Football Result Prediction

GOWTHAM S

Master of Computer Application

Coimbatore Institute of Technology

Coimbatore, Tamil Nadu, India

SURIYA PRAKASH T

Master of Computer Application

Coimbatore Institute of Technology

Coimbatore, Tamil Nadu, India

VIGNESH A

Master of Computer Application

Coimbatore Institute of Technology

Coimbatore, Tamil Nadu, India

Dr .J.B.JONA, Associate Professor

Department of Compute Application

Coimbatore Institute of Technology

Coimbatore, Tamil Nadu, India

ABSTRACT:

Data analysis (or Data mining) has become a major part of our businesses, governments and lives. The concept of analyzing the things that happen around us in our everyday life could be said to be human nature making sense of our environment, the people around us, how and or the people around us livetheir lives and drawing conclusions from these observations. Data mining, in this sense, is observing a stream of data or a static dataset, using algorithms to make sense of the data or find patterns, have a target or conclusion in mind want the algorithm to draw from the data (inductive)or allow the algorithm to draw its own conclusions from the data (deductive); whichever way, new insights and conclusions are drawn.

There are different methods that have been proposed for predicting a football match but in this thesis paper, do not propose one method or any methodfor predicting a football match. show acomparison of different methods using feature selection from mainly historical data; and show if using football team formations (or tactics) as dependent features add any significant value to the predicting a football match. However, the team formation data used in this thesis has limited information due to the difficulty in acquiring football data in the world today.

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

1.INTRODUCTION

Data analysis (or Data mining) has become a major part of our businesses, governments and lives. The concept of analyzing the things that happen around us in our everyday life could be said to be human nature, making sense of our environment, the people around us, how and/or the people around us livetheir lives and drawing conclusions from these observations. Data mining, in this sense, is observing a stream of data or a static dataset, using algorithms to make sense of the data or find patterns, have a target or conclusion in mind want the algorithm to draw from the data (inductive) or allow the algorithm to draw its own conclusions from the data (deductive); whichever way, new insights and conclusions are drawn. Data mining is said to give answers to questions never thoughtwere there. Data mining is now being used in sports.In Association Football (football1), the sports seen as the most popular sports in the world, data mininghas been a big part of the sport as a cooperatebusiness, a betting business, gaming industry business. Data mining is a big part of the football cooperate business by determining the worth of players bought into a club and sold out of a club, bydetermining how players' health are managed individually as well as collectively, how players aretold to play and many more. The football betting industry business, however, has been an active business long before data mining was introduced to it but as every other industry, with the introduction of technology and more specifically, data mining, there becomes an exponential growth in the amountof business that can be done in a quicker but more efficient way. Having said this, data mining in the football betting industry has brought about more efficient ways of predicting the outcome of a singlefootball game, to the extent that even the minor details that go on in a football game like the amountof corners, who gets the first goal kick or which player gets the first booking of the game is being predicted on a daily basis; league positions are also

being predicted and bookmaker odds have become more profitable with the use of data mining techniques amongst many advantages data mining has brought to the football betting industry business.

2. LITERATURE

Alpaydın [1] defines machine learning as programming computers to optimize an execution model utilizing case information or past encounters. We have a model with a few parameters, and learning is an execution of a computer program that optimize these parameters. The model may be prescient to create expectations within the future, or graphic to collect information from information, or both. Machine learning uses the hypothesis of insights in building numerical models, since the main task is making deduction from a sample.

Murphy [2] defines machine learning as an evolving domain of informatics whose algorithms are outlined to imitate human insights by learning from the encompassing environment. Machine learning techniques have been implemented effectively in different areas including spacecraft engineering, pattern recognition, computer vision, financing, entertainment and medical applications. Mitchell [3] defines machine learning as searching for the answers, which are related to construction of computer systems that consequently learn with experience, and defining fundamental processes of learning patterns. Also, Mitchell mentioned that statistics, human and animal learning methodologies in psychology and neuroscience are the fields that are closely related to machine learning. In future, it is can be anticipated the harmony between the studies of human learning and machine learning. These fields have a close relationship for the fundamental scientific questions. Speech recognition, computer vision, bio-surveillance, robot control and accelerating empirical sciences are the real world application fields of machine learning.

3. DATASET

Our dataset consists of data derived from the past three football seasons (2012/2013, 2013,2014 and 2014/2015) from the five (5) big leagues in Europe, which are the English Premier League, German Bundesliga, French Ligue 1, Italian Serie A and the Spanish La Liga, in no order of relevance. Before going into the main experiments of this research, would like to point out the diversity of the different leagues in dataset based on one attribute, whichis the total form. The total form is representative of the cumulative points gathered by each team in the course of a season; these points (cumulative) are then divided by the number of games played (also cumulative). The figure below shows two column bars for each of the five leagues; the first bar represents the default accuracy value of that league while the second bar shows the accuracy value based on team league positions of the league in question.

4. TEAM FORMATION

This part of the experiments is carried out on the team forms. The team forms are representative of a team's recent performance. The team form values will be represented using four (4) different representations, which will be explained below. The team form consists of the Home and Away team forms as shown below:

Home Team Form: The total amount of points the home team got from the last n (the value of n will be determined experimentally) games.

Away Team Form: The equivalent information for the away team. The representations of the Team Forms are:

Representation 1 (r1): This represents the numeric values of the team forms, normalized to interval [0,3].

Representation 2 (r2): This represents the discretized value of the team forms. had reason to believe that the classifiers do not distinguish between values well enough while using r1, so discretized r1 using the set of rules

Numeric Values	Discretized Values
[0,1>	Bad Form
[1,2>	Good Form
[2,3]	Best Form

5. METHODOLOGY

The proposed prediction technique uses a **Neural Network** approach to predict the results of football matches. The neural network detects patterns from a number of factors affecting the outcome of a match making use of historical cases of training.

6. RESULTS AND DISCUSSION

Algorithm	Accuracy	Loss%
Logistic regression	0.66	0.44
Gaussain naive classifier	0.65	0.45
Random forest classifier	0.62	0.48

Figure no 1

The main goal of this study is to investigate to what extent a neural network is able to predict the outcome of football matches as a home win, away win or a draw. To investigate this matter different neural networks have been created that differ in deepness, number of neurons per layer and output. An interesting main finding that emerges from the analysis, is that the proposed networks with 3 output neurons are able to predict football matches with an accuracy of 48% while the proposed models with 1 output neuron achieves a 43% accuracy for predicting matches of the Premier League, Championship, League 1 and League 2. These results suggest that neural networks with 3 output neurons are more appropriate for football result prediction than neural networks consisting of 1 output neuron. This suggestion is somewhat unexpected as prior studies only created networks that consist of 1 output neuron. Results also suggest

that for football prediction models 1 hidden layer is sufficient which is in line with the network architectures of Aslan and Inceoglu (2007),Huang and Chang (2010) and McCabe and Trevathan (2008), and that the number of neurons per layer is sufficient when it is between the size neurons which is in line with Heaton (2009) and McCabe and Trevathan (2008).

7.CONCLUSION

This thesis paper was set out to show a comparison between different methods a football match can be predicted using feature selection, whether or not team formations add value in predicting a football match, what features would perform better than the others, what combination of features are best used together and their overall performances in predicting football matches. The general theoretical literature regarding football match predictions propose one method for predicting football games using quantitative data, qualitative data or a mixture of both, whichever type of data is used, a comparison of different methods and the use of team formations as a dependent feature to predict a football match is a gap usually left unanswered.

REFERENCES

[1] E. Alpaydin, "Introduction to machine learning," Cambridge, MA: The MIT Press, 2014.

[2] El Naqa I., Murphy M.J, "What Is Machine Learning?," Machine Learning in Radiation Oncology. Springer, Cham, 2015

[3] T. M. Mitchell, "The discipline of machine learning," Pittsburgh, PA: Carnegie Mellon University, School of Computer Science, Machine Learning Dept., 2006

[4] Hucaljuk, J & Rakipović, A. (2011). Predicting football scores using machine learning techniques. 2011 Proceedings of the 34th International Convention MIPRO,, 1623-1627.

[5] Abel Hijmans. Dutch football prediction using machine learning classifiers (unpublished).

[6] Darwin, P & Dra, H (2016). Predicting Football Match Results with Logistic Regression. 2016 International Conference On Advanced Informatics: Concepts, Theory And Application (ICAICTA).

[7] Farzin, O., Parinaz, E., &Faezeh, S. M. Football result prediction with Bayesian network in Spanish league-Barcelona team. International Journal of Computer Theory and Engineering, 5(5), 2013, 812-815.

[8] T. Cheng, D. Cui, Z. Fan, J. Zhou, and S. Lu, "A new model to forecast the results of matches based on hybrid neural networks in the soccer rating system," Proceedings Fifth International Conference on Computational Intelligence and Multimedia Applications. ICCIMA 2003.

Aslan, B., & Inceoglu, M. (2007). A Comparative Study on Neural Network Based Soccer Result Prediction. Seventh International Conference on Intelligent Systems Design and Applications (ISDA 2007).

[9] Boice, J. (2018, August 10). How Our Club Soccer Predictions Work. Retrieved April 15, 2019, from https://fivethirtyeight.com/methodology/how-our-club-soccer-predictions-work

[10] Boyko, R. H., Boyko, A. R., & Boyko, M. G. (2007). Referee bias contributes to home advantage in English Premiership football. Journal of Sports Sciences,25(11), 1185-1194.

[2]