GUEST EDITORIAL

Sports Analytics in the Era of Big Data: Moving Toward the Next Frontier

Renato Assunção¹ and Konstantinos Pelechrinis²

Data and analytics have been part of the sports industry from as early as the 1870s, when the first box score in baseball was recorded. However, it is only recently that advanced data mining and machine learning techniques have been utilized for facilitating the operations of sports franchises. One of the main reasons for this is the technological advancements that have allowed us to collect more fine-grained data. In particular, optical or sensor-based player tracking data allow us to know the location of every player on the field, court or rink at every point in time. These data basically capture everything that happens during the game, and when appropriately analyzed, they can provide valuable insights to coaches. However, apart from sports clubs, other stakeholders in the industry (the leagues' offices, media, etc.) invest in analytics. For example, the leagues increasingly rely on data to decide on potential rule changes.

In this special issue, our goal has been to collect research contributions and perspective articles for the design, development, and evaluation of methods and applications in the area of sports analytics as they pertain to both the business side and (pre or in) game strategy and sports science in general. Although some of the contributions might not have a traditional angle of big data that readers of the *Big Data* journal might expect, these articles deal with novel applications that have traditionally completed without the use of data. This special issue is divided into two parts, each containing five articles. This first part of the special issue includes articles that deal with sports such as racing to topics pertaining to betting and fantasy sports.

For example, in the article "Effects of Pacing Properties on Performance in Long-Distance Running," Arie-Willem de Leeuw et al. used publicly available data from ~120,000 races—10 km and half and full marathons—to analyze the primary pacing profiles for runners based on a set of covariates—for example, professional versus recreational, gender, and age. The au-

thors further consider the various pacing properties together to identify the main characteristics that distinguish fast finishers from underperformers.

The recent ruling of the U.S. Supreme Court in May 2018 that struck down federal laws prohibiting sports gambling has sparked the interest of the research community in topics pertaining to the area. In this first part of the special issue on sports analytics, Gary Smith and Andrew Capron in their article "Overreaction in Football Wagers" examined how gamblers might ignore reversion to the mean when it comes to team performance and overreact to recent score fluctuations. Using data from 1993 to 2017—for both the point spread and the over/under bet—the authors find that gamblers do not take regression to the mean into account. This is reflected in the closing lines sets by the bookmakers, and consequently a simple strategy where one bet on the team that has not been successful at covering the spread can be profitable. In the case of over/under bets, this would mean betting on the under when the two teams have been scoring way over their total lines (and vice versa).

The next article in this special issue is "Visualizing a Team's Goal Chances in Soccer from Attacking Events: A Bayesian Inference Approach" by Gavin Whitaker, Ricardo Silva, and Daniel Edwards. This article examines how to model soccer chance events during a game using features capturing the local context during the chance event. The authors use a sophisticated hierarchical probabilistic model to describe how the probability of a scoring chance varies. The building block is a Poisson random variable that depends on several observed and unobserved or latent factors, as well as contextual factors (position on field, ball controller, etc.). The inference of the model utilizes the Gibbs sampler algorithm and prior distribution choices made by the authors. The rest of the article shows how the model can be used to extract rich information about chance events in soccer.

¹Universidade Federal de Minas Gerais, Brazil.

²University of Pittsburgh, Pittsburgh, Pennsylvania.

Boris Bačić and Patria A. Hume wrote “Computational Intelligence for Qualitative Coaching Diagnostics: Automated Assessment of Tennis Swings to Improve Performance and Safety.” They are motivated by the spread of wearable computing devices integrated into augmented coaching systems and technology that find applications in the domains of physical activity, sport, rehabilitation, exergames, and health care. In this work, they describe a proof-of-concept prototype system for the assessment of tennis swings. The authors use recorded data of tennis swings to investigate whether a machine learning algorithm can learn to classify three coaching rules: stance, “low-to-high,” and swing width into labels from good to bad. They present a graphical user interface that can give simple visual cues about the quality of the swing to a user. Such a system would certainly be a very interesting and potentially valuable to players and coaches.

Finally, fantasy sports are an integral part of fan experience in today’s sports world. Urbaczewski and Elmore

contributed a perspective article in this special issue entitled “Big data, efficient markets, and the end of daily fantasy sports as we know it?” In this article, the authors make the case that the availability of big data and computing power will change the landscape of daily fantasy sports markets dramatically when viewed from the lens of the Efficient Market Hypothesis. This amount of information and computing power can provide an even larger skew in the distribution of earnings from daily fantasy sports. They bring analogies and comparisons with other markets—casino gaming to stock trading—that went through similar transformations, while they also provide some propositions on how the daily fantasy sports market can be *protected*.

We hope you will enjoy reading the articles published in this special issue!

Cite this article as: Assunção R, Pelechris K (2018) Sports analytics in the era of big data: moving toward the next frontier. *Big Data* 6:4, 237–238, DOI: 10.1089/big.2018.29028.edi.