A Deep Learning Approach to Predict Football Match Result

4 authors, including:

Dwijen Rudrapal
National Institute of Technology Agartala
42 PUBLICATIONS   **298** CITATIONS

A Deep Learning Approach to Predict Football Match Result

Dwijen Rudrapal*, Sasank Boro, Jatin Srivastava and Shyamu Singh

National Institute of Technology, Agartala, Tripura-799046, India
`dwijen.rudrapal@gmail.com,sambsasank@gmail.com,srijatin1997@gmail.com,`
`shyamsingh188@gmail.com`

Abstract. Predicting a match result is a very challenging task and has its own features. Automatic prediction of a football match result is extensively studied in last two decades and provided the probabilities of outcomes of a scheduled match. In this paper we proposed a deep neural network based model to automatically predict result of a football match. The model is trained on selective features and evaluated through experiment results. We compared our proposed approach with the performance of feature based classical machine learning algorithms. We also reported the challenges and situations where proposed system could not predict the outcome of a match.

Keywords: Predictive Modelling, Match result, Machine Learning, Multi-layer Perceptron, Deep Learning

1 Introduction

Football is the most played and most watched sport in the world. Footballs governing body, the International Federation of Association Football (FIFA), estimated that at the turn of the 21st century there were approximately 250 million football players and over 1.3 billion people interested in football [1]. Popularity of this game created huge number of people's association. Therefore, prediction of match result in advance is very attractive to the experts and researchers. But it is very difficult to guess the result of a football match by experts or past statistics. There are many factors who can influence the match outcome like, skills, player combination, key players forms, teamwork, home advantage and many others. Prediction becomes more challenging when match played with extra time or substitute players or injuries.

Football sports related data is now publicly available and grown interest in developing intelligent system to forecast the outcomes of matches. In last two decades, researchers proposed several state-of-the-art methodology to predict outcome using past and available data. Most of the previous research devise the prediction problem as a classification problem [1, 13]. The classifier predicts the result with one class among win, loss or draw class [9].

---

[1] https://www.britannica.com/sports/football-soccer

In this paper we focus on deep neural network for this problem. Deep neural network technique has shown to be effective in various classification problems in many domains [4] including in this domain [5, 7].

The main contribution of this paper are as follows:

- We propose Multi Layer Perceptron based prediction model to predict football match result. We experimented our model with classical machine learning based models and report our evaluation result.
- We propose new set of features to train our predictive model by detail analysis of features performance and their role on match outcome and from head-to-head match result.

The remainder of this paper is organized as follows. A summary of previous promising work on football predictions is presented in Section 2. In Section 3, the proposed approach including the dataset, is presented followed by experiment result in section 4. We discuss error analysis of our approach in section 5 and finally, the conclusion is in Section 6.

2 Related Work

In this section we have discussed promising research work related to predictive model for football match result. The research for predicting the results of football matches outcome as started as early 1977 by [11]. proposed a model to predict the match outcome using matrix of goal scoring distribution. Purucker proposed [10] an ANN model to predict results in the National Football League (NFL) using hamming, adaptive resonance theory (ART), self-organizing map (SOM) and back-propagation (BP). The approach uses 5 features and experimented on data of 8 round matches in the competition. The work extended in the approach proposed by Kahn [5] achieve greater accuracy by using large dataset of 208 matches in the 2003 season of National Football League (NFL). The work includes features like home team and away team indicator indicator. The approach [8] presents a framework for football match result prediction based on two components: rules based component and Bayesian network component. The performance of the system depends on sufficient expert knowledge of past data. McCabe and Trevathan predict result in their approach [7] for 4 different kind of sports. Proposed approach targeted Rugby League, Australian Rules football, Rugby Union and EPL. A multi-layer perceptron model trained using BP and conjugative-gradient algorithms for predicting task.

The system in the work [3] proposed a prediction model to generate forecasts of football match outcome for EPL matches using objective and subjective information of teams like team strength, team form, psychological impact and fatigue. Ben Ulmer and Matthew Fernandez used features like game day data and current team performance in the proposed work [14]. Albina Yezus[15] used both static and dynamic features in proposed approach. Static features includes forms of the players and teams, concentration and motivation, while dynamic features includes goal difference, score difference and history.

Tax and Joustra proposed a prediction system [12] for Dutch football competition based on past 13 years data to predict the match results. Koopman and Lit developed a statistical model in their work [6] for predicting football match results. The work used Bayesian Networks and assumes a bivariate Poisson distribution with intensity coefficients that change randomly over time. Recently Bunker and Thabtah [2] investigated machine learning approaches for sport result prediction and focused on the neural network for prediction problem. The work identifies the learning methodologies, data sources, evaluation and challenges proposed a prediction framework using machine learning approach as a learning strategy.

3 Proposed Methods for Football Match Prediction

3.1 Data Collection

English Premier League (EPL) is the most watched football league in the world with almost 4.7 billion viewers. Its competitiveness as well as its outcomes are random in nature. For example, in the season of 2010-11, the distribution of wins, losses and draws was 35.5%, 35.5% and 29% respectively. For our current research work, we concentrated on EPL matches. We collected different data like team statistics and squad, player and result from various online sources [2,3,4,5]. The dataset includes information of 11,400 matches from the season 2000-01 to the season 2015-2016. We analyzed what parameters or features are influential or highly influential in determining the outcome of a match. Detail statistics of the dataset is reported in table 1.

Table 1. Statistics of dataset

Season count	Match count	Team count	Player count
15	11400	42	462

3.2 Feature Selection

From literature survey of previous promising research work, we have drawn 40 features from our prepared dataset. Features are drawn from different class of statistics like team, player and head-to-head match statistics. Each match is classified into two classes, home match and away match. Through our analysis we found that this classification plays crucial role to predict match result due to the advantages of venue, crowd support and known environment to the home team.

---

[2] football-data.co.uk
[3] fifaindex.com
[4] skysports.com
[5] kaggle.com

So, all the features are calculated for home as well as away matches. Distribution of features are reported in table 2. Details explanation of the features are discuss in following subsections.

Table 2. Features distribution details

Features class	Features count		
	Home team	Away team	Total
Team	9	9	18
Player	7	7	14
Head-to-head match	4	4	8
Total	20	20	40

Team Related Features

1. **Attack Point:** It describes the attacking ability of a team.
2. **Mid Field Point:** It describes the mid field ability of a team.
3. **Defence Point:** It describes the defensive ability of a team.
4. **Team Rating:** Calculated considering the attack, mid field and defence point.
5. **EPL Times Point:** It is the times that a team played EPL during the 18 seasons. If a team played all the 18 seasons they are given 4 points.
6. **EPL Streak Point:** It describes how consistent a team is playing in EPL.
7. **EPL Performance Point:** It describes the best performance of a team in EPL during these 18 seasons. A champion is given maximum of 4 points.
8. **EPL Point:** It is the average of EPL times Point, EPL Streak Point and EPL Performance Point.
9. **Standings Point:** It is the standing point of each team in whichever league they are. For example point can vary from 1-4 according to their position in the points table.

Player Related Features

1. **Palyer Rating:** It describes the overall rating of a player considering parameters like aggression, tackle, shooting accuracy, dribbling and many more. This gives an overview of the players performance. 94 was the highest overall rating of a player in our dataset.
2. **Player Potential:** Player potential describes the maximum ability of a player. The player rating can have a maximum value equal to player potential when the player is at his best form.
3. **Market Value:** Market Value gives an idea of the form of player, how dependable he is and how many seasons still left in his career. The maximum value of this feature is 123 million in our corpus.

4. **Goal Keeper Point:** It describes the ability of the Goal keeper of each team.
5. **Defence Point:** It includes total of 11 defence parameters of a player like aggression, marking, positioning, sliding tackle, standing tackle. The average value of these parameters is taken and calculated with respect to 4.
6. **Mid Field Point:** It includes total of 13 mid field parameters of a player like acceleration, agility, ball control, crossing, dribbling.
7. **Striking Point:** It gives an idea of the striking or scoring ability of each player. Includes total of 19 striking parameters of a player like acceleration, curve, finishing, long shots, power shot.

Head to Head Match Related Features

1. **Team Points:** Team point is the average of the points achieved by a home or away team. Every win conceded 3 points, draw and lost conceded 1 and 0 points respectively.
2. **Form Points:** It describes the form of a team during head to head matches. Last 5 home and 5 away matches prior to the match is considered.
3. **Goal Difference:** It is the difference between goals scored and goals conceded divided by the total number of matches.
4. **Form Points Difference:** It is the difference between the home form point and away form point.

3.3 Classifier Selection

In this paper we have proposed Multi-Layer Perceptron (MLP) for football match result prediction. In earlier similar research work [5] [10] MLP has shown comparable performance. We have also experimented with popular statistical machine learning algorithms like Support vector machine (SVM), Gaussian Naive Bayes and Random Forest to evaluate our approach.

4 Experiment and Result

Proposed MLP classifier having 10 hidden states trained on our dataset. We used holdout method to split our dataset into training set and testing set. So, 70% of the dataset is tagged as training dataset and rest 30% tagged as test dataset. In table 3, we report the accuracy as 73.57% on the test dataset. We also report sensitivity, specificity, precision, recall and F1-score of our proposed approach.

Table 3. Performance evaluation of proposed approach

Parameter	Accuracy	Sensitivity	Specificity	F1-score
Score	73.57	71.02	75.48	71.45

As sensitivity and specificity of our model does not differ so much, therefore we can conclude that our model has a decent accuracy. We have compared the accuracy of our model with promising statistical machine learning algorithms in this domain research like SVM, Gaussian Naive Bayes and Random Forest.

Table 4. Analysis of proposed approach

Algorithm	Accuracy (in %)	F1-score (in %)
MLP	73.57	71.45
SVM	58.77	50.07
Gaussian Naive Bayes	65.84	64.26
Random Forrest	72.92	66.07

5 Error Analysis

In this section we make an in depth analysis of our model and report the situations where the model failed to predict match outcome.

1. In our current research work, we are considering the features of only the starting 11 players. But sometimes substitute of a player also determines very important role in the match. Sometimes, playing 11 may also change in another match. So, these challenges can be resolved by considering all the players playing 11 and substitutes.
2. There are many hidden features like crowd support, unfairness of a referee, unpredictable nature of the game etc. In our current work we could not include those features. Inclusion of those features will improve our model.
3. As the features which we have considered are dynamic in nature, the value changes slowly and gradually with the passage of seasons. So, performance of our approach is mostly dependant on recent past matches.

6 Conclusion

In this work we have achieved a decent accuracy of predicting the outcome of a football match. Our model works very well considering recent past match statistics. We made detail analysis of features and drawn effective features to train our model. We have also shown that match wise features like home and away match features are effective to the model.

Our future scope of the work is to study the importance of the features like team's reserve player strength, coach rating and inclusion of those to make our model more robust.

References

1. Abdelhamid, N., Ayesh, A., Thabtah, F., Ahmadi, S., Hadi, W.: Mac: A multiclass associative classification algorithm. Journal of Information & Knowledge Management 11(02), 1250011 (2012)
2. Bunker, R.P., Thabtah, F.: A machine learning framework for sport result prediction. Applied Computing and Informatics (2017)
3. Constantinou, A.C., Fenton, N.E., Neil, M.: pi-football: A bayesian network model for forecasting association football match outcomes. Knowledge-Based Systems 36, 322–339 (2012)
4. Davoodi, E., Khanteymoori, A.R.: Horse racing prediction using artificial neural networks. Recent Advances in Neural Networks, Fuzzy Systems & Evolutionary Computing 2010, 155–160 (2010)
5. Kahn, J.: Neural network prediction of nfl football games. pp. 9–15 (2003)
6. Koopman, S.J., Lit, R.: A dynamic bivariate poisson model for analysing and forecasting match results in the english premier league. Journal of the Royal Statistical Society: Series A (Statistics in Society) 178(1), 167–186 (2015)
7. McCabe, A., Trevathan, J.: Artificial intelligence in sports prediction. In: Information Technology: New Generations, 2008. ITNG 2008. Fifth International Conference on. pp. 1194–1197. IEEE (2008)
8. Min, B., Kim, J., Choe, C., Eom, H., McKay, R.B.: A compound framework for sports results prediction: A football case study. Knowledge-Based Systems 21(7), 551–562 (2008)
9. Prasetio, D., et al.: Predicting football match results with logistic regression. In: Advanced Informatics: Concepts, Theory And Application (ICAICTA), 2016 International Conference On. pp. 1–5. IEEE (2016)
10. Purucker, M.C.: Neural network quarterbacking. IEEE Potentials 15(3), 9–15 (Aug 1996)
11. Stefani, R.T.: Football and basketball predictions using least squares. IEEE Transactions on systems, man, and cybernetics 7, 117–121 (1977)
12. Tax, N., Joustra, Y.: Predicting the dutch football competition using public data: A machine learning approach. Transactions on Knowledge and Data Engineering 10(10), 1–13 (2015)
13. Thabtah, F., Hammoud, S., Abdel-Jaber, H.: Parallel associative classification data mining frameworks based mapreduce. Parallel Processing Letters 25(02), 1550002 (2015)
14. Ulmer, B., Fernandez, M., Peterson, M.: Predicting Soccer Match Results in the English Premier League. Ph.D. thesis, Doctoral dissertation, Ph. D. dissertation, Stanford (2013)
15. Yezus, A.: Predicting outcome of soccer matches using machine learning. Saint-Petersburg University (2014)