Week Seven Lecture Videos

Indiana University Bloomington: Big Data Applications

Naimesh Chaudhari: naichaud@iu.edu

Assignment 6 – November 02, 2019

1. Abstract

In this paper, I provide a brief introduction about the information presented in the week seven lecture videos. I also provide additional information on a subtopic that I am interested in from the videos, ideas that could improve some of the existing work, and finally a deeper dive into a specific sport I have great interest in.

2. Summary

We start by discussion how big data and analytics have been utilized in sports. Baseball was the early adopter of this, and other sport later noticed the success of implementing analytics and adopted the same methodology. Some of the sport discussed are Soccer, Olympics, NFL Football, Basketball, Tennis and Horse racing.

In Baseball the movie Moneyball depicted how analytics can be utilized in sports. There are several actions within a game which we can pull statistics from. We can analyze real time video, and signal from specific positions. This is defined as sabermetrics, the empirical analysis of baseball, especially baseball statistics that measure in-game activity.

Within sabermetrics some of the big data analytics discussed are PITCHf/x, HITSf/x, FIELDf/x, and Commandf/x. These analytics are highly doable due to mass amounts of clean data that baseball has for over 140 years. Some of the metrics discussed are OPS, OBP, SLG, wOBA, ERA, ERC, FIP, & UZR.

We discussed a specific issue of wins above replacement (WAR). WAR is a sophisticated sabermetrics in baseball statistic developed to sum up the extent of a player total contribution to their team. In war a replacement level player is defined as contributing 20.5 fewer run than a player of league-average performance.

In unit two we start discussing advanced sabermetrics. Vince Gennaro introduced a pitcher clustering method. The model data takes in 12 different properties/features of a pitcher. The pitchers are split in left and right-handed. We can use this model to profile pitchers against batters to select the best option. Some of the factors used to decide are pitching style, pitching quality, hitting style, hitter quality, and ballpark effect. Analyzing these factors is extremely important due to the huge ROI that can be achieved. One win is worth close to five million dollars.

In unit three we discuss sports wearable and other Sports like soccer, NFL, NBA, Tennis and Horse racing. Some of the available wearable are wristbands, google, gold swing analyzer, jawbone, google glass, and Babolat play & connect. Within soccer big data analytics is used to create field visualization. For dynamic ball sports we can embed sensors within to ball to do real time analytics. Olympics does not use big data analytics well and has a huge opportunity to do so. Within NBA and NFL spatial visualization is used heavily. NFL also uses video to analyze plays and predict next ones. NBA for each player

uses shooting location success. Within tennis spatial visualization is used to see the ball trajectory hit by each player. Within horse racing tags are attached to horse to analyze their performance.

3. Interested Subtopic

I am very excited to see that analytics is not only being adopter in typical business but also in sports. Basketball is one of my favorite sports, so I wanted to research this further. Randerson112358 states that "*each years the basketball leagues host a hackathon to allow them to get some new great ideas and find new data analysis talent."*

Sports franchises such as the Golden State Warriors have openly credited their success due to their analytics team. Common practices such as resting players was suggested by their analytics team. They were able to analyze a player's performance based on how rested they were. Watching Basketball games, we see this practice utilized more often by coaches. Star players are given some rest time so during the important games they can perform at optimal levels.

In 2009 the league started collecting data though a state-of-the-art video system that captures players movements on court with or without the ball. This collected data allowed data scientist to come in and develop models that can analyze how important a player is to a win. Due to its heavy investment into analytics, Randerson112358 states that, "*there might come a time when the NBA surpasses MLB in the amount of analytics they use*."

One of the major coincidences between the Golden State Warriors (team that lead the charge in advanced analytics in NBA) and the Oakland Athletics (team that led the change in advanced analytics in MLB) is that they are both in the same location, Oakland! The major change that analytics brought to Basketball was the amount of three points a team took. From 2017 to 2017 it rose by nearly 46%.

Other types of studies though Bayesian statics reveal additional insights, such as how well does a team's defense do when a player is on court vs not. Through these additional metrics teams are no longer restricted to always access the top overvalued players. They can strategically build their roster to give them the best probability of success based on their cap rate availability. This in turn will optimize the amount of revenue an owner can gain.

4. Ideas that can improves existing work

Within Basketball, one of the major debacles of last year was allowing Kevin Durant to play one of the major finals game when he was recovering from a calf injury. As we all know, the team took a risk through their doctors' advice, and toward the end of the game Kevin Durant suffered from a torn achilleas tendon, which is one of the most damaging injury a player can get in Basketball. Many players have never been able to come back to themselves after sustaining this injury.

What I would like to see in Basketball and many other sports is have solid models developed that monitor a player's muscles and tendons through sensors as they go though a physical therapy program. This would include how their vitals are doing, how the tendons and muscles are moving, how is this compared to when they were in optimal shape.

The model would take all this input data and should push out a probability of sustaining another injury if this player was asked to play a game tomorrow.

Additionally, if a player is asked to play the game, they should be required to wear these sensors embedded in a cotton like material that monitors how they are doing during the game. If at any point the analysts see data that is not normal for the player, they should inform the coaching staff so that the player can be pulled out.

I believe in Kevin Durant's case they would have been able to pinpoint that due to his calf not being fully recovered he is putting much more strain on the tendon. If this was not picked up during his training sessions, surely if he was wearing a monitoring device during the game there would have been anomalies that the analyst would have been able to pick up in real time and pull him out of the game. This would have saved the player and the franchise a very costly and damaging blow.

5. Video Deep Dive

Unit 3

I found unit three to be the most interesting video section for me as it talked about the sports that I am most interested in, Tennis & Basketball. In this section I am going to give a summary of all the videos in Unit 3. The video name has been stated in bold letter to allows the reader to distinguish between the sections.

Wearables

Wearables have advanced quite heavily over the last few years. When wearables are utilized the players can be analyzed in real time. Questions that can be answered are what speed they are going at, what is their heartrate, are they hydrated, how is their breathing, are they fatigued, and do they have any pain in a specific area.

Some of the major technologies and brands used in wearable are accelerometer wristbands, snow googles, golf swing analyzer, up-Jawbone, Google glasses, and Babolat play&connect. Other companies like Liveathos have invented full wearable workout gear with sensor that monitors a player's muscle utilization and fatigue. There are many in-equipment sensors that can be implemented into devices/clothing. Some of them are thermometer, gyroscope, galvanic skin response, accelerometer, hygrometer, and many more.

An interesting example is where a company embedded sensors in skis and ski shoes. They measured the orientation and acceleration on 3-axis of a skier. A gyroscope and accelerometer were utilized to capture this data. The information is sent via a 2.4 GHz RF frequency. The device runs on triple A batteries and sends back 10-25 measures per second. Due to running on battery it has a maximum battery life of 8 hours, which is reasonable for a ski day.

Soccer & The Olympics

One needs to be careful when solely making decision on analytics. Certain things like how a player motivates other can not be captured yet. Major governing bodies in soccer have this belief which is why it has not been utilized as much. Some of thoughts that support their belief are, humans are not rational, humans are risk adverse, and under pressure humans fail.

There are some companies that still use data to generate insight. One of the examples shown in the video is from a company called O2-uk, which is a mobile phone network. They utilized data to show player positions on a field as they play

during the English Premier League. Although there isn't much analytics here, it does use data to show information on a 2D plane. Due to the ball being a dynamic, there is additional stuff we can add to the ball to detect some of its statistics. These monitors are called Catapult monitors., which the players wear as well. Analytics of how a player moves with the ball is captured to gain insight from.

Adidas came out with a system that a team can wear to monitor their data. It is an open platform that measures heart rate, speed, distance, location, and acceleration. This is done through the combinations of shirts, pods, and iPad.

In Olympics there was an initiative to take gymnast and make them into aerial skiers. They utilized a social media platform to find these candidates. Due to how little data they are using this is one of major areas where analytics and make a big impact.

Spatial Visualization in NFL & NBA

Within NFL and NBA, the data collected is very qualitative. One spatial visualization showed the completion percentage of a pass a QB throws. We can see that passes thrown to the sides and back within a short proximity of the QB have much higher probability to be completed then passes that are through the middle or "hail marries." There was also a second visualization shown that depicted where all QB's pass most frequently. NFL also uses computer vision and machine learning to classify plays and predict next ones. The system can recognize players which allows it to utilize additional data to make predictions.

In NBA, shooting percentages are tracked for players. They measure how likely a player can make a shot from a specific location. The size in the visualization represented frequency and color represented the accuracy. Another visualization shows the locations where shooting attempts are made. We can clearly see that the middle and 3 point line areas are prioritized.

We can utilize player shooting locations and build a team that covers all the important areas of shooting. This can allow teams to maximize their point generation.

Tennis & Horse Racing

One of the things that regularly gets tracked in tennis is the ball trajectory. This information is put on a 3D chart to show speed, height, and trajectory. This information also allows us to measure stroke pattern frequency. The service that tracks this information is called Hawk-Eye. Which has been utilized since 2005. One of the major issues with Tennis is that the governing bodies tend to keep the collected data secretive. This is a major issue and opportunity as there are not many analytics services with tennis and can be an area of huge improvement.

In horse racing, a system called Trakus tracked horse location during a race. It is attached to a horse's saddle and weighs 2.8 ounces. It is believed to be more accurate than GPS positioning. This information is over-layed on the screen to allow viewers to see the exact positioning of a horse.

Tennis: How data can be used to improve the game.

As stated in previous videos, Tennis is one of the sports that does not do a great job of releasing the data it collects. This in my opinion hinders the sport from benefiting from crowd sourced knowledge, that can

greatly improve the type of analytics they utilize.

I chose Tennis as my subject to discuss because I played competitive Tennis during my high school & college years. I also trained many individuals ranging from ages 10-50. It is a sport I am extremely passionate about and think can benefit greatly from advanced analytics. The ideas I will discuss generally are from the coach's perspective, because they are the ones that understand, motivate, and improve a player the most. I will split the suggestion into two sections before a match and during the match.

Before a Match

Although this is being utilized in many other sports, I do not see it being done regularly in Tennis. The open platform that Adidas developed to track a player's vitals during practice could be very beneficial. A coach is always concerned about the amount of effort his players are giving during practice, currently their option is to make an educated assessment. Having the Adidas platform in Tennis would allow them to track their players and make data driven suggestion to players to improve their ability.

Other things coaches can do with this information is compare it to top athletes that have gone through their program. This would allow both coaches and players to understand where their game is compared to other top athletes and create strategies to improve in areas they lack. One examples of this would be developing a specific training regimen for a player based on his vitals data. If a player has great endurance but weak lateral movements, we can focus on that, if a player has great lateral movement but no endurance, we can focus on that.

Companies like Babolat have taken the lead in racket analytics, by developing a racket that tracks the swing path & the amount of topspin a player puts on the ball. This sort of information is invaluable to a player as he is trying to become the best player he can. A combination of vital and racquet data presented in a understandable manner can definitely be a big win for the sport.

During the Match

Although we can analyze a player's strengths and weakness from their historical data, I think it is equally important to be able to asses this during a match. Players can change, develop different strategies, and do well in different areas during different days. A simple example of this would be one day a player may be returning great but another day he might not be.

If we can use live video data to analyze the players tendencies and success rates in different areas of the game, we can utilize this information to coach our players during a water break as to where they should hit and what weakness of the other player they should try and exploit. It is true that both side of competition will have this data to work off. I do not think that is an issue as this will force player to become better versions of themselves. The sport will evolve and overall will generate better competition!

5. References

1 randerson112358. "How The NBA Uses Data & Analytics." *Medium*, Medium, 22 May 2019, https://medium.com/@randerson112358/how-the-nba-uses-data-analytics-6eac3c43a096.