

Review on Winning and Progress Modelling On Real Time Soccer Data

Mr. Swapnil Yadav

*Abha Gaikwad – Patil College of Engineering & Technology,
Nagpur, India*

Received on: 2 Dec,2020 , Revised on: 30 Dec,2020 , Published on: 06 Jan ,2021

Abstract – Sports Analytics is prominent research area with several applications in a variety of fields. These could be the prediction of an athlete's performance or a team performance, the estimation of an athlete's talent and market value, as well as the prediction of a possible injury. Teams and coaches are increasingly willing to install such "tools" in their training, in order to improve their tactics. This paper analyzes the literature on Sports Analytics and proposes a new approach for prediction. We have conducted experiments using suitable algorithms mainly on football related data in order to predict a player's position in the field. Current evaluation methods for soccer player performance relies either on rating from soccer experts or structured statistics of the match, such as shots and tackles. The former requires a lot of manpower and the evaluation is inevitably subjective. The latter can only record the quantity of a player's match events, but cannot reflect the quality (e.g., a wonderful shot or a terrible shot is marked as a shot). In order to solve the above problems, an improved evaluation method for soccer player performance using affective computing is proposed. All the players from Chinese Super League 2019 season are selected as evaluation objects, and the results show that the improved method can evaluate player performance more effectively and reasonably.

INTRODUCTION

Sports analytics is the application of Machine Learning methods and implementations to sports in order to draw

useful conclusions. Such conclusions may affect the performance of an individual athlete, the team as a whole for a specific game or for the whole season. It can also help teams make predictions for upcoming talents, a player's market value and the possibility of an injury. It is a field that is becoming more and more popular nowadays and it is likely to be adopted by a plethora of teams, coaches, individual athletes and companies. Furthermore, in order to be more focused, the number of a player's shots can be predicted in each match, something that has a correlation with goal scoring possibility. In order to achieve good accuracy, not only do we need to find an accurate database, but also the attributes of the database to be relevant to research. So, sometimes we had to combine data from more than one database and create our in order to conduct experiments. In the end, we managed to obtain promising results, encouraging us to continue our work in the future.

There are many ways that a team can use data and there are many kinds. First each player and team as a whole could be related to the way that the game is played (in this case football). This kind of data have to do with the players average stats, such as the number of goals that they score, the number of fouls they commit, with how many red and yellow cards they are booked, how many tackle ins they do, how many kilometers they run during a match (the use of cameras helps for these stats) and many more.

However, it is difficult to compare all those stats in successive matches, because a player’s performance depends on his opponents as well. In other words, a striker may score many goals playing versus a badly organized defense, or a bad goalkeeper. Moreover, there are data that give information about how a team manages to score a goal. For example, how many passes they achieve before they score a goal, what is the average ball possession, and the field that they occur. Obviously, it is different to keep the ball close to the opponent’s goalpost. Sometimes, though, there are outliers that show that other factors play important roles as well, because teams win in football even if they have little ball possession. So, it is needed to use data that are more difficult to collect. These data have to do with players’ physical condition, such as pulse rate when being calm and when sprinting, measure sweatiness and track a player’s sleep. However, most of these data were impossible to collect until recently because these devices were not allowed in football games.

II-AFFECTIVE COMPUTING MODULE

To quantify the quality of players completing each event during a match, the affective computing module is constructed, whose ultimate target is to extract triples of player names, events and affective scores from the post match report. In the data preparation phase, the post-match reports of all the matched in Chinese Super League 2018 season on sports websites are collected as training corpus. 14 soccer fans have labeled to all the sentences in post-match reports make up the dataset as positive and negative.

Due to serious imbalance between positive and negative samples, the data augmentation methods proposed by are utilized to obtain 6,182 positives and 6,073 negatives. Besides, in preparation for text factorization, 131,604 sports news provided by THUCNews is used as training corpus to generate the Word2vec model with 100-dimensional output vector based on a CBOW model.

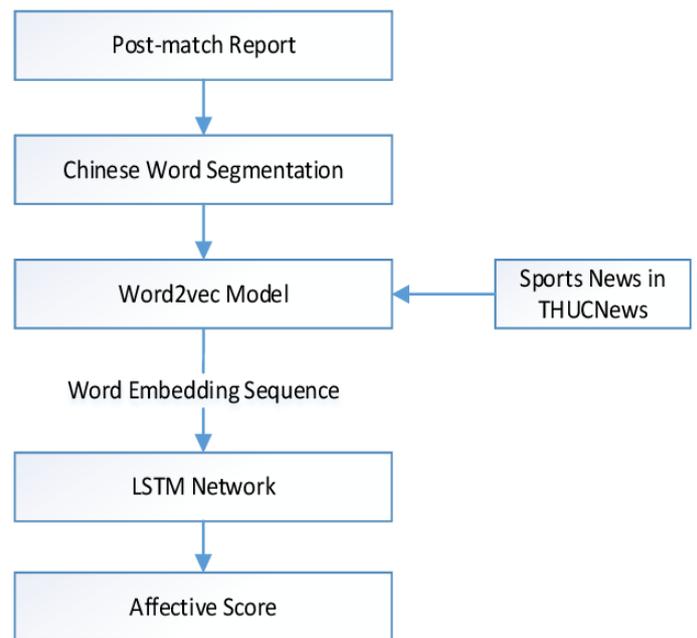


Fig.2- Affective computing model based on LSTM

III- OBJECTIVES

A goal to use and obtain data from wearable devices and conduct experiments there as well. Although, it is difficult to obtain these data. Even sport associations for football and basketball do not allow the use of these devices to great extent, probably because they want to protect sports, but also the athletes. So, at first the data will be obtained by training sessions and not in real matches. Teams are not expected to provide easily such data.

IV-PROBLEM DEFINITION

There are still some problems that need to be address. For example, some behaviors (such as off-ball movement, position selection during defense) in the match are difficult to be counted and quantified, the corresponding time and situation of the events are not taken into account (e.g., the final hit in injury time is more valuable than the goal after a 3-0 lead) and so on.

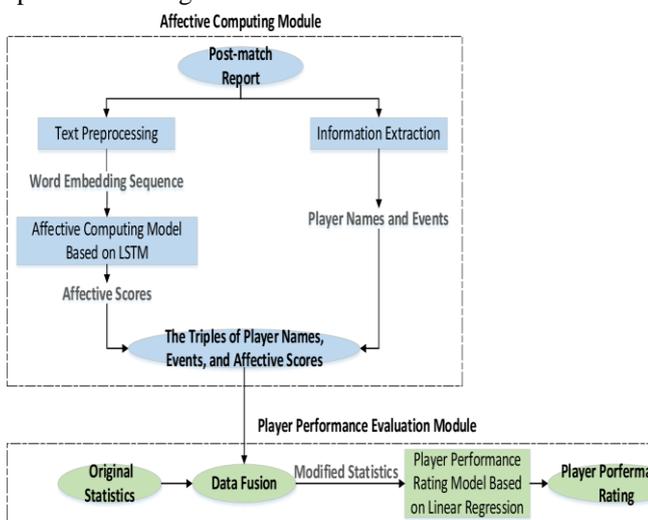


Fig 1- Framework of the improved evaluation method for soccer player performance using affective computing

V-PROPOSED RESEARCH METHODOLOGY TECHNIQUES

For the purposes of this work, we focused on specific athletes. It is understandable, that these approaches may not be generalizable to all athletes, because they might have different characteristics. We must treat every athlete as an individual, in order to achieve good results. The reason that the chosen players were Messi and Suarez is that they are more consistent in their performance. They also play for Barcelona for a few years, so they have developed a certain style. Results would probably be different if it as Messi's first year in Barcelona. It would also be interesting to have the results of individual football players merged and studied as a whole team, in order to predict the team's performance.

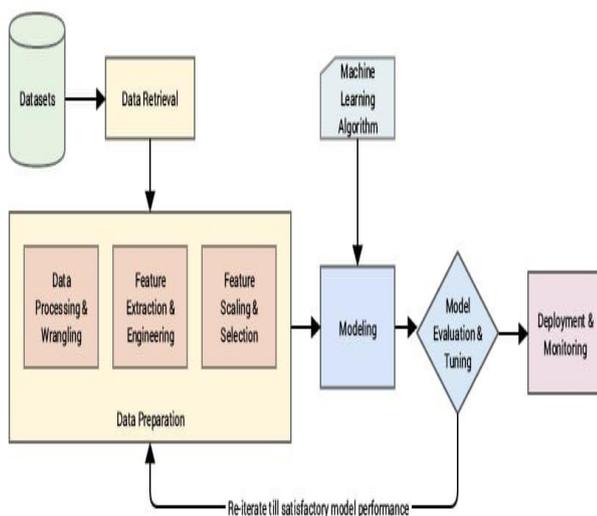


Fig 3- Propose Work Model

VI-DATA DESCRIPTION

One of the most complete websites that contains information about football players individually, but also about the teams, is whoscored.com. Stats about a player's appearances, how many minutes he was in a game for whole season, the number of goals he scored, his assists and completed passes, even how many red and yellow cards he has been booked with, and many more are provided. In this website, we found the necessary data about Messi and Suarez who were the main football players for the experiments that were going to be conducted.

- Since we find the data that are related to our topic, we need a software tools to scrap them and have them in a csv file.
- Now, we have a database to work with. However, it is not "clean" and includes attributes that we do not need. So, the csv file that we obtained from the scraping tools needs to be edited and processed before we can use it for classification. We also had to eliminate missing values, so that the database could be more easily managed while conducting experiments.
- After preprocessing, we can use classification algorithms known to provide good results of course, there should be a variety of classifiers that are going to be used, in order to be able to compare the results and chose the classifier that is more accurate. However, the same classifier should be used for all the players, otherwise the "method" will not be fair, and the results will be biased.

VII-CONCLUSION

In this paper, an improved evaluation method for player performance using affective computing is proposed. By introducing the text information of the post-match report, the affective computing technology is employed to obtain the affective scores of \events mentioned in the post-match report, which are combined with structured statistics to make both the quantity and quality of events can be considered, so that the evaluation of the player performance is more reasonable. Via conducting the player market value comparison experiment and national team player ranking comparison experiment, it is proved that our proposed improved method is superior to the traditional method. Our work brings new ideas and inspiration to the task of the player performance evaluation, and it can be extended to evaluate player performance in various complex and popular sports, such as basketball, ice hockey, American football and so on.

REFERENCES

- [1] "Beyond Moneyball: The future of sports analytics / Analytics Magazine," [Online]. Available: <http://analyticsmagazine.org/beyond-moneyball-the-future-of-sports-analytics/>. [Accessed 10 March 2019].
- [2] Z. McCann, "Player tracking transforming NBA analytics - Tech - ESPN Playbook-

- ESPN,"19May2012.[Online].Available:<http://www.espn.com/blog/playbook/tech/post/id/492/492>. [Accessed 10 March 2019].
- [3] "Competitive fire helps Kirk Lacob make his own name with Warriors," 20 June 2015. [Online].Available: <https://www.sfgate.com/warriors/article/Competitive-fire-helps-Kirk-Lacob-make-his-own-6339796.php>.Accessed 10 March 2019].
- [4] Decroos, Tom and Bransen, Lotte and Van Haaren, Jan and Davis, Jesse, "Actions speak louder than goals: Valuing player actions in soccer," *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 1851-1861, doi: 10.1145/3292500.3330758.
- [5] Bransen, Lotte and Van Haaren, Jan and van de Velden, Michel, "Measuring soccer players' contributions to chance creation by valuing their passes," *Journal of Quantitative Analysis in Sports*, vol. 15(2), pp. 97-116, 2019, doi: 10.1515/jqas-2018-0020.
- [6] Pappalardo, Luca and Cintia, Paolo and Ferragina, Paolo and Massucco, Emanuele and Pedreschi, Dino and Giannotti, Fosca, "PlayeRank: data-driven performance evaluation and player ranking in soccer via a machine learning approach," *ACM Transactions on Intelligent Systems and Technology (TIST)*,vol. 10(5), pp. 1-27 , 2019.
- [7] J. Corscadden, R. Eastman, R. Echelberger, C. Hagan, C. Kipp, E. Magnusson, G. Muller, S. Adams, J. Valeiras and W. T. Scherer, "Developing Analytical Tools to Impact U.Va.Football Performance," in *Systems and Information Engineering Design Symposium (SIEDS)*, 2018.
- [8] J. F. Drazan, A. K. Loya, B. D. Horne and R. Eglash, "From Sports to Science_: Using Basketball Analytics to Broaden the Appeal of Math and Science Among Youth," *MIT-Sloan Sport. Anal. Conf.*, p. 1-16, 2017.
- [9] Gudmundsson, Joachim and Horton, Michael, "Spatio-Temporal Analysis of Team Sports," *ACM Computing Surveys (CSUR)*, vol 50(2), pp. 1-34, Apr. 2017, doi:10.1145/3054132.
- [10] Müller, Oliver, Alexander Simons, and Markus Weinmann, "Beyond crowd judgments: Data-driven estimation of market value in association football," *European Journal of Operational Research*, vol. 263(2), pp. 611-624, 2017, doi: 10.1016/j.ejor.2017.05.005.
- [11] Gyarmati, László, and Hefeeda, Mohamed, "Competition-Wide Evaluation of Individual and Team Movements in Soccer," 2016 *IEEE 16th International Conference on Data Mining Workshops (ICDMW)*. IEEE, 2016, pp. 144-151, doi:10.1109/ICDMW.2016.77.
- [12] Stanojevic, Rade and Gyarmati, Laszlo, "Towards data-driven football player assessment," 2016 *IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, pp. 167-172, 2016
- [13] Y. Yang, "Predicting Regular Season Results of NBA Teams Based on Regression Analysis of Common Basketball Statistics", *PhD Thesis, University of California at Berkeley*, 2015.
- [14] Mikolov, Tomas and Chen, Kai and Corrado, Greg and Dean, Jeffrey, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.
- [15] V. A. Tatsis, C. Tjortjis and P. Tzirakis, "Evaluating data mining algorithms using molecular dynamics trajectories," *Int'l Journal of Data Mining and Bioinformatics* 8 (2), pp. 169-187, 2013.
- [16] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. C. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. Vanderplas, A. Joly, B. Holt and G. Varoquaux, *European Conference on Machine Learning and Principles and Practices of Knowledge Discovery in Databases*, 2013.
- [17] Duch, Jordi and Waitzman, Joshua S and Amaral, Luís A Nunes "Quantifying the performance of individual players in a team activity," *PloS one*, vol. 5(6), 2010, doi: 10.1371/journal.pone.0010937.
- [18] D. Miljkovi_, L. Gaji_, A. Kova_evi_ and Z. Konjovi_, "The use of data mining for basketball matches outcomes prediction," *SIISY 2010 - 8th IEEE Int. Symp. Intell. Syst. Informatics*, pp. 309-312.
- [19] J. Wilson, *Inverting the Pyramid: The History of Football Tactics*, Orion, 2009, pp. 138-144.
- [20] S. Zhang, C. Tjortjis, X. Zeng, H. Qiao, I. Buchan and J. Keane, "Comparing data mining methods with logistic regression in childhood obesity prediction," *Information Systems Frontiers* 11 (4), pp. 449-460, 2009