Proceedings of the 2023 International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/8/20230178

An overview of machine learning applications in the football field

Dasheng Sun

Beijing University of Technology, Beijing, 100124, China

1047870710@qq.com

Abstract. Machine learning has become one of the hottest research directions in computing today and is widely used in various fields, including the field of football, which is considered the world's number one sport. The application of machine learning in the football field has brought great convenience to players, coaches, club managers, spectators, and various professions engaged in the football industry, driving the development of the football industry. As the applications of machine learning are continuously developed and explored, machine learning has a broad development prospect in the field of football. This paper outlines various applications of machine learning in the match process, football odds, and team management, analyzes the shortcomings of previous research, and explores the future directions for the development of machine learning in the field of football.

Keywords: machine learning, football, gaming odds, match analysis.

1. Introduction

The 2022 FIFA (International Federation of Association Football) World Cup in Qatar has received great attention, proving once again that football is the number one sport in the world. There are also many applications of machine learning in the field of football, such as post-match scoring systems for players, odds prediction by betting companies, and scouting systems for teams to evaluate players' abilities. With the rapid development of machine learning, its application in the field of football is also more extensive. A great deal of research has been conducted in this field by scholars worldwide. For example, model fusion-based prediction of football match results [1], the automatic scoring system for player performance [2], and so on.

This paper analyzes specific applications of machine learning in the field of football. It explores the current state of machine learning applications in football and compares the differences in the performance of different algorithms on specific problems. Some shortcomings of existing applications will be identified and possible future trends will be looked into. This paper summarizes various applications of machine learning in football, providing new ideas for future directions. Though lacking experimental support for the feasibility of future applications, this paper still provides directions for future research.

DOI: 10.54254/2755-2721/8/20230178

2. Machine learning in the competition

2.1. Player performance analysis

Player performance analysis on the field is a very important application of machine learning in football. The system can automatically score a player's performance based on his on-field data such as goals, assists, pass success rates, and defensive contribution. Machine learning analysis of players' on-field contributions has also become a focus for many teams, who use the scores as a reference to select the better players to play.

Since football is an offense-driven sport, most player evaluation models are based on offensive data. The existing models mainly consider the players' behavior with the ball, and do not consider the non-intuitive data of running without the ball [2]. The public tends to focus more on the contribution of players in terms of goals and performance on the offensive end, and the contribution of defensive players is severely underestimated, so that offensive players usually receive higher scores in manual scoring. Merhej et al. focus on modeling defensive actions in a game to assess a player's defensive contribution [3]. The authors model defensive players, assigning different weights to different defensive actions, and the final training results show that the contribution value of a defensive player is basically proportional to the value of the player. However, for more complex situations on the court, the model does not give more rewards for certain situations where defending is more difficult, and this is a direction that can be further investigated in the future.

2.2. Competition result analysis

Machine learning has been widely applied in the analysis of football match results. There are many factors that affect the result analysis of football matches, including home and away games, weather, player fatigue, past head-to-head records of both sides, recent team form, etc. Therefore, a major difficulty for machine learning in score prediction is data collecting. With the rapid development of machine learning in football, different models have been applied to score prediction. The literature Pugsee and Pattawong compared the performance of the Random Forest Classifier and multilayer perceptron and finds that Random Forest Classifier is more accurate and can achieve an average of 80% accuracy for match outcome prediction, which is a very satisfactory result [4]. For the prediction based on the BP neural network, Wang et al. used eleven neurons, each representing a factor affecting the match, and ended up with an accuracy of 81.8% [5]. They used a small dataset because the performance with a large dataset may make the results less accurate. A more comprehensive study was carried out by Mrigank Vashist [6], who used the England Premier League dataset consisting of 380 matches and consisting of 45 features. To compare the advantages and disadvantages of algorithms on this problem, he used nine algorithms and compared them. Finally, it was found that by fusing three algorithms, namely XGBoost, gradient boosting classifier, and logistic regression, the best results can be produced. The results are shown in Table 1.

Table 1. Comparison of accuracy.

Model Name	Accuracy Score	Cross Val Score
XBG Classifier	0.75	0.637276
Logistic Regression	0.75	0.628674
Random Forest Classifier	0.70	0.624373
Decision Tree Classifier	0.65	0.539964
K Neighbors Classifier	0.60	0.564337
SVC	0.65	0.635842
GaussianNB	0.50	0.463620
Ada Boost Classifier	0.70	0.627240
Gradient Boosting Classifier	0.70	0.636380

DOI: 10.54254/2755-2721/8/20230178

However, the data sets used in most models so far are small and cannot be applied to all football matches, which is a direction for future improvement. This application can better help teams to find their own deficiencies and address the problems through targeted training or reinforcement.

3. Machine learning in football betting

The betting industry has grown very rapidly over the past few decades and the economic benefits it brings cannot be underestimated. With football receiving more and more attention in the betting arena, the major bookmakers have turned their attention to this promising market. Tens of thousands of football matches around the world each week attract large amounts of money for betting and generate huge profits for bookmakers. Machine learning has also been applied in developing sound betting strategies. As early as 1976, Stefani [7] introduced the least squares betting method and applied it to the 1976 World Cup. This shows that people have been trying to use some betting strategies to obtain stable returns long ago. A data-driven approach was proposed in the literature to predict the outcome of matches and place bets accordingly [8]. This method produced an average financial return of 5.2% per bet. In contrast, the simple use of linear regression methods failed to produce a positive return. In the research of Tsampazis and Tefas, the official odds provided by the bookies were used as data and an auto-encoder was used to build the model [9]. This is a non-recursive multi-layer perceptron that is trained using backpropagation. They are all characterized by not focusing too much on the correct rate, but rather on the return rate, since a high correct rate does not necessarily mean a high profit. This is something that subsequent studies can learn from.

But at the moment, football betting is still a field full of uncertainty, and it is still not realistic to rely on machine learning to get a steady return. Machine learning is still mostly used by betting companies in football betting. No influential models or algorithms have emerged for investors. The question of whether machine learning can be successful in investment strategies remains to be studied in depth by future generations.

4. Machine learning in team management

4.1. Player value

With the professionalization of the football industry, player transfers have become an integral part of clubs' operations. Big teams often spend a lot of money to buy well-known players to improve the strength of the team and the influence of the club. Small teams that do not have the same budget to invest choose to put young players with potential on the shelves for a good price and then use the money to keep the club running. Currently, football researchers prefer to use some traditional linear regression models to evaluate the value of players, where they mainly consider the age, position, performance, and commercial value of the player. Li et al. [10] focus on two new machine learning models to measure the value of players in the market. One is a conventional multiple linear regression model, while the other is a random forest model approach. Both methods are successful and the results show that 69.6% of players are considered normal estimates, 28.9% of players are overvalued, and 1.6% of players are undervalued.

Most player value estimation models are still based on player performance and age, while lacking consideration of factors such as injuries and market value. This is a direction that can continue to be improved in subsequent studies.

4.2. Youth players

Another characteristic of the professionalization of the football industry is the improvement of the club system. Clubs are no longer just competing for established players in the transfer market, they are also targeting young players with potential. This is why youth players are also one of the most valuable assets of a club. Machine learning is also widely used in the youth training industry, including assisting coaches to develop more effective training content and personalized development programs based on the characteristics of different players. A hot topic that has risen in recent years is player

DOI: 10.54254/2755-2721/8/20230178

health management. As the football industry becomes more and more professional, clubs are spending a lot of energy on maintaining the health of their players, especially young players. Youth is the most critical time for players to improve their skills, and one serious injury can potentially kill a player's future. In recent years, a growing number of studies have used contemporary machine learning algorithms including random forest and regression algorithms to predict the injury risk of young players [11]. They have constructed models based on anthropometrics that effectively predict a player's risk of injury. Clubs can apply this technology to properly protect youth players. This application has not been promoted yet, but there is a demand for player protection from various clubs, which shows that the application is very promising.

5. Football robot

Another application of machine learning in the field of football is football robots. Robotic football matches have attracted a lot of attention in recent years, and many researchers have been involved in the development of technology. The environment of a robot football game is very complex, and the robot needs to judge the situation on the field and make the best choice. How to train a robot to learn to play football requires machine learning. Reinforcement learning is the most widely used learning method in football robots [12]. Reinforcement learning has features that are well suited for football robots, such as fast decision making, delayed rewards, ability to handle randomness.

In addition to analyzing which is better to handle on the football field, some robots need to have image recognition capabilities. They need to accurately identify which is the football ball, which are the teammates, which are the opponents, and where are the items in the field. This requires the use of image detection techniques in artificial intelligence. The object detection techniques that need to be used by football robots are the most basic. The literature [13] uses the YOLOv3 network architecture to design a model with reliable detection capabilities. The results are shown in Table 2.

Table 2. Comparison of accuracy.

Class	AP	F1	Recall
Goal	0.975	0.96	94.17%
Robot	0.984	0.97	94.96%
Point	0.912	0.92	85.76%
Ball	0.987	0.97	95.64%

Football robots are a relatively niche area in football. Its spectacle and technicality are far less exciting than that of a live football game. But the application of machine learning on football robots is already relatively mature, and its development space is very limited.

6. Conclusion

Although there are many applications of machine learning in football, the performance is not the best and there is still a lot of room for improvement. At this stage, machine learning is widely applied in match prediction. Betting companies, sports media, or even some individuals who want to make money by predicting match results are applying machine learning. The application of machine learning in the football field will be more and more in the future, and the reliability and timeliness of the models and algorithms will be more and more demanding. The future development of machine learning in the football field will probably be biased toward training and player transfer, where machine learning has not been widely used but has a broad prospect. In the field of training, the application of machine learning can help determine which treatments are better and thus correct some bad habits of players. In addition, it can also help coaches develop reasonable training methods to maximize the improvement of players. In the field of player transfer, machine learning can more accurately determine which players have more potential and which players are more cost-effective. By doing some good operations in the transfer market, the application of machine learning helps clubs gain a lot of benefits.

DOI: 10.54254/2755-2721/8/20230178

This paper mostly discusses the general direction of machine learning applications in football, with less discussion on specific applications. The actual applications face more complex scenarios, and the differences in situations make it impossible to generalize the techniques needed, which is a direction where subsequent research can continue in depth.

References

[1] Zhang, Q., Xu, H., Wei, L. and Zhou, L. Prediction of football match results based on model fusion (2019).

[2] Cui. Y. X., Li, Y. S. and Shen, Y. F. Research status and prospect of football match performance evaluation model. In compilation of abstracts of the 12th National Sports Science Conference - Poster Exchange (Sports Engineering Branch), 110–111 (2022).

[3] Merhej, C., Beal, R. J., Matthews, T. and Ramchurn, S. What happened next: using deep learning to value defensive actions in football event-data. In Proceedings of the 27th ACMSIGKDD Conference on Knowledge Discovery amp; Data Mining, KDD'21, (New York, NY, USA), Association for Computing Machinery, 3394–3403 (2021).

[4] Pugsee, P. and Pattawong, P. Football match result prediction using the random forest classifier. In Proceedings of the 2nd International Conference on Big Data Technologies, ICBDT'19, (New York, NY, USA), Association for Computing Machinery, 154–158 (2019).

[5] Wang, Y. Q., Zhao, H. R., Zhao, H. W. and Xu, X. C. Soccer match prediction based on social network analysis and bp neural network. Weifang Engineering Vocational College Journal 34(03), 104–108 (2021).

[6] Vashist, M., Bahl, V., Sengar, N. and Goel, A. Machine learning for football matches and tournaments. In 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON) 1, 489–496 (2022).

[7] Improved least squares football, basketball, and soccer predictions. IEEE Transactions on Systems, Man, and Cybernetics 10(2), 116–123 (1980). doi: 10.1109/TSMC.1980.4308442.

[8] Knoll, J. and Stubinger, J. Machine-learning-based statistical arbitrage football betting. KI-Kunstliche Intelligenz 34(1), 69–80 (2020).

[9] Tsampazis, K. and Tefas, A. Deep sparse auto-encoders for football match and bet prediction. In Proceedings of the 10th Hellenic Conference on Artificial Intelligence, SETN'18, (New York, NY, USA), Association for Computing Machinery (2018).

[10] Li, C. Y., Kampakis, S. and Treleaven, P. Machine learning modeling to evaluate the value of football players (2022).

[11] Robles-Palaz´on, F. J., Puerta-Callej´on, J. M., G´amez, J. A., De Ste Croix, M., Cejudo, A., Santonja, F., Sainz de Baranda, P. and Ayala, F. Predicting injury risk using machine learning in male youth soccer players. Chaos, Solitons and Fractals 167, (2023).

[12] Cheng, X. Y. and Yang, C. Y. Research status and development of robot soccer learning mechanism. Journal of Jiangnan University (Natural Science Edition) (06), 642-647 (2007).

[13] Wang, Z. Deep learning-based approach for object detection in robot football competition. In 2022 International Conference on Frontiers of Artificial Intelligence and Machine Learning (FAIML), 179–185 (2022).