International Journal of Decision Support System Technology
Volume 14 • Issue 1

Machine Learning on Soccer Player Positions

Umberto Di Giacomo, Università degli Studi del Molise, Italy*
Francesco Mercaldo, Università degli Studi del Molise, Italy
Antonella Santone, Università degli Studi del Molise, Italy
Giovanni Capobianco, Università degli Studi del Molise, Italy
https://orcid.org/0000-0002-9888-3627

ABSTRACT

During the last few years, sports analytics has been growing rapidly. The main usage of this discipline is the prediction of soccer match results, even if it can be applied with interesting results in different areas, such as analysis based on the player position information. In this paper, the authors propose an approach aimed to recognize the player position in a soccer match, predicting the specific zone in which the player is located in a specific moment. Similar objectives have not yet been considered. The authors consider supervised machine learning techniques by considering a dataset obtained through video capturing and tracking system. The data analyzed refer to several professional soccer games captured at the Alfheim Stadium in Tromso, Norway. The approach can be used in real time in order to verify if a player is playing according to the guidelines of the coach. In the experimental analysis, three different types of classification have been performed (i.e., three different divisions of the field), reaching the best results with random tree algorithm.

KEYWORDS

Machine Learning, Performance Analysis, Soccer Analytics, Sport Analytics

1. INTRODUCTION AND RELATED WORK

Sport analytics refers to the use of data and advanced statistics, for example machine learning techniques, to measure performance with the aim to take informed decisions and gain a competitive sports advantage. In other words, sport analytics is the practice of applying mathematical and statistical principles to different sports, such as baseball (Dietrich et al., 2014), basketball (Jain and Kaur,2017) and hockey (Liu and Schulte,2018). In soccer, the most usage is about the prediction of results and the definition of strategies that can be used to win a game or to obtain an improvement of the team performances. Usually, the models constructed in these analysis are based on several aspects about the game, such as tactical, technical or physical information.

However, although each sport has its own characteristics, sport analytics uses the same basic methods and approaches as any other kind of data analysis and, when properly applied, can yield tremendous competitive advantages to a team or an individual player.

The analysis that can be performed with sport analytics is typically divided into two different parts: bio-mechanical and notational analysis (Hughes and Franks,2004). Both techniques involve the analysis and improvement of the sport performance giving good feedbacks to coaches and athletes.

DOI: 10.4018/IJDSST.286678 *Corresponding Author

Sport biomechanics is concerned with fine detail about individual sport techniques while, on the other hand, notational analysis is more concerned with gross movements or movement patterns in games or teams, and is primarily interested in strategy and tactics. These types of analysis are useful because, if we consider the well- chosen performance indicators to evaluate a specific game, it can be possible to highlight advantageous strategy or important aspect of team performance. In other words, they help coaches to identify good and bad performances of team member or the whole team. (Bartlett,2001).

As we have said before, one of the most widespread use of sport analytics, in soccer environment, is related to the prediction of soccer match results. In literature there are several works focused about the most important factors that influence the results of a game. In (Capobianco et al.,2019), the authors propose a new feature set aimed to model a soccer match. The set is related to characteristics obtainable not only at the end of the match, and it is used to predict the results of the match and the number of goal

scored by the team that won the game. In (Joseph et al.,2006), an approach based on Bayesian Networks to predict match results has been presented. The analysis showed that the Bayesian networks is generally superior to other techniques such as the MC4, a decision tree learner, naive Bayesian learner (NB), and k-nearest neighbor learner (KNN) for this domain in terms of predictive accuracy. Specifically, authors obtain an accuracy equal to 59% which outperformed other machine learning models i.e., 41.7% (obtained by the MC4 classification algorithm), 47.86% (with the NB algorithm) and 50.58% (with the KNN algorithm). A similar analysis has been proposed in (Liti et al.,2017), where the authors predict the outcome of soccer matches finished with a draw at the end of the first half using the information stored during the first part of the match; while, in (Razali et al.,2017) a Bayesian Network approach to predict the outcome of English Premier League matches has been constructed. In (Berrar and Dubitsky,2019) the authors suggest that a key factor in soccer match outcome prediction lies in the successful incorporation of domain knowledge into the machine learning modeling process.

In soccer, there are other types of work concerning specific aspects of the game or player performance analysis. For example, in (Fernandez and Cervone,2019) it has been presented a model that quantifies the expected outcome of a soccer possession at any time during the possession, driven by a fine-grained evaluation of the full spatio-temporal characteristics of the 22 players and the ball. In (Kharrat et al.,2017) the authors try to examine who are the best players in European football, and demonstrate how the players' ratings evolve over time, using plus-minus rating. Finally, in (Schultze and Wellbrock,2017) it has been proposed a weighted plus/minus metric to be used as an instrument to evaluate player performance.

In our work, we propose an approach to predict the player positions in a soccer match that can be used to verify, also in real-time, if a specific player observes the guidelines given by the coach. Additionally, this method can be used after the match, to analyze the behavior of the team and make considerations on several aspects to improve performances during the training; or to analyze the next opponent team in order to get some kind of information that can be used to obtain a strategical advantage before the match. Similar objectives have never been considered yet with our best knowledge. In this method, we

exploit supervised machine learning techniques by considering several classification algorithms to enforce the conclusion validity. In detail, the proposed method exploits features related to the relative positions of the ball in x and y axis, other features as, for instance, the player speed.

The paper proceeds as follows: Section 2 contains the proposed methodology for soccer player position detection; Section 3 presents the experimental results; in Section 4 conclusions and future research directions are presented.