International Journal of Mobile Computing and Multimedia Communications
Volume 13 • Issue 2

Prediction of Football Match Results Based on Edge Computing and Machine Learning Technology

Yunfei Li, Jilin Institute of Physical Education, China
Yubin Hong, Jilin Institute of Physical Education, China*

ABSTRACT

With the rapid development of artificial intelligence, various machine learning algorithms have been widely used in the task of football match result prediction and have achieved certain results. However, traditional machine learning methods usually upload the results of previous competitions to the cloud server in a centralized manner, which brings problems such as network congestion, server computing pressure, and computing delay. This paper proposes a football match result prediction method based on edge computing and machine learning technology. Specifically, the authors first extract some game data from the results of the previous games to construct the common features and characteristic features, respectively. Then, the feature extraction and classification task are deployed to multiple edge nodes. Finally, the results in all the edge nodes are uploaded to the cloud server and fused to make a decision. Experimental results have demonstrated the effectiveness of the proposed method.

KEYWORDS

Edge Computing, Football Result Prediction, Machine Learning, Sport Match Result Prediction

1. INTRODUCTION

Football is one of the most popular sports. Predicting the results of football matches is interesting to many, from fans to punters. It is also interesting as a research problem, in part due to its difficulty, because the result of a football match is dependent on many factors, such as a team's morale, skills, current score, etc. So even for football experts, it is very hard to predict the exact results of football matches (Owramipur 2013). However, since various types of football matches have outstanding similarities in some respects, in theory, it is possible to find the law from a large number of football matches to find a way to judge the level of victory or defeat (Guan 2021).

With the rapid development of artificial intelligence, machine learning algorithms have been widely used in real life, such as face recognition, stock price prediction, etc. Hence, how to build a football match prediction model based on machine learning algorithms and use scientific methods to solve the prediction problem has become a topic of interest to experts and scholars. Fortunately, researchers have constructed a variety of football match result prediction methods, and have achieved some results. For example, literature (Igiri 2015) used the SVM algorithm to study the factors that affecting the results of the British Championship. Lei used the logistics regression algorithm in machine learning to analyze and process the historical results of football matches, and realized the prediction of the football match results (Lei 2019). Guan also achieved effective result prediction of the Chinese Super League team by suing fuzzy neural network and extreme learning machine (Guan 2021).

DOI: 10.4018/IJMCMC.293749 *Corresponding Author

However, traditional machine learning methods usually use centralized computing, which need to upload previous matches results to the cloud server for centralized processing. When there is a lot of data to be processed, the current methods will face new problems, that is, with the increase of match data, the amount of data that needs to be uploaded to the cloud server will increase significantly, which will occupy a large amount of bandwidth and cause network congestion. And if the classification tasks are all completed by the cloud server, the computing resources cannot be dynamically scheduled, which will occupy a large amount of server computing resources. In addition, in order to improve the learning ability of the classification model, it is usually necessary to increase the number of parameters. However, the increase in the amount of model parameters will cause the classification model to occupy too much computing and storage resources, which will further increase the computing pressure of the server.

In order to solve the above problems, this paper proposes a football match result prediction method based on edge computing and machine learning technology. Edge computing is a new computing method that appears to solve the problems of network congestion and cloud center computing pressure faced by cloud computing. Specifically, edge computing realizes the nearby processing of data by deploying edge computing nodes at the end close to the collection device, and these nodes have the capabilities of computing, storage and communication. And this manner can effectively alleviate network congestion, and sharing the computing tasks of the cloud center by edge nodes can effectively reduce the computing pressure of the cloud center (Wang 2020). Inspired by this, this paper splits the football match result prediction task into some sub-tasks that can be executed at multiple edge nodes, so that the task of feature extraction and classification is deployed on the edge computing nodes.

The remainder of this paper is organized as follows: Section 2 gives a summary of previous work on football result prediction. In Section 3, we describe the proposed architecture of football match result prediction method based on edge computing and machine learning technology. The experiments are provided in Section 4. Section 5 is the conclusion.

2. RELATED WORKS AND ANALYSIS

With the increasing popularity of football, the problem of predicting football match results has become a hot topic in the field of football sports, and it is also an important research content in academic circles. As early as 1990, a football result prediction modal was first proposed based on the average goal rate per football match at the International Gaming Conference. With the progress of the times and the development of technology, there are more and more ways to predict the results of football matches (Arabzad 2014). It can be seen from this that with the popularity of football worldwide, the commercialization of football has also entered a fierce struggle. Whether it is a fan or a data analyst who is interested in football matches, their research on predicting the results of football matches has never stopped. As a result, based on strict mathematical statistics, many scientific methods have been emerged. According to different fields of researchers, there are mainly the following prediction methods.

Bayesian model: The probability of both sides winning the match is determined by various uncertain factors such as the strength of both sides, including the number of historical battles and records of the two teams, and offensive and defensive capabilities. These three factors are used to compare the relative strength of any teams that ultimately scores more than the competing team. Other factors related to team strength may also affect the winning rate, but the Bayesian model believes that their influences are very small. The disadvantage of the Bayesian model is that it places too much emphasis on historical records. With the development of the times, it is impossible to predict the outcome of the current match by only relying on past play data (Razali 2017).

Fuzzy comprehensive evaluation: The fuzzy comprehensive evaluation method analyzes the factors that are likely to attract everyone's attention in the football match as an index to predict the football results, these factors include the number of goals, the number of shots, the weather temperature,

the overall state of the team. However, fuzzy comprehensive evaluation requires too many indicator dimensions. The difficulty of using this type of method to predict the football match result is the data acquisition and data cleaning in the early stage (Cao 2011, Rao 2014).

Machine learning/data mining: In today's rapid development of artificial intelligence, how to build a football match prediction model based on machine learning algorithms and use scientific methods to solve the prediction problem has become a topic of interest to experts and scholars. Machine learning algorithms can obtain features from the original data through feature extraction operations, and try to obtain the relevant laws of the data, then use the obtained laws to predict the unknown data set (Sun 2017). Commonly used machine learning algorithms mainly include k-nearest neighbor algorithm (KNN), support vector machine (SVM), logistic regression (LR), decision tree, random forest, etc. (Mlouk 2017). SVM can effectively adjust the contradiction between algorithm complexity and generalization ability, so it has better promotion ability than traditional pattern recognition methods in the field of small sample learning (Yao 2017). However, when dealing with larger data sets, it usually takes longer to train. KNN is an example-based learning method that can generate decision boundaries of any shape without the need to build a model, but its classification overhead is very large, because similarities need to be calculated one by one. In addition, when the value of k is small, it is also very sensitive to noise (Chatzigeorgakidis 2017). In response to the above shortcomings, the researchers have made many improvements, but there are no classification methods that can simultaneously achieve short training time, strong predictive ability, easy rule extraction and strong adaptability. In comparison, logistic regression has better performance. Because the data characteristics in the football match, such as the number of shots, the number of free kicks in the front court, the number of corner kicks, the number of goals, etc., have an obvious non-linear relationship with the result of the match, the neural network method is very suitable for dealing with this non-linear prediction problem. Guan achieved effective result prediction of the Chinese Super League team by suing fuzzy neural network and extreme learning machine (Guan 2021). Literature (Sujatha 2018) realized football match result prediction based on BP neural network, and analyzed various factors that affect the match result based on historical data.

However, traditional machine learning models need to upload the match results to cloud server in a centralized manner, and the cloud server performs classification calculations separately, which is likely to cause network congestion and occupy a large amount of computing resources on the server. Fortunately, edge computing can effectively solve this defect.

Based on the enlightenment of the above content, facing the problem of football match result prediction, borrowing the idea of edge computing, this paper established three different prediction models based on KNN, logistic regression and neural network, respectively.

3. ARCHITECTURE OF FOOTBALL MATCH RESULT PREDICTION METHOD BASED ON EDGE COMPUTING AND MACHINE LEARNING TECHNOLOGY

3.1 Problem Formulation

From the cybernetic point of view, the task of creating football result prediction models reduced to find out functional mapping from input features to match results:

$$X = \{x_1, \ldots, x_t\} \rightarrow D \in \{d_1, d_2, d_3\}, \tag{1}$$

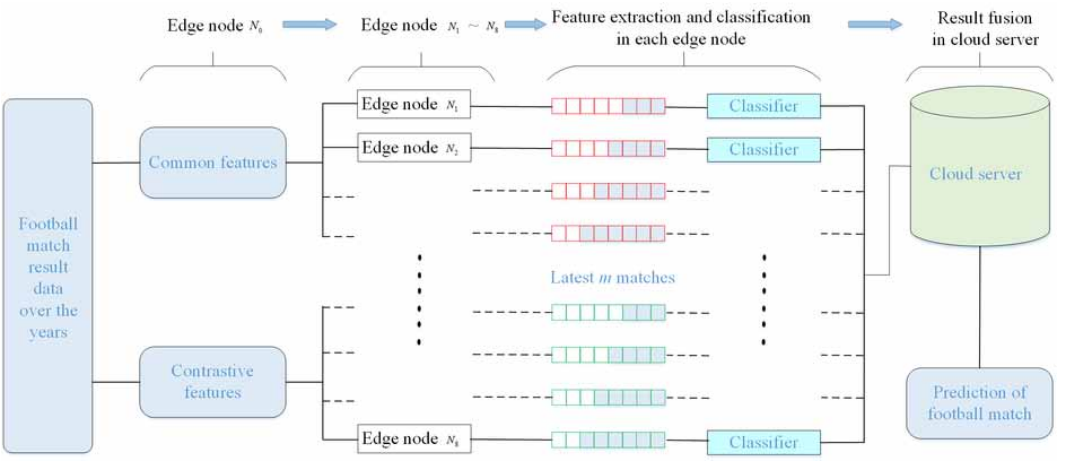
where, X denotes a vector of features (i.e. influence factors), such as number of goals, the number of shots, etc. D denotes the football match result for assessment of one of the terms, and d_1 denotes the host team's win, d_2 denotes draw, d_3 denotes the guest team's win, usually $d_1 = 1$, $d_2 = 0.5$, $d_1 = 0$.

3.2 Architecture of The Proposed Framework

In order to solve the problems of traditional machine learning-based football match result prediction methods that cannot respond in real time, occupy storage and data transmission volume is too large, etc., an architecture of football match result prediction method based on edge computing and machine learning technology is proposed, which is shown in Figure 1.

In Figure 1, the architecture proposed in this paper can be divided into two parts: Edge computing nodes (hereinafter referred to as edge nodes) and cloud server. Among them, edge nodes undertake the main computing task. Specifically, the N_0 node extracts features from historical match results: Including past match results features with all teams (called common features), and match results features with opponent teams (called contrastive features). And transmit these features to edge nodes $N_1 \sim N_m$. On this basis, edge nodes $N_1 \sim N_m$ continue to extract features of different time scales, where $m = 8$. Specifically, edge nodes $N_1 \sim N_4$ are used to extract common features such as the number of points, the number of home goals, the number of away goals, etc., from the last 3, 4, 5, 6 matches, respectively. And edge nodes $N_5 \sim N_8$ are used to extract contrastive features such as the number of goals scored, the number of goals conceded, the number of points, the number of home goals, the number of away goals, etc., from the last 3, 4, 5, 6 matches with a specific opponent, respectively.

Figure 1. The illustration of the architecture for football match result prediction



Our method in this article can be divided into two stages: The training stage and the classification (prediction) stage. In the training stage, the edge nodes $N_1 \sim N_m$ respectively upload the extracted features of different time scale to the cloud server, and receive the model F issued by the cloud server. In the classification stage, the edge nodes input the features into the classification model and output the prediction result, and then upload the result to the cloud server. The features in the classification stage are not uploaded to the cloud server, thereby reducing network transmission and alleviating network congestion. Besides, the computing tasks of edge nodes are independent of each other, and adding nodes can obtain more scale features and improve classification accuracy. In addition, when computing resources are limited, the number of edge nodes can be adjusted to achieve dynamic scheduling of computing resources.

3.3 Football Match Result Prediction Method Based on Edge Computing and KNN

After obtaining the above model, combined with different classifiers, different match result prediction models can be realized. And this subsection uses KNN as the classifier. The basic idea of KNN is: In the feature space, if most of the k nearest (i.e., nearest neighbors in the feature space) samples near a sample belong to a certain category, the sample also belongs to this category.

Assuming there are n teams in a football league, then based on the results of the past few seasons, we can construct $n(n-1)$ samples:

$$\{(X^1, y^1), (X^2, y^1), ..., (X^{n(n-1)}, y^{n(n-1)})\}, \tag{2}$$

where $X^i = [X_1^i, X_2^i, ..., X_8^i]$ denotes the multi-time scale features (here $X_1^i, X_2^i, ..., X_4^i$ are common features, $X_5^i, X_6^i, ..., X_8^i$ are contrastive features), and $y^i = \{0, 0.5, 1\}$ denotes the corresponding label, here 0 represents loss, 0.5 represents draw, and 1 represents win.

When testing, we first construct the multi-time scale features of the two testing teams:

$$X^{t1} = [X_1^{t1}, X_2^{t1}, ..., X_8^{t1}] \text{ and } X^{t2} = [X_1^{t2}, X_2^{t2}, ..., X_8^{t2}]. \tag{3}$$

Then calculating their nearest neighbors and getting their corresponding output o^{t1} and o^{t2}. Finally, the team's label y^{t1} and y^{t2} can be calculated by the following equation:

$$\begin{cases} if\ o^{t1} > o^{t2}, & y^{t1} = 1, y^{t2} = 0, \\ if\ o^{t1} = o^{t2}, & y^{t1} = y^{t2} = 0.5, \\ if\ o^{t1} < o^{t2}, & y^{t1} = 0, y^{t2} = 1. \end{cases} \tag{4}$$

3.4 Football Match Result Prediction Method Based on Edge Computing and Logistic Regression

Logistic regression is a generalization of linear regression, and is mainly used to deal with classification problems. The input of logistic regression is the sample features, and the output is the probability that the sample belongs to the positive class, hence it is a standard binary classification algorithm.

Since there are three categories, we first use the one-against-one strategy to construct a three-classification logistic regression model based on the data samples in Eq. (1). And marking them as models F_1, F_2, and F_3. When testing, we first construct the multi-time scale features of the two testing teams, then substituting them into the multi-class logistic regression model, the corresponding output o^{t1} and o^{t2} can be obtained respectively. Finally, using Eq. (4) we can get their labels.

3.5 Football Match Result Prediction Method Based on Edge Computing and BP Neural Networks

BP neural network is generally composed of three parts: input layer, hidden layer and output layer. And each layer is composed of several neurons. The learning of BP neural network is a supervised learning process. Specifically, the model continuously adjusts the parameters in the neural network by learning to propagate the error backwards while correcting the error, and finally achieve or close to the expected mapping relationship between input and output. The learning of BP neural network can be divided into two stages.

Forward propagation stage: The input signal passes through the input layer, passes through the hidden layer to the output layer, and finally produces an output result. In the process of information being propagated, the weights between connected neurons remain unchanged, and the output of each layer of neurons only affects the input of the next layer of neurons. When the difference between the output information of the final output layer and the supervision signal (error signal) is less than the set threshold, no changes will be made when the learning is completed, otherwise the error signal will be propagated.

Back propagation stage: The error signal generated in the network will propagate backwards from the output layer to the input layer. During the propagation of the error signal, the weights between the neurons in the network will have error signals for feedback adjustment. The weights will gradually stabilize during the adjustment process, and finally the BP neural network will gradually stabilize to obtain the most ideal output.

Different from the previous models, in this model, the output label is the probability that the sample belongs to the positive and negative classes, that is, the probability of the team winning and losing, so the output is a 2-dimensional vector. If the team wins, the corresponding label is set as $[1;0]$. If the team losses, the corresponding label is set as $[0;1]$. If the team draws, the corresponding label is set as $[0.5;0.5]$. Where the first element in this vector represents the probability of the team winning, the second element represents the probability of the team losing, and their sum is 1.

After obtaining the corresponding labels, substituting the training data in Eq. (1) into the model, we can get the trained neural network model F_N. When testing, we first construct the multi-time scale features of the two testing teams, then substituting them into F_N, the corresponding output $o^{t1}=[A_1;B_1]$ and $o^{t2}=[A_2;B_2]$ can be obtained respectively. Finally, the team's label y^{t1} and y^{t2} can be calculated by the following equation:

$$\begin{cases} if\ A_1 - A_2 > 0.5, & y^{t1} = [1;0],\ y^{t2} = [0;1], \\ if\ A_1 - A_2 < -0.5, & y^{t1} = [0;1],\ y^{t2} = [1;0], \\ if\ \left|A_1 - A_2\right| \le 0.5, & y^{t1} = y^{t2} = [0.5;0.5]. \end{cases} \tag{5}$$

4. Experimental results and analysis

This section includes four parts. The main objective of the first part is to introduce the database. In the second part, we will describe in detail how to construct data features. The football match result prediction experiments are performed in the third part. In the last part, we will conduct ablation experiments to verify the influence of edge nodes on the prediction results. All experiments are carried out on a PC with system configuration Inter(R) Core (TM) 2.60GHz with 8GB RAM.

4.1 Database

In this subsection, we use the Chinese Football Super League for the seasons 2008-2018 as an example to verify the effectiveness of the proposed method. In this league, there are 16 teams in each season. However, in the 2008-2018 seasons, there are only 11 teams that have participated in at least 8 seasons, and in our experiments, we select 10 teams to construct our database. Thus, according to Eq. (1) we can construct totally 90 data samples. In the experiment, 80 samples were selected as the training set, and the remaining 10 samples were used as the testing set.

4.2 Sample Features

In this article, there are two types of features need to extract, namely common features and contrastive features.

The common features are extracted from the results of all matches in the past few seasons, their elements mainly include: the transfer money spent, League rank, total goals scored, total goals conceded, home goals scored or conceded, away goals scored or conceded, team cost, year of match, League points, home points, away points, home advantage (if the match is being played at home for team A, feature value of A will be 1 and correspondingly will be 0 for team B) 14 factors in total.

The contrastive features are extracted from the results of matches against specific opponents in the past few seasons, their elements mainly include: the total goals scored, total goals conceded, home goals scored or conceded, away goals scored or conceded, mutual points, home points, away points, home advantage, total shots, total ball possession 12 factors in total. In addition, in order to reduce the influence of different feature sizes on the experimental results and speed up the convergence speed, the features will be normalized first.

4.3 Experimental Results

In KNN model, the nearest neighbor algorithm is used, i.e., set $k = 1$. For neural network model, the number of hidden layer nodes of the network is set to 32, and the Adam optimizer with learning rate of 0.01 is used. Then, the football match result prediction method based on edge computing and KNN, football match result prediction method based on edge computing and logistic regression, and football match result prediction method based on edge computing and BP neural networks constructed in this paper are used to predict the results of these games, and the final fusion result is obtained by averaging the output. The results obtained are shown in Table 1.

Table 1. Prediction results of different methods

Methods	KNN		Logistic regression		BP neural networks	
	Number	Rate	Number	Rate	Number	Rate
Correct prediction	4	40%	6	60%	8	80%
Wrong prediction	6	60%	4	40%	2	20%

From the prediction results in Table 1, we can observe that among the models constructed in this article, the KNN-based model has the lowest accuracy in predicting the result of the matches, and the logistic regression-based model has a higher prediction accuracy than the KNN-based model. The prediction accuracy of BP neural networks-based model is 80%, which is a higher probability in football match prediction. This means that our proposed BP neural networks-based model is very suitable for predicting the football match result.

4.4 Ablation Experiments

In this subsection, based on the BP neural network model, different numbers of edge nodes are used to verified the influence of edge computing on prediction accuracy. And three experiments are designed: Use common features and contrastive features at the same time, only common features are used, and only contrastive features are used. The experimental results against varying number of edge nodes are shown in Figures 2-4.

From these figures, it can be observed that increasing the number of edge nodes can effectively improve the prediction accuracy, or the use of different time scale features can effectively improve the prediction accuracy. In addition, we also find that when the number of nodes is small, the contribution of contrastive features is greater. As the number of nodes increases, the contribution of common features gradually increases.

Figure 2. The prediction accuracy versus the variation of the number of edge nodes when both common features and contrastive features are used

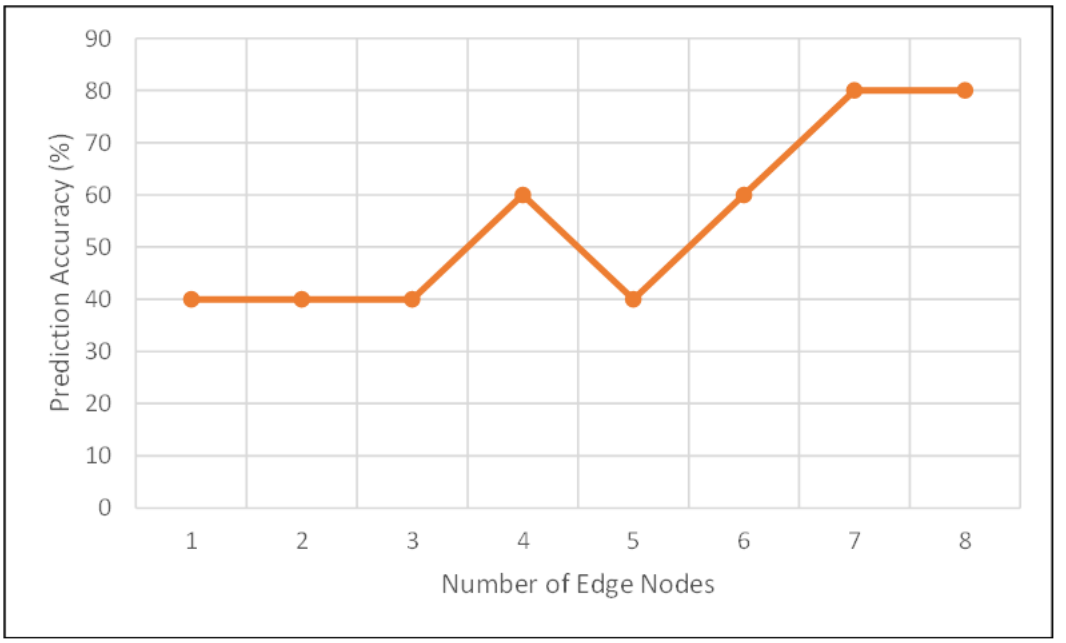
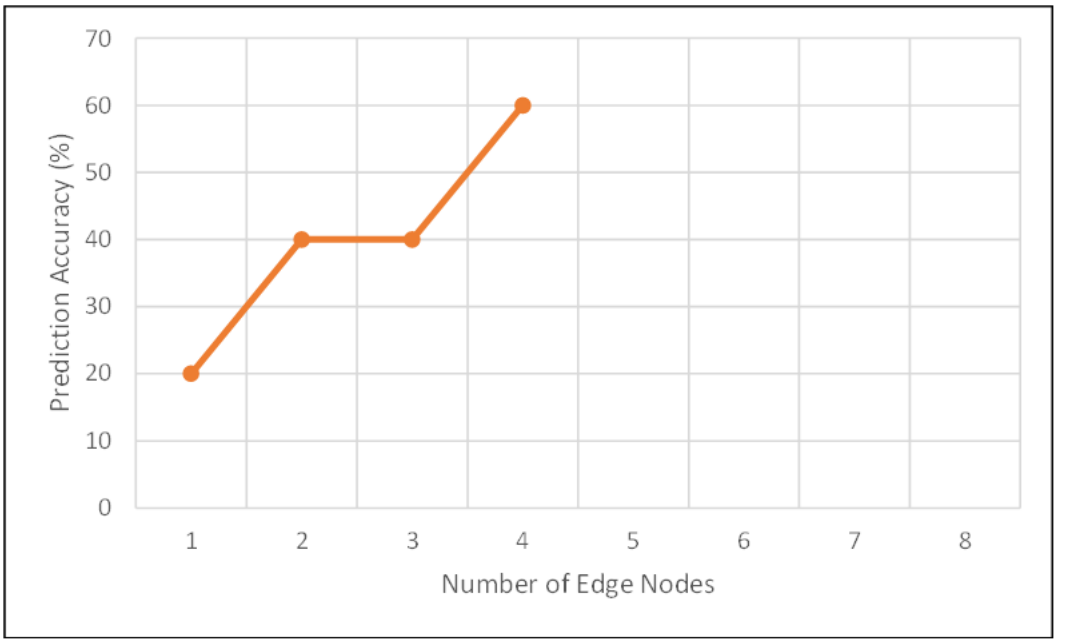
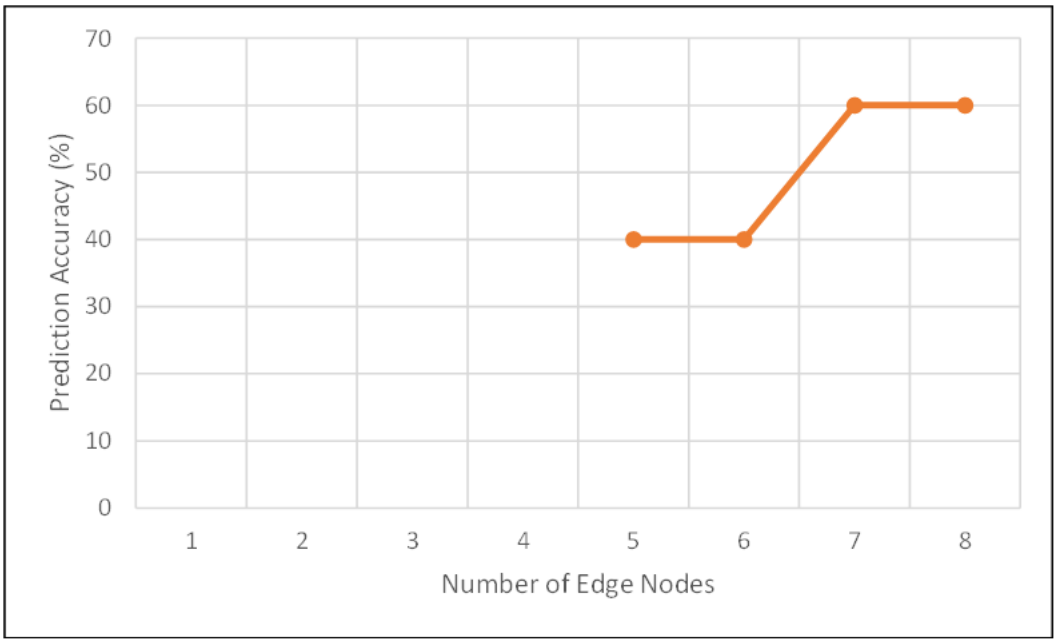


Figure 3. The prediction accuracy versus the variation of the number of edge nodes when only common features are used



5. CONCLUSIONS

Figure 4. The prediction accuracy versus the variation of the number of edge nodes when only contrastive features are used



Aiming at the problem of sports match result prediction, this paper proposes a new football match result prediction method based on edge computing and machine learning technology, which has three different classification models by combining edge computing and three machine learning technology, and realizes effective prediction of football match results. Experimental results on real data sets show that the method proposed in this paper can effectively predict sport match results and provide reference for subsequent research. Future research work includes how to construct a deeper model to obtain more accurate prediction results. In addition, we also interested in selecting more discriminative features.

ACKNOWLEDGMENT

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

Arabzad, S. M., Tayebi Araghi, M. E., & Sadi-Nezhad, S. (2014). Football match results prediction using artificial neural networks: The case of Iran Pro League. *Journal of Applied Research on Industrial Engineering*, *1*(3), 159–179.

Cao, K. (2011). Applied Research on Fuzzy Comprehensive Evaluation in Predicting the Relationship Between Football Matches. *Sports World (Scholarly)*, *7*, 50–51.

Chatzigeorgakidis, G., Karagiorgou, S., & Athanasio, U. S. (2018). FML-KNN: Scalable machine learning on big data using K-nearest neighbor joins. *Journal of Big Data*, *5*(1), 62–77. doi:10.1186/s40537-018-0115-x

Guan, S., & Wang, X. (2021). Optimization analysis of football match prediction model based on neural network. *Neural Computing & Applications*. Advance online publication. doi:10.1007/s00521-021-05930-x

Igiri, C. P. (2015). Support vector machine-based prediction system for a football match result. *IOSR Journal of Computer Engineering Ver. III*, *17*(3), 2278–2661.

Lei, G. (2019). Multi-classification Forecasting Model Based on World Cup Competition. *Software Guide*, *18*(7), 45–48.

Mlouk, A. A., Agouti, T., & Gharnati, F. (2017). Mining and prioritization of association rules for big data: Multi-criteria decision analysis approach. *Journal of Big Data*, *4*(1), 1541–1553.

Owramipur, F., Eskandarian, P., & Mozneb, F. (2013). Football Result Prediction with Bayesian Network in Spanish League-Barcelona Team. *International Journal of Computer Theory and Engineering*, *5*(5), 812–815. doi:10.7763/IJCTE.2013.V5.802

Rao, C. J., & Dong, J. H. (2014). Fuzzy Comprehensive Evaluation Model of Selecting Football Coach. *Advanced Materials Research, 989-994*, 1821-1824.

Razali, N., Mustapha, A., Yatim, F. A., & Ab Aziz, R. (2017). Predicting Football Matches Results using Bayesian Networks for English Premier League (EPL). *IOP Conference Series. Materials Science and Engineering*, *226*, 012099. doi:10.1088/1757-899X/226/1/012099

Sujatha, K., Godhavari, T., & Bhavani, N. P. G. (2018). Football Match Statistics Prediction using Artificial Neural Networks. *International Journal of Mathematical and Computational Methods*, *3*, 1–8.

Sun, L. F., Qin, J., Wang, K. L., & Zhang, J. (2017). Expansion of pathogen recognition specificity in plants using pattern recognition receptors and artificially designed decoys. *Science China. Life Sciences*, *60*(8), 336–352. doi:10.1007/s11427-017-9064-5 PMID:28699103

Wang, X., Han, Y., Leung, V., Niyato, D., Yan, X., & Chen, X. (2020). Convergence of edge computing and deep learning: A comprehensive survey. *IEEE Communications Surveys and Tutorials*, *22*(2), 869–904. doi:10.1109/COMST.2020.2970550

Yao, H. P., Liu, C., Zhang, P. Y., & Wang, L. (2017). A feature selection method based on synonym merging in text classification system. *EURASIP Journal on Wireless Communications and Networking*, *2017*(1), 964–971. doi:10.1186/s13638-017-0950-z