Football Result Prediction with Bayesian Network in Spanish League-Barcelona Team

Farzin Owramipur, Parinaz Eskandarian, and Faezeh Sadat Mozneb

Abstract—The problem of modeling football data has become increasingly popular in the last few years and many different models have been proposed with the aim of estimating the characteristics that bring a team to lose or win a game, or to predict the score of a particular match. We propose a Bayesian Network (BN) to predict results of football matches. During the last decade, Bayesian networks (and probabilistic graphical models in general) have become very popular in artificial intelligence. In this paper, we look at the performance of a BN in the area of predicting the result of football matches involving Barcelona FC. The period under study was the 2008-2009 season in Spanish football league and we test its performance. We get necessary information about football states from valid websites. This BN is used for prediction of football results in future matches.

Index Terms—Bayesian network, football result prediction, football data, simulation, prediction

I. 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. Bayesian networks (and probabilistic graphical models in general) have become very popular in predictions works such as weather, games, etc [1], [2]. In this paper we present a new approach for football result prediction. In doing so, we introduce a football factors that effect on results.

The remainder of this paper is organized as follows: Section II gives a summary of previous work on football prediction. In Section III, we describe Bayesian network. Section IV is about our approach in using of Bayesian network. In Section V, we present simulation of BN in NETICA software. For evaluation, we use the results of the 2008-2009 Spanish leagues, comparing a historic predictor with the predictions of our network's results and finally, conclusion is in Section VI.

II. RELATED WORKS

Football is perhaps the World's pre-eminent sport, so it is

not surprising that there has been a substantial amount of research on football prediction. Rothstein et al. [3] used fuzzy logic in formalizing football predictions, and genetic and neural optimization techniques in tuning their fuzzy model. Koning [4] took a Bayesian network approach with Markov chains and the Monte-Carlo method, estimating the quality of football teams using this model. Rue et al. [5] suggested a Bayesian dynamic generalized linear model to estimate the time dependent skills of football teams, and to predict football matches. Andersson et al. [6] estimated the expected number of goals in a football match, based on the scoring intensity, and predicted the probability that a team wins a tournament using the estimated scoring intensities.

Halicioglu [7] analyzed football matches statistically and suggested a method to predict the winner of the Euro 2000 football tournament. His method is based on the seasonal coefficients of variation (CVs) of the end-of-season points. Falter et al. proposed an updating process for the intra-match winning probability. Falter et al. [8] and Forrest et al. [9] focused more on the analysis of football matches rather than on prediction. Goddard et al. [10] suggested an ordered probity regression model and using it, forecast English league football results.

Karlis et al. [11] used the plausibility of Poisson regression models for the interpretation and prediction of football scores. Crowder et al. [12] modeled the 92 teams in the English Football Association League using refinements of the independent Poisson model and predicted the probabilities of home win, draw or away win. Koning et al [13] and Kuonen [14] specifically address tournament-specific issues, and/or directly model football tournaments. Martinich [15], Amor, et al. [16], and Yang et al. [17] investigated other sports such as American football, major league baseball.

Among the existing works, the approach of Min et al. [18] is most similar to ours. They present an approach for result prediction. Their system (called FRES) consists of two major components: a rule-based reasoner and a Bayesian network component. This approach is a compound one in the sense that two different methods cooperate in predicting the result of a football match. Second, contrary to most previous works on football prediction they use an in-game time-series approach to predict football matches.

III. BAYESIAN NETWORKS

In the last few years, Bayesian networks (BN) have become a popular way of modeling probabilistic relationships among a set of variables for a given domain [19]-[21]. Bayesian networks (BNs) are graphical models [22] for reasoning under uncertainty, where the nodes represent

variables (discrete or continuous) and arcs represent direct connections between them. These direct connections are often causal connections. In addition, BNs model the quantitative strength of the connections between variables, allowing probabilistic beliefs about them to be updated automatically as new information becomes available. A Bayesian network for a set of variables X = {X1,.....,Xn} consists of 1) a network structure S that encodes a set of conditional independence assertions about variables in X, and 2) a set P of local probability distributions associated with each variable. Together, these components define the joint probability distribution for X. The network structure S is a directed acyclic graph.

The probabilities encoded by a Bayesian network may be Bayesian or physical. When building Bayesian networks from prior knowledge alone, the probabilities will be Bayesian. When learning these networks from data, the probabilities will be physical (and their values may be uncertain). To illustrate the process of building a Bayesian network, consider the problem of detecting credit-card fraud. We begin by determining the variables to model.

One possible choice of variables for our problem is Fraud (F), Gas (G), Jewelry (J), Age (A), and Sex (S), representing whether or not the current purchase is fraudulent, whether or not there was a gas purchase in the last 24 hours, whether or not there was a jewelry purchase in the last 24 hours, and the age and sex of the card holder, respectively. The states of these in the next phase of Bayesian-network construction, we build a directed acyclic graph that encodes assertions of conditional independence. One approach for doing so is based on the following observations. Variables are shown in Fig. 1.

IV. SELECTING MAIN FACTORS AND RELEVANT INFORMATION

There are a large number of factors which could affect the outcome of a football match from the perspective of one of the teams involved. One of the difficulties in any investigation of the relationships involved in a given effect is that to a large extent the assumption of a particular model determines the attributes to study and predetermines the possible relationships that can be found. So, the act of choosing which model and attributes to study sets a boundary on what can be discovered.

So we try to find main factors that affect on football results. We divide them into two groups:
1) Non-psychological factors
2) Psychological factors

Betting experts use them to predict match results. Table I shows main factors of football match prediction that we used in BN.

We consider the Spanish football league for the season 2008-2009. The league is made by a total of T=20 teams, playing each other twice in a season (one at home and one away). We consider specially Barcelona team's results. We gather data from football websites such as [23]-[25]. We found necessary information for BN in their content.

After that, we categorize them and make a criterion for

each factor. For example, in the number of goals is various amounts (average) that Barcelona scores in each match, so we categorize its scoring goal into three groups:
1) Less than one.
2) Between one and three.
3) More than three.

In BN, these results affected on final result (win/lost/draw). Fig. 2 shows a Diagram of factors and their Dependences.
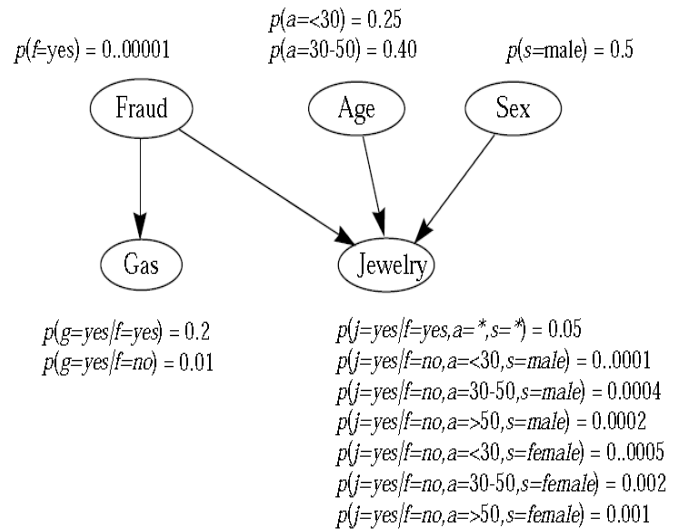

$p(a=<30) = 0.25$
$p(f=yes) = 0..00001$
$p(a=30-50) = 0.40$
$p(s=male) = 0.5$

Fraud   Age   Sex

Gas   Jewelry

$p(g=yes|f=yes) = 0.2$
$p(g=yes|f=no) = 0.01$

$p(j=yes|f=yes,a=*,s=*) = 0.05$
$p(j=yes|f=no,a=<30,s=male) = 0..0001$
$p(j=yes|f=no,a=30-50,s=male) = 0.0004$
$p(j=yes|f=no,a=>50,s=male) = 0.0002$
$p(j=yes|f=no,a=<30,s=female) = 0..0005$
$p(j=yes|f=no,a=30-50,s=female) = 0.002$
$p(j=yes|f=no,a=>50,s=female) = 0.001$

Fig. 1. A Bayesian network for detecting credit-card fraud; Arcs are drawn from cause to effect.

TABLE I: MAIN FACTORS IN FOOTBALL MATCH PREDICTION

Main factors in Football match prediction	
Psychological	Non Psychological
weather	Average_of_players_age
History_of_5last_games	Injuried_main_players
Result_against_for_teams	ave_match_in_week
Home_game	Performane_of_main_players
ability_front_team	performance_of_all_players
Psychological_state	ave_goal_in_all_home
	ave_goal_for_Home

TABLE II: VALUES OF MAIN FACTORS THAT AFFECT ON FINAL RESULT

Some values of BN factors
Average_of_players_age = medium
History_of_5last_games = win
Injuried_main_players = No
Psychological_state = good
performance_of_all_players = high
weather = good

V. IMPLEMENTING BN IN THE NETICA SOFTWARE

In this section, we describe how implemented our BN in NETICA software [26]. NETICA is one of the best software for BN. The diagram in Fig. 2 is implemented in NETICA exactly. CPT tables for every factor are filled based on information and others are filled by NETICA. In this approach, we consider every matches separately because every match has different values for factors and they change based on their situations in matches. In Fig. 3, a state of BN for a match is shown and it is related to in Spanish league (Barcelona- Sevilla). We gathered that match's data and inserted into CPT tables and Fig. 3 shows the results of match.
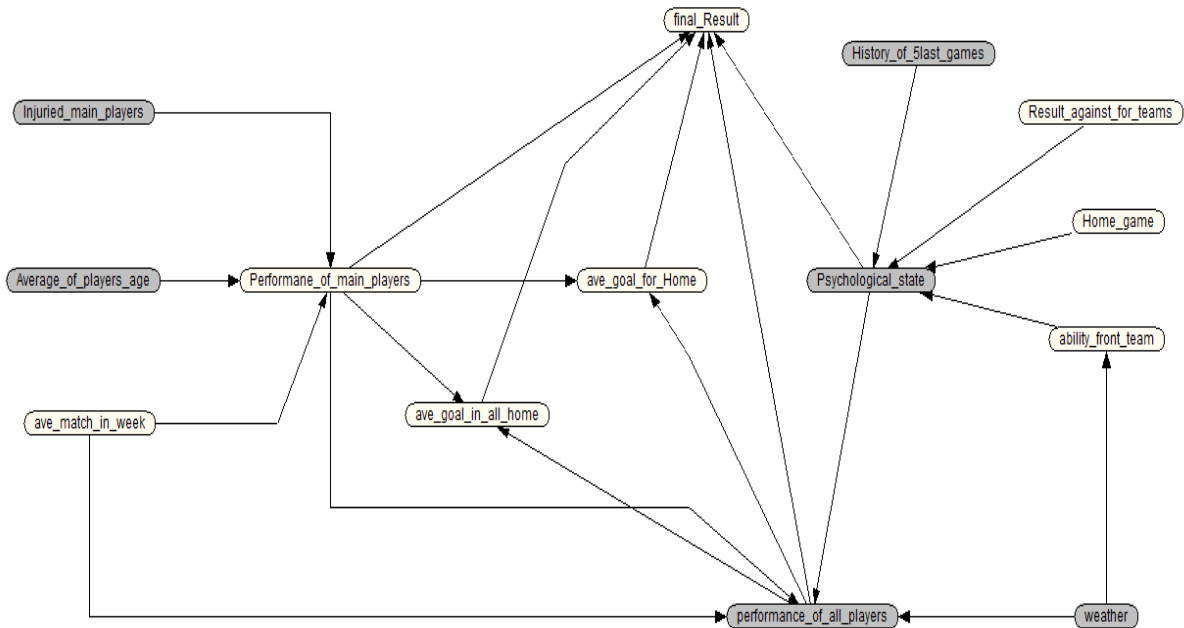
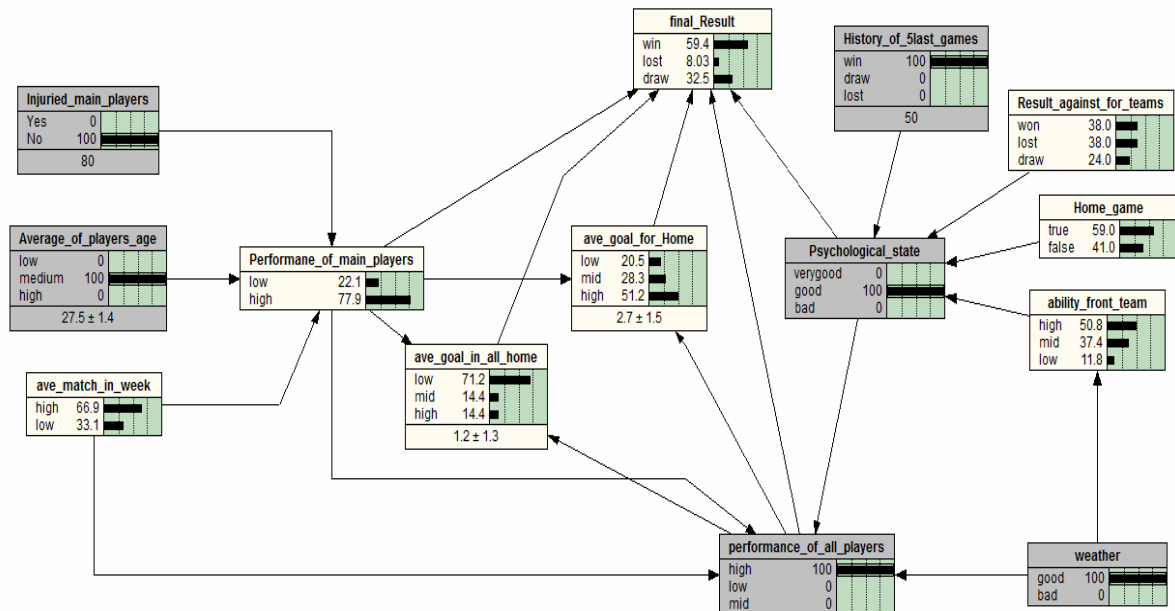Fig. 2. Labeled box view of football result prediction's BN

Fig. 3. A view of BN for Barcelona- Sevilla match in 2008-2009 football season in Spanish league

For every match for Barcelona in league, we do this process. In Table II, we show the value s of that match.

We repeated this process for all 38 matches that did in 2008-2009 season. The results of these operations were gathered and the final point of Barcelona team was determined. After that, this result compared with 2008-2009 season and we saw the final result was correct in 92%.

VI. CONCLUSION

In this paper, we described a Bayesian network approach for football results prediction. The prediction in football has been an interesting problem and researchers have tried to find a solution for it. So we offered a Bayesian network approach for it and showed how it worked. All used data in BN are gathered from valid websites that offer football statistics in the world.

REFERENCES

[1] G. Baio and M. A. Blangiardo, "Bayesian hierarchical model for the prediction of football results," *Journal of Applied Statistics*, vol. 37, no. 2, pp. 253-264, 2010.
[2] H. Rue and O. Salvensen, "Prediction and retrospective analysis of soccer matches in a league," *The Statistician*, vol. 49, pp. 399-418, 2000.
[3] A. P. Rotshtein, M. Posner, and A. B. Rakityanskaya, "Football predictions based on a fuzzy model with genetic and neural tuning," *Cybernetics and Systems Analysis*, vol. 41, no. 4, pp. 619-630, 2005.
[4] R. H. Koning, "Balance in Competition in Dutch Soccer," *The Statistician*, vol. 49, no. 3, pp. 419-431, 2000.
[5] H. Rue and O. Salvesen, "Prediction and retrospective analysis of soccer matches in a league," *The Statistician*, vol. 49, no. 3, pp. 399-418, 2000.

[6] P. Andersson, M. Ekman, and J. Edman, "Forecasting the fast and frugal way: a study of performance and information-processing strategies of experts and non-experts when predicting the world cup 2002 in soccer," SSE/EFI Working Paper Series in Business Administration No. 2003:9, Stockholm School of Economics, Stockholm, 2003.

[7] F. Halicioglu, "Can we predict the outcome of the international football tournaments: The case of EURO 2000," *Doğuş Üniversitesi Dergisi*, vol. 6, no. 1, 2005.

[8] J. M. Falter and C. Perignon, "Demand for Football and Intra-Match Winning Probability: An Essay on the Glorious Uncertainty of Sports," *Applied Economics*, vol. 32, pp. 1757-1765, 2000.

[9] D. Forrest and R. Simmons, "Outcome uncertainty and attendance demand in sport: the case of English soccer," *The Statistician*, vol. 51, no. 2, pp. 291-241, 2002.

[10] J. Goddard and I. Asimakopoulos, "Forecasting football results and the efficiency of fixed-odds betting," *Journal of Forecasting*, vol. 23, pp. 51-66, 2004.

[11] D. Karlis and I. Ntzoufras, "On Modeling association football data," Technical Report, Dept. of Statistics, AUEB, 1999.

[12] M. Crowder, M. Dixon, A. Ledford, and M. Robinson, "Dynamic modeling and prediction of English football league matches for betting," *The Statistician*, vol. 51, pp. 157-168, 2002.

[13] R. H. Koning, M. Koolhaas, G. Renes, and G. Ridder, "A Simulation Model for Football Championships," *European Journal of Operational Research*, vol. 148, no. 2, pp. 268-276, 2003.

[14] D. Kuonen, "Statistical Models for Knock-Out Soccer Tournaments," Technical Report, Department of Mathematics, École Polytechnique Federale de Lausanne, 1996.

[15] J. Martinich, "College football rankings: do the computers know best," *Interfaces*, vol. 32, no. 5, pp. 85-94, 2002.

[16] M. Amor and W. Griffiths, "Modeling the Behaviour and Performance of Australian Football Tipsters," Working Paper, Dept. of Economics, Univ. of Melbourne, 2003.

[17] T. Y. Yang and T. Swartz, "A two-stage bayesian model for Predicting Winners in Major League Baseball," *Journal of Data Science*, vol. 2, no. 1, 2004.

[18] B. Min, C. Choe, and R. I. McKay, "A Compound Approach for Football Result Prediction, " in *Proc. the 2006 Asia-Pacific Workshop on Intelligent and Evolutionary Systems*, November 2006, pp. 215-225.

[19] A. R Khantaymoori and A. Heydari, "Field advice system using Bayesian Network," in *Proc. 15th Annual International Conference of Iranian Computer Forum*, Iran, 2010.

[20] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3nd edition, Prentice Hall, 2009.

[21] K. B. Korb and A. E. Nicholson, *Bayesian Artificial Intelligence*, *Computer Science and Data Analysis*, CRC/Chapman Hall, Boca Raton, 2004.

[22] K. P. Murphy. (2001). An introduction to graphical models. Technical Report, Intel Research. [Online]. Available: http://www.cs.ubc.ca/~murphyk/Papers/intro_gm.pdf

[23] Football356 Website, European Football Statistics. [Online].Available: http://stats.football365.com

[24] Soccer-Spain Website, Spanish Football Statistics. [Online].Available: http://www.soccer-spain.com

[25] Stat-Football Website, World Football Statistics. [Online].Available: http://www.stat-football.com/en/

[26] Netica, Bayesian network development software, Norsys Software Corp, Vancouver, Canada. (2012). [Online]. Available: http://www.norsys.com/netica.html

Farzin Owramipour was born in 1987 in Iran. Mr. Owramipour received his B.Engr. degree from University of Nabi Akram (UCNA), and his M.S. degree from Islamic Azad University, Zanjan branch (Zanjan , Iran) in computer engineering in 2012.

Parinaz Eskandarian was born in 1987 in Iran. Ms. Eskandarian received her B.Engr. degree from University of Tabriz Jahad Daneshgahi , and her M.S. degree from Islamic Azad University, Zanjan branch (Zanjan, Iran) in computer engineering in 2012.

Faezeh Sadat Mozneb obtained B.Eng. in hardware Computer Science and Engineering from Islamic Azad University, Central Tehran Branch (Tehran, Iran) in 2008. She obtained MSc in Computer Science and Engineering from Islamic Azad University, Qazvin branch (Qazvin, Iran) in 2012. During the last four years, she is doing research on wireless networks and teaching at Islamic Azad University. She is interested in innovative approaches in Computer Communications and Networks. Her focus is on architectures that unify connectivity, mobility and services; development of programmable service platforms, efficient overlays and P2P networks, end-to-end QoS control mechanisms, and security schemes for wireless mobile ad hoc networks (MANETs).