How Data Science and Analytics revolutionised the NBA

Andreas Kyratzis

University of Sussex

Contents

Abstract

The central aspect of this research is to understand the impact of Data Science and Analytics in the basketball world and specifically in the National Basketball Association (NBA). Through the various literature analysed the main focus is to understand how the game has changed through the insights from the data. The changes that erupted through the use of Data Science can be seen as two different approaches, approaches to change the ways teams select players for their rosters and changes on game time approaches to coaching. Further techniques are highlighted that perform data analysis on the statistical side of the data generated by the NBA.

Introduction

Imagine a scenario of Lebron James having the ball in his hand on the top of the three-point line with 1 second to go down by 2 points. Wouldn't it be impressive to know if the shot he would take to win the game would go in the basket giving his team the win? In the basketball world, statistics of shots and players are ways of determining good shots or bad shots and good players or bad players. However, that is where the National Basketball Association has been at for almost all its lifetime. Statistics are not enough to win games today in the era of impressive three-point shooters who can also dunk the ball in the basket or pass to the most open player on the court. Since the players have revolutionised through the years, the statistics must improve as well to make sense of all this sports' data.

During the last few years, NBA teams have employed Data scientists to make sense of all the new data that comes in about players and on-court data of games to improve the teams and the rosters. The objectives of the work of these individuals are not set in stone as there are many different types of data that is handled by an NBA team. The likes of the data that an NBA data scientist would work with can be video data from games, scoreboard spreadsheets and sensor data from body sensors or cameras. All of these different types of data are processed for the purpose of building better teams and creating better in-game coaching of the individual players and team as a whole.

Given the above example, Lebron James might have been given a pick n roll play a few seconds earlier leaving him with a defender on him that is of different position meaning of different body type; therefore, that would impact positively or negatively Lebron's shot. Having many variables like that makes the work of a Data scientist in the industry a hard but well acknowledged and respected position for a team. Many teams today credit a lot of their success to their Data scientists and analysts who deal with their data.

Movement data is a big part of the development of analytics in the sports industry. It enables plays to be thoroughly analysed to prove how they could be better defended or for the opposite purpose; how the team on offence could score easier or more efficiently. In the following chapter, the focus will be on how teams capture movement and how they improve play using the insights of such data using data science methods.

Athletes as moving dots

Basketball players' movements on the court are mostly mined through video cameras or movement sensors spread around the court. What can be done with the data mined is what makes the area of this topic of exceptional importance. As the players on the court can be seen as moving dots through the sensory data, the best way to make sense of the dots is to find a way to make the machine understand the patterns of the moving dots.

Spatiotemporal pattern recognition

Spatiotemporal pattern recognition, which is mostly used in fields such as the military, surveillance and biology is now a trending way of making sense of the moving dots in the basketball world. In this stage is where Data scientists come in as masters of machine learning modelling. Modelling movement can be tricky as the research area in recent times tries to feed the machine with realistic basketball movements and nonrealistic basketball movements, therefore, enabling the machine to understand what a player can do on the floor and consequently finding all the patterns that may occur at any given situation as discussed in the literature (Kotleba, 2017).

Going more in-depth, extracting features such as basketball move' pick n roll' and all its combinations, the machine is able to see the game the same way as a coach and while it incrementally starts learning new moves it sees the game in more detail than the coach of a team as it lacks emotions and player favouritism. Teams are not trying to dismiss coaching masterminds such as Gregg Popovich but rather empower their knowledge thru the use of machine learning and thorough exploratory data analysis.

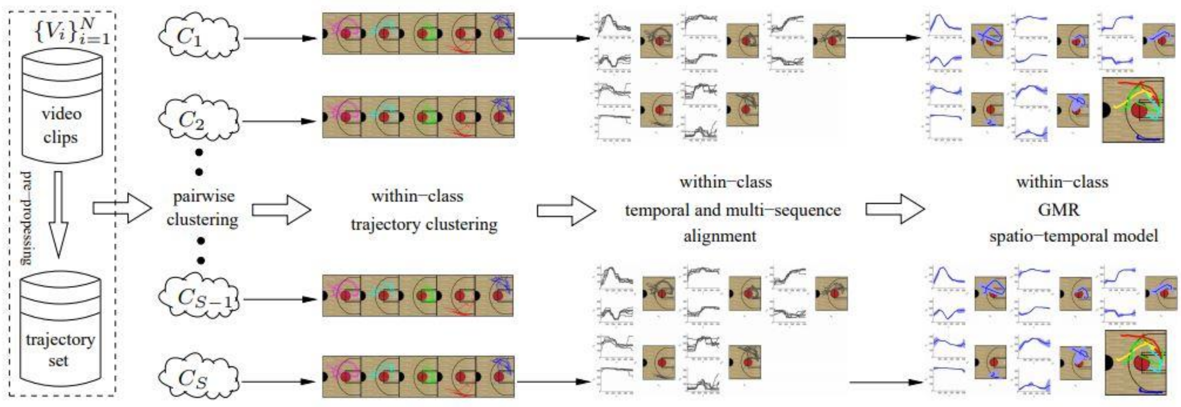

Figure 1 - Spatiotemporal modelling on NBA video clips (Chen et al, 2015)

Figure 1 (Chen et al., 2015) above shows the efforts in using Spatiotemporal Learning techniques to learn basketball offensive strategies. Their research highlights the use of video data to understand how various basketball moves are executed with the use of Gaussian mixture regression learning. The steps taken in their research are as follows; taking all video clips of basketball games and creating a trajectory set where they perform pairwise clustering to identify similar clips to each clip. Next, they use trajectory clustering in classes to identify each different type of play in the clips, continuing with identifying different combinations or sub-plays from the trajectory clustering using temporal and multi-sequence alignment in the classes and finally using Gaussian mixture regression to perform classification of all the different types of plays and their sub-categories. The above technique uses unsupervised learning to learn basketball offensive strategic plays with a classification rate of almost 90% identifying ten different clusters of strategic plays. When considering its successful learning of offensive strategies, we can see that spatiotemporal pattern recognition has very felicitous applications in the world of basketball.

Kerner (2004) uses spatiotemporal pattern recognition in traffic movement data to understand traffic congestions in freeway bottlenecks highlighting that spatiotemporal modelling is not a newly researched area as there are multiple examples of its uses in various fields while using movement data (Kurogi, 1987) (Hecht-Nielsen, 1987).

After looking at various literature on Exploratory Data Analysis of NBA data (Tjortjoglou, 2015), (Papadopoulos, 2018), it appears that working with play to play data is very hard to track variations of movement as there are 10 players on the floor and tracking each one's movement can be extremely hard as each has to identified and separated from each other's movements. Having ten moving dots plus the ball which is being moved around at all times means that there is too much work to be done if there was a data analyst working on the sideline doing data analysis of plays as they are being performed on the floor. Thus, companies like Second Spectrum are trying to revolutionise the field by using machine learning techniques to get the analysis of all plays done almost in real-time scenarios.

Finding patterns

Teams today are using analytics to search for new strategies. Analytics are used to find the value of players that their stats may not be splashy on the scoreboard but are vital for a team's success by making an impact on the floor. Scoreboards show who scored the most points and who got the most rebounds but do not show how good a player on offence plays when guarded by a particular player on defence. While the defence is a difficult area of the sport to be tracked, analysts in the NBA have worked with analysing the above example to showcase defensive superiority of players. A video explaining the aforesaid (iNerdSome, 2016), shows how a player who may not have exceptional stats on scoreboards, can still get a big contract from an NBA team allowing him to perform at the highest level.

Second Spectrum, a company which focuses on machine learning modelling using optical tracking data of the NBA is one of the pioneers in the field generating much focus on their end by their use of spatiotemporal modelling to track plays and strategies as well as individual player efficiency. The company by analysing basketball moves thru the tracking data have set up rules in their modelling that match specific patterns such as a screen or pick-n-roll to identify the individual variations of the moves to help in-game coaching. By getting outstanding accuracy on their predictions, NBA teams have agreed to work with the software solution company to develop and help coaching staff understand the players' capabilities in more depth. The company is working hard in putting new rules in their machine learning algorithm to understand each different basketball move which could provide better detail on the insights they offer to the NBA teams. Maheswaran (2015), discusses how due to their success, almost all NBA teams that head to the playoffs every year are using their services in order to win championships.

Precision and Recall

When dealing with tracking data in spatiotemporal modelling, it is essential to make some sense of the precision and recall of the metrics calculated to use a well-performing machine learning technique which will make the correct data points thrive. As seen in other applications of Spatiotemporal modelling, very high precision is required in parallel with a large amount of recall in order to get the correct patterns recognised (Dereszynski and Dietterich, 2011). Hence, league coaches and front offices are expecting the same kind of results on machine learning techniques driven by NBA data. While it is essential to gain insights of good plays and players, it is also essential to understand every single mistake in the game that could have an impact on the team's ability to play a game.

How Analytics eradicated the mid-range two-point shot

From the introduction of the three-point shot in the NBA in the late 1980s till only a few years ago, not many players where shooting three-pointers as they were inefficient in the players' and coaches' eyes. Players that were shooting three-pointers were mostly doing just that and not much else as being able to shoot three-pointers efficiently or having any other skill or position was very much separated from any other position. Today in the modern era of the game, players should be able to adhere to multiple positions and have multiple offensive and defensive skills.

The game did not just magically change. Houston Rockets' general manager, Daryl Morey, a computer scientist with a background in statistical methods, has created a trend of pattern recognition to understand the better shots that his team would need to take to bring in success on the court. By using video movement tracking on their home court, the Houston Rockets mined data that came out to be extremely valuable in changing the future of the game. By analysing the percentages of different shots by different players, it was found that the shots with the most significant likelihood of going in and having the highest return in points per possession were two-point dunks and three-point shots. Therefore, a straightforward two-point mid-range shot was found to be a non-good shot to take as it had a lower percentage of going in compared to a dunk and had less return in points when comparing with a three-point shot. That is the reason the Houston Rockets very rarely take shots inside the three-point line that are not dunks and thus have risen in the NBA leaderboards over the recent years (The Economist, 2018).

It was not years till other teams started doing the same; changing the shots players are shooting today. If it were not for data analytics tools being employed in offices of NBA teams, the players would have still been shooting inefficient two-point shots.

Maheswaran (2015), addresses how this change in gameplay has changed coaches' and general managers' views who are now looking for leaner and faster power-forwards and centres who can shoot three-point shots making their all-around offence more difficult to defend.

League independent analytics

Avalon, Balci and Guzman (2016), research on how score margins of NBA games can be predicted using machine learning techniques. In their research, they tested five different techniques, linear regression, Principal Component Analysis with Support Vector Machines, Random Forest, Adaptive Boosting and Gaussian Discriminant Analysis. After their benchmarking and testing, they found that Linear Regression, which is prone to fail with outliers in data points, has not worked well and had the worst performance of all five methods used when doing regression tasks. Outliers in NBA scoring are very frequent as teams can win any game and a situation where a bad team wins a challenging game happens daily. Another emphasised discovery by their study is that Gaussian Discriminant Analysis has exceptional performance and thrives upon understanding the distribution of the data when doing classification tasks. The best all-around modelling technique used in the study was the Principal Component Analysis with the Support Vector Machines. Literature such as above highlights that analytics based on the NBA are continuously being performed on the statistical side of the data generated by the league.

Future of Data in the industry

While coaches in the NBA are exceptional at what they are doing, the machine will eventually be able to understand the game of basketball in more detail than the average NBA coach. It will continue getting better at analysing plays and players, but its purpose is to make the game better by using the data to its advantage. If it were not for data in the NBA, Stephen Curry might not have been the star he is today as well as James Harden. The analysts of teams used what these players were bringing on the court and created a system around them which enabled them and their teams to thrive.

Pattern recognition is the focus that the general managers of teams are expecting of Data scientists, patterns that either make their team better or their team worse. In return, the coaches can later work on those patterns being avoided or being repeated at a higher frequency. It all comes down to gaining insights into the game being played, and teams are interested in that; highlighting the importance of using data science as a tool of knowledge in the NBA.

Various sports leagues and teams are starting to use and embrace data science solutions after seeing the success it has had in the NBA the past few years, unfolding the data their leagues are sourcing to make their teams better at their game. It will be soon enough that all sports leagues will be capturing data for use by their teams and general offices.

On the other hand, the data-driven revolution of the NBA is killing the careers of certain NBA players that are not fitting the standards of this new age as exploratory data analysis by the teams shows that they are not good candidates for the position they are playing as they may lack certain qualities which are found valuable through their studies (Shekhar, 2018).

Conclusion

While this is a fairly modern approach to coaching and managing basketball teams, gaining these insights today can bring championships to teams that want to use them to improve play.

Golden State Warriors which is the team with the most championships in later years credit much of their success on data analysis and their efforts in using the technology of this day to their advantage in winning NBA titles. The focus on using Data Science methods in the NBA by such teams emphasises that most if not all teams will be spending resource money int trying to keep up with teams such as the Warriors.

As long as the NBA stands, so will analytics teams in the background of it. Considering the above, there will be more job opportunities for Data Scientist roles in the area of sports and more importantly in the NBA. It is uncertain what the future may bring to technology and the NBA, but it is with certainty being stated that coaches will never cease to exist as their role is vital for a team's success.

References

Avalon, G., Balci, B. and Guzman, J. (2016). *Various Machine Learning Approaches to Predicting NBA Score Margins*. Stanford University.

Cervone, D., D'Amour, A., Bornn, L. and Goldsberry, K. (2016). A Multiresolution Stochastic Process Model for Predicting Basketball Possession Outcomes. *Journal of the American Statistical Association*, [online] 111(514), pp.585-599. Available at: https://arxiv.org/pdf/1408.0777.pdf [Accessed 14 Apr. 2019].

Chen, C., Chu, H., Liu, T., Wang, Y., Liao, H. and Tang, N. (2015). *Spatio-Temporal Learning of Basketball Offensive Strategies*. 1st ed. [ebook] Available at: https://people.cs.nctu.edu.tw/~yushuen/data/BasketballTactic15.pdf [Accessed 5 Apr. 2019].

Dereszynski, E. and Dietterich, T. (2011). Spatiotemporal Models for Data-Anomaly Detection in Dynamic Environmental Monitoring Campaigns. *ACM Transactions on Sensor Networks*, 8(1), pp.1-36.

Franks, A., Miller, A., Bornn, L. and Goldsberry, K. (2015). Characterizing the spatial structure of defensive skill in professional basketball. *The Annals of Applied Statistics*, 9(1), pp.94-121.

Hecht-Nielsen, R. (1987). Nearest matched filter classification of spatiotemporal patterns. *Applied Optics*, 26(10), p.1892.

iNerdSome (2016). *How NBA Analytics are changing the Game*. [video] Available at: https://www.youtube.com/watch?v=1F-pEeSTPkE [Accessed 9 Apr. 2019].

Kerner, B. (2004). *The physics of traffic*. Berlin: Springer.

Kopf, D. (2017). Data analytics have made the NBA unrecognizable. [Blog] *Quartz*. Available at: https://qz.com/1104922/data-analytics-have-revolutionized-the-nba/ [Accessed 12 Apr. 2019].

Kotleba, A. (2017). The Science of Moving Dots. [Blog] *Americal Mathematical Society*. Available at: https://blogs.ams.org/mathgradblog/2017/02/25/science-moving-dots/ [Accessed 6 Apr. 2019].

Kurogi, S. (1987). A model of neural network for spatiotemporal pattern recognition. *Biological Cybernetics*, 57(1-2), pp.103-114.

Miller, A., Bornn, L., Adams, R. and Goldsberry, K. (2014). FACTORIZED POINT PROCESS INTENSITIES: A SPATIAL ANALYSIS OF PROFESSIONAL BASKETBALL. [online] Available at: https://arxiv.org/pdf/1401.0942.pdf [Accessed 18 Apr. 2019].

Papadopoulos, A. (2018). How to build a simple classification model to predict if it's in or rim. [Blog] *Towards Data Science*. Available at: https://towardsdatascience.com/nba-data-science-93e0314bb45e [Accessed 6 Apr. 2019].

Shekhar, A. (2018). 5 NBA stars who lost their shine because of analytics. [Blog] *Sportskeeda*. Available at: https://www.sportskeeda.com/basketball/5-nba-stars-who-lost-their-shine-because-of-analytics [Accessed 5 Apr. 2019].

TED (2015). *The Math Behind Basketball's Wildest Moves | Rajiv Maheswaran | TED Talks*. [video] Available at: https://www.youtube.com/watch?v=66ko_cWSHBU [Accessed 9 Apr. 2019].

The Economist (2018). *How data transformed the NBA | The Economist*. [video] Available at: https://www.youtube.com/watch?v=oUvvfHkXyOA [Accessed 7 Apr. 2019].

Tjortjoglou, S. (2015). How to Track NBA Player Movements in Python. [Blog] *Medium*. Available at: http://savvastjortjoglou.com/nba-play-by-play-movements.html [Accessed 6 Apr. 2019].