



Exploring polynomial classifier to predict match results in football championships



Rodrigo G. Martins^{a,c}, Alessandro S. Martins^a, Leandro A. Neves^b, Luciano V. Lima^c,
Edna L. Flores^c, Marcelo Z. do Nascimento^{d,*}

^aIFTM, r. Belarmino Vilela Junqueira S/N, 38305-200, Ituiutaba, MG, Brazil

^bUNESP - DCSS, r. Cristóvão Colombo 2265, S. J. Rio Preto, 15054-000, SP, Brazil

^cUFU - FEELT, av. João Neves de Ávila 2121, BLX, Uberlândia, 38400-902, MG, Brazil

^dUFU - FACOM, av. João Neves de Ávila 2121, BLB, Uberlândia, 38400-902, MG, Brazil

ARTICLE INFO

Article history:

Received 8 October 2016

Revised 17 April 2017

Accepted 18 April 2017

Available online 19 April 2017

Keywords:

Polynomial classifier
Football championship
Prediction
Feature selection
Machine learning

ABSTRACT

Football is the team sport that mostly attracts great mass audience. Because of the detailed information about all football matches of championships over almost a century, matches build a huge and valuable database to test prediction of matches results. The problem of modeling football data has become increasingly popular in the last years and learning machine have been used to predict football matches results in many studies. Our present work brings a new approach to predict matches results of championships. This approach investigates data of matches in order to predict the results, which are win, draw and defeat. The investigated groups were different type of combinations of two by two pairs, win-draw, win-defeat and draw-defeat, of the possible matches results of each championship. In this study we employed the features obtained by scouts during a football match. The proposed system applies a polynomial algorithm to analyse and define matches results. Some machine-learning algorithms were compared with our approach, which includes experiments with information obtained from the football championships. The association between polynomial algorithm and machine learning techniques allowed a significant increase of the accuracy values. Our polynomial algorithm provided an accuracy superior to 96%, selecting the relevant features from the training and testing set.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Football is a game played by two teams with eleven players each and any scoring corresponds to kicking a ball into the opponent's goal. There are three possible results of a match: win, draw and defeat. One of the teams wins when it achieves the highest score within a period of 90 min (Sfeir, 2011). Namely, it is a simple and inexpensive game, which explains the reason why it became so popular world-wide. In 2014 Brazil hosted the last World Cup, whose audience reached more than 3.5 billion TV viewers (FIFA, 2014). The 2012 Olympic Games in London had an audience of 4 billion people (Olympic, 2012). The difference between the two events is that the World Cup was played among 32 countries for a

single sport whereas in the Olympic Games 204 countries participated among 26 types of sports.

The first World Cup happened in 1930, whereas the history of Olympic Games started much earlier, namely in 1896 (Cardinal, Boonchaythanasit, Cheung, Lee, & Si, 2016; Moore, 2016). From 1936 on they included football definitely. Hence, this long-lived and widespread interest in football yielded a huge dataset with detailed information about each single match in championships until now (Perin, Vuillemot, & Fekete, 2013).

The result of a football match has been considered as subject of numerous scientific efforts in the endeavour to improve the game tactics and team features. In literature there are many studies which focus on football matches forecasts (Constantinou, Fenton, & Neil, 2013). Forecasting football is comprised of result of a match (win, draw and defeat) and score, which can be used by the bookmakers on the bets. Much less effort has been devoted, to the understanding of football from the perspective of the predicting results. Predicting the results is a difficult problem because of the number of factors which must be taken into account may not represent quantitatively (Hucaljuk & Rakipović, 2011). For instance, a

* Corresponding author.

E-mail addresses: rodrigo.grassimartins@gmail.com (R.G. Martins), alessandrom2006@yahoo.com.br (A.S. Martins), neves.leandro@gmail.com (L.A. Neves), vieira@ufu.br (L.V. Lima), edna@ufu.br (E.L. Flores), marcelo.zanchetta@gmail.com (M.Z. do Nascimento).

team may completely dominate statistical measures such as number of good finish, number of successful pass or possessions in the offensive area, and fail to score one goal more than the opposing team to win a match (Brooks, Kerr, & Guttag, 2016).

In the literature we can find studies that involve predict matches results in a championship and of other team sports. Ulmer and Fernandez (2013) studied such techniques (baseline, Gaussian naïve Bayes, hidden Markov model, multimodal naïve Bayes, support vector machine (SVM), random forest (RF), One vs All SGD) to predict results by using the goals scored by each team in 10 seasons (from the 2002–03 season to the 2011–12 season) of the English Championship. Hucaljuk and Rakipović (2011) have surveyed the prediction of results to the UEFA Champions League also through scored goals with the following algorithms: naïve Bayes (NB), Bayesian networks, logitboost, k-nearest neighbours (KNN), RF and artificial neural networks. With SVM, Igiri (2015) studied the data related to scores in the English Championship.

With a much greater amount of information, Parinaz and Sadat (2013) employed data related to physiology and football strategies in order to analyse the Futbol Club Barcelona in the Spanish Championship. In that work the authors described a Bayesian network approach for football results prediction with the NETICA software. The model considered only one team to predict the result of football matches. Tax and Joustra (2015) used the following classification algorithms: CHIRP, logitboost, DTNB, FURIA, hyperpipes, J48, naïve Bayes, multilayer perceptron and RF. For selection they applied ReliefF, CfsSubsetEval and principle components to 65 features from the Dutch Championship, but they could not determine the relevant features in order to raise the accuracy of score predictions. Duarte, Soares, and Teixeira (2015) surveyed the C5.0, JRip, RF, KNN, SVM and NB algorithms to predict matches of the Portuguese Championship with the following pieces of information: previously scored goals, number of goals, lapse between matches and features of the competition. New features and also information about the psychology information of the football players were applied in Duarte et al. (2015), which however did not improve the performance of the model. For this reason, it is a challenge to investigate information and strategy of classification which facilitate the prediction of matches results. Thus, the focus of this research is to propose a new approach to predict football matches results of championships.

Others studies have been developed to predict matches results exploring machine learning algorithms. These algorithms are tools that receives as inputs the features and provides as output the prediction of matches results (win, draw and defeat). There are algorithms that can provide the most appropriate response to the problem (Pendharkar, Khosrowpour, & Rodger, 2000; Ramirez-Villegas & Ramirez-Moreno, 2012). Our decision for the polynomial classification was conditioned by its capacity of learning complex patterns that could be linearly inseparable and the success reached in other applications (Park, Oh, & Kim, 2008). A polynomial classification is a parameterized nonlinear map which non-linearly expands a sequence of input vectors to a higher dimension and maps them to a desired output sequence. This expansion can improve the separation of the different classes in an expanded vector space. Moreover, this strategy presents the advantages of providing only one model for optimal separation of classes and to consider this as the solution of the problem, which does not occur with the models presented in Campbell, Assaleh, and Broun (2002) and Ajmera and Holambe (2010).

Our present study introduces a novel approach to predict football matches results of championships. This approach investigate data of matches in order to predict the results, which are win, draw and defeat. The proposed system employs a polynomial classification algorithm that analyses and defines the matches results. We apply features related to information obtained by scouts. The

investigated groups were win-draw, win-defeat and draw-defeat. In this paper, we used the 10-fold cross-validation method and the sliding window technique, in which the training part and the group test used 4 and 2 rounds, respectively, to evaluate the ability of the model to predict football matches results of championships. The accuracy rates were calculated and evaluated considering datasets from real tournaments. Comparisons with machine learning classification algorithms were drawn and their new accuracy values showed a significant increase. Furthermore, the robustness of the proposed method was also investigated with addition of different noise levels on the features: the results were important even in adverse conditions. Here are the main contributions of this work:

- It presents a novel approach to identify football matches results based on the polynomial classifier with features obtained by scouts.
- It shows the usefulness of our approach when applied to datasets commonly investigated in the literature.
- It proposes the use of the polynomial classifier as an algorithm for feature selection and to improve the performance of machine learning techniques.
- It investigates the impact of noise at different attributes and the robustness of the proposed method given by the accuracy rates.

This paper is organised as follows. Section 2.1 describes details about how the datasets of the football championships were obtained. Sections 2.2 and 2.3 provide a brief description of each machine learning technique used in this survey and they also explain our polynomial algorithm. Experimental results, together with a discussion, are presented in Section 3. Section 3.4 describes some related to researches, and finally we draw our conclusions in Section 4.

2. Methodology

The proposed approach was organised into three stages, data collection, which is based on procedures related to the events that occur during a football match, feature selection for polynomial classification, and the classification through the polynomial algorithm and machine learning techniques. Fig. 1 presents the proposed flowchart of approach and in the following sections the main details are described.

The polynomial algorithm was implemented in MATLAB R2012b, the machine learning classifiers and feature selection were implemented applying the Waikato Environment for Knowledge Analysis (WEKA) software (Hall et al., 2009).

2.1. Data collection

The data used for the proposed approach were the football matches results, which were obtained from the different Championships: English Premier League (EPL), season 2014/2015; La Liga Primera Division (LLPD), season 2014/2015; and Brazilian League Championships, seasons 2010 (BLC 2010) and 2012 (BLC 2012). These different datasets can provide the predictive power of the proposed method. The championships were played on the point system accrued by 20 teams. Each team faced the other both home and away, playing in their own stadium as well as that of the visiting team. In this manner, there were a total of 38 rounds, and in each 10 matches were played, thus producing a total of 380 matches per championship.

EPL and *LLPD*: The data were the results (win, draw or defeat) of all English Premier League matches from season 2014/15. Similar number of instances were obtained of Liga Primera Division matches from season 2014/2015. This information is available online at <http://www.football-data.co.uk/>. These datasets were used

Download English Version:

<https://daneshyari.com/en/article/4943076>

Download Persian Version:

<https://daneshyari.com/article/4943076>

[Daneshyari.com](https://daneshyari.com)