepub.com/doi/abs/10.1177/1747954118765054?casa_token=XjAo0KBhEPQAAAAA%3A2XIKVFc6LGOo5KFKOuXWMbCuNrv6uMSfbcU5PgRYMcGFevhpN_SfZ3bj6hOKHb_X4-YkDPMU5TD7&journalCode=spoa

Rangel, W., Lamas, L., & Ugrinowitsch, C. (2019, August). *Basketball players' versatility: Assessing the diversity of tactical roles*. ResearchGate. Retrieved November 16, 2022, from https://www.researchgate.net/publication/335010180_Basketball_players'_versatility_Assessing_the_diversity_of_tactical_roles

Sandholtz, N., Mortensen, J., & Bornn, L. (2019, December). *Measuring Spatial Allocative Efficiency in Basketball*. Retrieved December 4, 2022, from https://www.researchgate.net/publication/337904719_Measuring_Spatial_Allocative_Efficiency_in_Basketball

Sarlis, V., & Tjortjis, C. (2020). Sports analytics — Evaluation of basketball players and team performance. *Information Systems*, *93*, 101562. https://doi.org/10.1016/j.is.2020.101562

Schoch, D. (2018, March 4). *Analyzing NBA Player Data II: Clustering Players*. http://blog.schochastics.net/post/analyzing-nba-player-data-ii-clustering/

Stephanos, D., Husari, G., Bennett, B., & Stephanos, E. (2021, April). *Machine learning predictive analytics for player movement prediction in NBA: applications, opportunities, and challenges*. Retrieved December 4, 2022, from https://dl.acm.org/doi/abs/10.1145/3409334.3452064

Vaci, N. (2019, January 25). *Large data and Bayesian modeling—aging curves of NBA players*. SpringerLink. https://link.springer.com/article/10.3758/s13428-018-1183-8?error=cookies_not_supported&code=21377040-6797-4877-bc76-6825a862e141

Wakim, A. (2014, March 28). *Functional Data Analysis of Aging Curves in Sports*. arXiv.org. https://arxiv.org/abs/1403.7548

Zhai, Z., Guo, Y., Zhang, S., Li, Y., & Liu, H. (2021, January 13). *Explaining Positional Differences of Performance Profiles for the Elite Female Basketball Players*. NCBI. Retrieved November 16, 2022, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7874149/

ACKNOWLEDGEMENTS

I would like to thank my faculty advisor for my thesis Dr. Rodney Paul for his assistance and guidance throughout the process of completing my undergraduate thesis. I also want to thank Dr. Jeremy Losak, Dr. Shane Sanders, and Dr. Justin Erlich for teaching me the skills necessary for this project, and the inspiration they offered through the years that allowed me to settle on this idea.