UnFOOT: Unsupervised Football Analytics Tool

José Carlos Coutinho[1,2](✉), João Mendes Moreira[2], and Cláudio Rebelo de Sá[1]

[1] University of Twente, Enschede, The Netherlands
j.c.milheirosoarescoutinho@student.utwente.nl,
c.f.pinhorebelodesa@utwente.nl
[2] University of Porto, Porto, Portugal
jmoreira@fe.up.pt

Abstract. Labelled football (soccer) data is hard to acquire and it usually needs humans to annotate the match events. This process makes it more expensive to be obtained by smaller clubs. UnFOOT (Unsupervised Football Analytics Tool) combines data mining techniques and basic statistics to measure the performance of players and teams from positional data. The capabilities of the tool involve preprocessing the match data, extraction of features, visualization of player and team performance. It also has built-in data mining techniques, such as association rule mining, subgroup discovery and a proposed approach to look for frequent distributions.

Keywords: Sports analytics · Association rules · Subgroup discovery · Data visualization

1 Introduction

There already exist tools that given the positional and event-labeled data can extract useful knowledge from the teams and players (Bialkowski *et al.* [1], Gudmundsson *et al.* [2]). However, these tools require event-labeled data, which can be more expensive to obtain than positional data.

UnFOOT[1] uses positional data from players in a football match and extracts different statistics as well as performance indicators of players and teams. This and other information can be explored in more detail in the data analysis section of the tool.

2 UnFOOT Tool

UnFOOT offers a simple and intuitive GUI for analyzing football matches only from spatiotemporal data of players[2]. The pipeline involves 3 stages: Processing of the data, Representation and Data Mining.

[1] This work was financed by the project *Kids First*, project number 68639.
[2] A demonstration can be watched at https://youtu.be/x86tg48qEs4.

U. Brefeld et al. (Eds.): ECML PKDD 2019, LNAI 11908, pp. 786–789, 2020.
https://doi.org/10.1007/978-3-030-46133-1_52

Table 1. Example of the records in the input dataset. *Period* corresponds to the period of the match (1 for the first half and 2 for the second half). *Timestamp* is the time elapsed since the beginning of the period. x and y correspond to the coordinates of the player which are relative to the center of the football field.

ID	Player_id	Period	Timestamp	x	y
1	2	1	15	−37	−45
2	2	1	16	−35	−49

Processing. UnFOOT loads the players' positional data, which should come in 6 columns with the format shown in Table 1. In these experiments, the measurement's frequency was 0.1 s (i.e. 10 record per second) and the $x = 0$ and $y = 0$ correspond to center of the field.

After loading the data, the tool makes one pass on the data and outputs a new dataset with extracted features. These features include the distance covered, the speed and the acceleration of the players. The dataset is divided into time windows of the same size. For each window, several internal modules extract different performance indicators and statistics from the positional data. One of the metrics, *pressure* uses a clustering technique (DBSCAN), from the python package *scikit-learn*[3]. With the clusters, we are able to identify moments of higher pressure of the players during the match. In the end of the analysis, the overall and detailed results are stored into a csv file to enable further analysis outside of the tool.

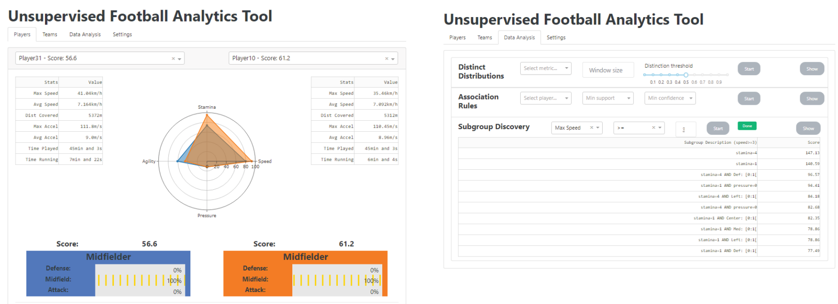
From these performance indicators UnFOOT produces an overall player score which is the mean of the indicators. These player scores are also added together to obtain the score of each team.

Representation and Structure. The GUI is divided in 4 different tabs: *Player*, *Team*, *Data Analysis* and *Settings*. *Player*, evaluates and compares players according to their overall score or specific performance indicators. (Fig. 1(a)); *Team*, displays and compares different team scores and shows the best players.; *Data Analysis*, allows the user to use an interface to execute data mining algorithms on the match data. (Fig. 1(b)); and *Settings*, lets the user load the dataset and define some basic settings before starting the analysis.

Data Mining. The UnFOOT tool has an interface with several data mining techniques to explore the features extracted. One module uses association rules mining to find relationships of performance indicators between consecutive time windows for a selected player. The last method uses subgroup discovery to find subgroups with unusual behaviour relatively to an user defined target. For the association rules mining module, we used *mlxtend*[4], and for the subgroup discovery module we used *pysubgroup*[5].

[3] https://scikit-learn.org/stable/modules/clustering.html.
[4] http://rasbt.github.io/mlxtend/.
[5] https://pypi.org/project/pysubgroup/.

Besides, we also propose a method to look for frequent distributions. This distributions can represent speed or distance covered by players. It is similar to frequent pattern mining, except that the items are distributions. For that, we use the Kolmogorov-Smirnov (KS) to verify if the distribution of one player is significantly different from the other players. In the positive case, the distribution is considered a new *item* and stored in a buffer. Then, UnFOOT counts how many times each distinct distribution is observed during the match to obtain the support (frequency) per player. Distributions which have less than 1% of support are discarded. The users can decide the minimum distance between the distributions to filter very similar distributions.



(a) Illustration of the player analysis interface (b) Illustration of the data analysis interface

Fig. 1. Player and data analysis interfaces

Use Cases. UnFOOT is targeted to football trainers, data analysts and data scientists. Trainers can use UnFOOT to support their analysis of players, teams and matches. This includes comparing different player's performance indicators, but also observing the variation of those indicators along the course of the match.

3 Results

Six real football games were analysed with the tool. Due to privacy issues, we are not able to provide more details about the match, such as the name of the best player per match or the names of the teams. According to some metrics obtained (Table 2), the best player of the match are usually found on the tool's top three players of the winning team. In two cases, they even had the best score overall. Even though the overall score was not originally designed to predict the best player of the match, we use it to validate the scoring function. However, this scoring function can only reasonably assess the quality of players, which are not goalkeepers. This is seen in Game 5, where the best player was actually a goalkeeper. We can also observe that the sum of the team players' individual performance may not be enough to evaluate the performance of the team, since in only half of the cases the highest team score corresponds to the winning team.

Let us now consider the association rules found with the Match 2 data using the association rules mining module. One of the best rules indicated that one striker of Team B was subject to a lot of pressure during the match. During 13% of the match, this player had an intermediate pressure score, which was followed by a high pressure score in 81% of the time. (Rule: Pressure = 6 $\rightarrow$ Pressure = 9 support = 13%, confidence = 81%) Also, the subgroup discovery module discovered two interesting subgroups. One indicates that the players playing on the attack with a high speed score tend to have high agility score. The other group indicates that when players have non-intermediate stamina score (high or low) they tend to have a high speed score.

Table 2. Results obtained with UnFOOT.

Match	Winner	Team A score	Team B score	Rank of best player of the match
1	A	758	778	3rd of Team A
2	A	814	811	1st overall
3	A	795	805	3rd of Team A
4	B	832	855	3rd of Team B
5	A	813	796	Last overall
6	A	816	819	1st overall

4 Conclusion

We proposed UnFOOT, a tool which allows a good understanding of players performance during a match or a training. Data analysts and data scientists can easily use the integrated data mining modules to perform more powerful data analysis. They can also modify the parameters of the algorithms, visualize the results and export the extracted features into a csv.

All results given by the tool were obtained only with player positional data, which presents an advantage over other methods. It also allows an easy extension to other invasion-based team sports since it is independent of the event data, which differs between sports. As future work, we would like to extend the tool to detect other events of the match.

References

1. Bialkowski, A., Lucey, P., Carr, P., Yue, Y., Sridharan, S., Matthews, I.: Identifying team style in soccer using formations learned from spatiotemporal tracking data. In: 2014 IEEE International Conference on Data Mining Workshop, pp. 9–14, December 2014. https://doi.org/10.1109/ICDMW.2014.167
2. Gudmundsson, J., Wolle, T.: Football analysis using spatio-temporal tools. Comput. Environ. Urban Syst. **47**, 16–27 (2014). https://doi.org/10.1016/j.compenvurbsys.2013.09.004