PAPER 3 - BDAA FALL 2016

Carlos Sathler
Graduate Student, MS in Data Science
Indiana University at Bloomington
cssathler@gmail.com

ABSTRACT

In this paper, we answer two questions about sports analytics, based on videos watched in week 3 of the BDAA Fall 2016 class.

1. BIG DATA (VIDEO) AND SMALL DATA: WHICH OFFER MOST OPPORTUNITY ON WHAT SPORTS?

Small data analytics is more suitable for application in sports in which analysts can model the game using a smaller number of features. For example, in baseball, it is possible to build a good representation of an actual game through a relatively small number of statistics [1]. These statistics can be captured at predefined breaks during the game action, and provide a good representation of the game development from beginning to end, when one team is finally victorious. One key aspect driving game development in baseball is pitcher vs. batter match-up. Many significant statistics are captured about the performance of these two types of players, and how they match against each other. There are also many ways in which these players can be categorized such that it is possible to match categories to maximize success. A simple, hypothetical example would be to categorize a pitcher as a "fast ball" pitcher and match him with a batter that's in the "slow hitter" category. All-in-all, for its characteristics baseball presents a good opportunity for the development of predictive models based on "small data" analytics.

Another sport with the potential to benefit from small data analytics is cricket, which is similar to baseball and also pairs a pitcher with a batter. In tennis, even though players are paired against each other there is a lot more variation as to where a player hits a ball and where the ball goes in the opponent's field. By contrast, in baseball, pitcher and batter are always located in the same place relative to each other. So tennis does not lend itself well to small data analytics.

The sports that offer most opportunity for "big data" analytics (analytics based on analysis of video, in the context of this paper) are the ones that cannot be modelled easily with a relatively small number of features. A good example of a sport in that category is soccer. Soccer is an extremely fluid game in which 22 players move rather freely inside a large field of play, and where the ball travels unconstrained in pretty much all directions. There are few statistics in soccer to begin with (shots on goal, passes taken, passes missed, corners, and a few others) and they have very limited power to capture the details of the game action or game development. Additionally, unlike baseball, there is no single type of interaction between players during the game which drives game development and ultimately game result. So it's not easy to create a model of the game and the game development, and predict game results, based on that model using small data.

What can be done instead, is analytics of a qualitative nature, which does not lend itself to prediction of game results, but can help coaches organize their teams based on teams' and players' moving patterns, for example. Video analysis is also used in sports like football, to help coaches identify most common completion zones or interceptions. Of course if this data is mapped to players, coaches can make decisions about team composition and play strategy. Video analysis is often done with images projected to the ground plane, which is telling of how much of the game is left out from analysis. Other examples of video analysis from the class lecture include basketball court maps of player scoring performance and three dimensional images of tennis balls trajectories during a game of tennis.

Video analytics for sports like soccer and basketball offer great potential for qualitative analytics but not so much for predictive analytics. These games are too dynamic in nature, have too many variables, and decisions during the game are too fluid and dispersed across too many players with a high potential to impact the game. In soccer, "set plays" could be an exception. With enough data, I can imagine it might be possible to feed enough video data to a program to predict the result of a particular "set play" against a given defense. Perhaps the same applies to recurring match-up situations that happen from a fixed position, such as a "tip-off" in basketball or a "scrum" in rugby.

Interestingly volleyball appears to have some characteristics common to both baseball and less structured games. It has rotations and match-ups at the net, server vs. defense match-ups, and yet it is a very fluid multidimensional sport as well. Volleymetrics [9] is a company that seems to be tapping on the potential of both qualitative and quantitative analytics in volleyball.

In conclusion, there are plenty of opportunities for big data qualitative (video) analytics in sports like basketball and soccer. Perhaps we will even see some predictive analytics done for these sports in the future based on video analysis of recurring game situations that start from a fixed state

(set-plays and tip-offs). But the potential for quantitative analytics to drive success in sports like baseball and cricket will most probably **not** be surpassed.

2. HOW IS BIG DATA AND LITTLE DATA APPLIED TO THE OLYMPICS 2016?

The blog entry "Changing the Game: Big Data Helps Bring Home Olympic Gold" [2] in the Sports Analytics Conference MIT Sloan website is a good starting point to explore the impact of big data, little data and analytics in the summer Olympics.

According to the blog entry [2], National Olympic Committees all over the world have improved their approach to athlete selection and preparation, investment decisions and competition selection, through the use of data analytics. The blog entry mentions a product called Podium [6] which aggregates data for a multitude of Olympic sports, for most countries in the world. The data goes back to 1896, the year of the first Olympic games.

It is not clear if the "most extensive sports database in the world" [6, p. 1] can be considered "big data". From the perspective of this assignment I would say not, since there is no indication Podium [6] offers analytics solutions using video footage from sports events or athletes' performances.

Another example of small data in the Olympics comes from the Forbes article "How Can Big Data And Analytics Help Athletes Win Olympic Gold In Rio 2016?" [3]. It describes how the British rowing team uses analytics against data collected on athletes' training activities and performance to make decisions around athlete selection and training strategies. The article also highlights the importance of historic data to drive decisions today; what can we learn from old successful athletes' data?

A very good example of big data in the Olympics comes from the US cycling team [5]. The team used a solution developed by IBM [4] to improve the performance of "team pursuit", an Olympic track cycling event. The solution uses sensors, including wearable sensors, power meters and health data monitors, and employs real time data transmission to the cloud, where analytics is performed on the data and transmitted to coaches' tablets for decision making at individual and team level, on the fly. It is a fascinating state of the art solution that combines, IoT, big data, analytics and the cloud.

Finally, there is little to be found online specifically about the use of qualitative use of video analytics in the Olympics. However, it can be inferred from the article "Why Analytics Matter for the U.S. Olympic Men's Basketball Team" that it was indeed used [10]. In this article Jerry Colangelo, the US basketball managing director for the 2016 Olympic team speaks of how important shot-charts would be, among other analytics tools, for the success of the US team in the competition.

A couple other ways in which big data was applied to the Olympics include pervasive use of sensors [7] and video analytics for surveillance [8].

3. REFERENCES

[1] Baseball statistics - wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Baseball_statistics. (Accessed on 09/08/2016).

[2] Changing the Game: Big Data Helps Bring Home Olympic Gold - MIT Sloan Analytics Conference News. http://www.sloansportsconference.com/mit_news/changing-gamebigdatahelpsbringhomeolympicgold/. (Accessed on 09/09/2016).

[3] How Can Big Data And Analytics Help Athletes Win Olympic Gold In Rio 2016? http://www.forbes.com/sites/bernardmarr/2016/08/09/how-big-data-and-analytics-help-athletes-win-olympic-gold-in-rio-2016/3/#4da861f27369. (Accessed on 09/08/2016).

[4] IBM Emerging Technologies - jStart - Portfolio - USA Cycling. http://www-01.ibm.com/software/ebusiness/jstart/portfolio/usacycling.html. (Accessed on 09/09/2016).

[5] IBM Is Helping The U.S. Women's Cycling Team Prepare To Ride Past The Competition At The 2016 Rio Olympics. http://www.sporttechie.com/2016/03/28/analytics/ibm-helping-u-s-womens-cycling-team-prepare-ride-past-competition-2016-rio-olympics/. (Accessed on 09/09/2016).

[6] Podium the Performance Manager. http://www.infostradasports.com/podium/. (Accessed on 09/09/2016).

[7] Rio Olympics: Sensors, big data to be the next revolution. http://www.rcrwireless.com/20160803/americas/rio-olympics-sensors-big-data-tag5. (Accessed on 09/08/2016).

[8] UAV and Video Analytics Technologies for Olympics Safety | iHLS Israel Homeland Security. http://i-hls.com/2016/08/uav-and-video-analytics-technologies-for-olympics-safety/. (Accessed on 09/09/2016).

[9] VolleyMetrics | Smart analytics for volleyball. http://volleymetrics.com/. (Accessed on 09/08/2016).

[10] Why Analytics Matter for the U.S. Olympic Men's Basketball Team. http://theseason.gc.com/why-analytics-matter-for-the-u.s.-olympic-mens-basketball-team. (Accessed on 09/09/2016).