Knowledge-Based Systems 50 (2013) 60–86

Contents lists available at SciVerse ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Profiting from an inefficient association football gambling market: Prediction, risk and uncertainty using Bayesian networks ☆

Anthony Costa Constantinou *, Norman Elliott Fenton, Martin Neil

Risk & Information Management (RIM) Research Group, Department of Electronic Engineering and Computer Science, Queen Mary, University of London, London E1 4NS, United Kingdom

ARTICLE INFO

Article history:
Received 4 December 2012
Received in revised form 20 May 2013
Accepted 21 May 2013
Available online 4 June 2013

Keywords:
Bayesian networks
Expert systems
Football betting
Football forecasts
Subjective information

ABSTRACT

We present a Bayesian network (BN) model for forecasting Association Football match outcomes. Both objective and subjective information are considered for prediction, and we demonstrate how probabilities transform at each level of model component, whereby predictive distributions follow hierarchical levels of Bayesian inference. The model was used to generate forecasts for each match of the 2011/2012 English Premier League (EPL) season, and forecasts were published online prior to the start of each match. Profitability, risk and uncertainty are evaluated by considering various unit-based betting procedures against published market odds. Compared to a previously published successful BN model, the model presented in this paper is less complex and is able to generate even more profitable returns.

1. Introduction

Association Football (hereafter referred to as simply *football*) is the most popular sport internationally [10,27,11], and attracts an increasing share of the multi-billion dollar gambling industry; particularly after its introduction online [6]. This is one of the primary reasons why we currently observe extensive attention paid to football odds by both academic research groups and industrial organisations who look to profit from potential market inefficiencies. While numerous academic papers exist which focus on football match forecasts, only a few of them appear to consider profitability as an assessment tool for determining a model's forecasting capability.

Pope and Peel [30] evaluated a simulation of bets against published market odds in accordance with the recommendations of a panel of newspapers experts. They showed that even though there was no evidence of abnormal returns, there was some indication that the expert opinions were more valuable towards the end of the football season. Dixon and Coles [8] were the first to evaluate the strength of football teams for the purpose of generating profit against published market odds with the use of a time-dependent Poisson regression model based on Maher's [26] model. They formed a simple betting strategy for which the model was profitable at sufficiently high levels of discrepancy between the model and the bookmakers' probabilities. However, these high discrepancy levels returns were based on as low as 10 sample values; at lower discrepancy levels and with a larger sample size the model was unprofitable. The authors suggested that for a football forecast model to generate profit against bookmakers' odds without eliminating the in-built profit margin, *"it requires a determination of probabilities that is sufficiently more accurate from those obtained by published odds"*. A similar paper by Dixon and Pope [9] was also published on the basis of 1993–1996 data and reported similar results. Rue and Salvesen [32] suggested a Bayesian dynamic generalised linear model to estimate the time-dependent skills of all the teams in the English Premier League (EPL) and English Division 1. They assessed the model against the odds provided by Intertops, a firm which is located in Antigua in the West Indies, and demonstrated profits of 39.6% after winning 15 bets out of a total of 48 for EPL matches, and 54% after winning 27 bets out of a total of 64 for Division 1 matches.

In an attempt to exploit the favourite-longshot bias for profitable opportunities, Poisson and Negative Binomial models have been used to estimate the number of goals scored by a team [3]. The conclusion was that even though the fixed odds offered against particular score outcomes did seem to offer profitable betting opportunities in some cases, these were few in number. Goddard and Asimakopoulos [17] proposed an ordered probit regression model to forecast EPL match results in an attempt to test the

* Corresponding author.
E-mail addresses: anthony@constantinou.info (A.C. Constantinou), norman@eecs.qmul.ac.uk (N.E. Fenton), martin@eecs.qmul.ac.uk (M. Neil) .

0950-7051/$ - see front matter © 2013 The Authors. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.knosys.2013.05.008

Table 1
How $S \rightarrow S_R$ is defined in 14 predetermined ranks, based on [7].

S	>89	85–89	80–84	75–79	70–74	...(intervals of 5 points)	25–29	20–24	<20
S_R	1	2	3	4	5	...	12	13	14

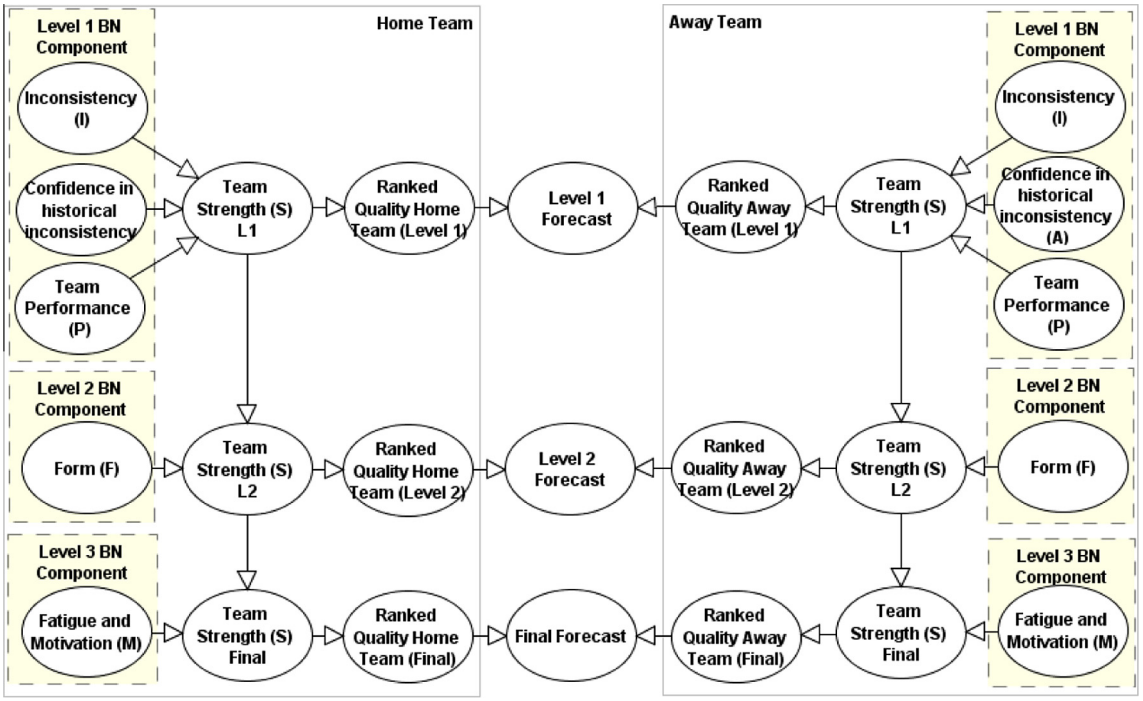

Fig. 1. Simplified model topology of the overall Bayesian network.

weak-form efficiency of prices in the fixed-odds betting market. To evaluate the model they considered seasons 1999 and 2000. Even though they reported a loss of −10.5% for overall performance, the model appeared to be profitable (on a pre-tax gross basis) at the start and at the end of every season.[1] Using a benchmark statistical model with a large number of quantifiable variables relevant to match outcomes Forrest et al. [15] examined the effectiveness of forecasts based on published odds and forecasts generated. They considered five different bookmaking firms for five consecutive seasons (1998–2003) and demonstrated that the model generated negative returns ranging from −10% to −12% depending on the bookmaking firm, but the loss was reduced to −6.6% when using the best available odds by exploiting arbitrage between bookmaking firms.

[19] attempted to investigate the rationality of bookmakers' odds using an ordered probit model to generate predictions for EPL matches. By considering William Hill odds, they followed the betting strategy introduced in [8,9] and reported negative returns ranging from −2.5% to −15% for all discrepancy levels during seasons 2004–2006. In the absence of any consistently successful model against market odds, the authors claimed that "*if it was successful, it would not have been published*". [21] considered the ELO rating system for football match prediction, although it was initially developed by [12] for assessing the strength of international chess players. Even though the ratings appeared to be useful in encoding the information of past results for measuring the strength of a team, resulting forecasts reported negative expected returns against numerous seasons of published odds using various betting strategies. However, Constantinou and Fenton [5] later developed a novel rating technique (called pi-rating) that outperformed considerably the two ELO rating variants of [21], in terms of profitability, over a period of five EPL season.

[7] recently presented a Bayesian network model that was used to generate forecasts about the EPL matches during season 2010/2011, by considering both objective and subjective information for prediction. Forecasts were published online [29] prior to the start of each match, and this was the first academic study to demonstrate profitability that was consistent against published market odds over a sufficiently high number of betting trials without eliminating the bookmakers' profit margin.

In this paper we present a Bayesian network model for forecasting football outcomes that is based on the approach in [7], but with reduced complexity and higher forecasting capability (which we explain in detail in Sections 2–4). The paper is organised as follows: Section 2 describes the model; Section 3 presents the various betting procedures along with a Bayesian network component for assessing the risks involved under each of the procedures; Section 4 discusses the results; Section 5 provides our concluding remarks.

2. The model

In this section we first provide a brief overview of the model summarising the main differences to the approach in [7]. We then describe the technical components of the model in subsections.

As in [7] we have used the AgenaRisk Bayesian network tool to build the model. The most important differentiator between AgenaRisk and other Bayesian network tools is its ability to properly incorporate continuous variables, without any constraint, and without the need for static discretisation. It does this through

[1] Gross pre-taxed returns of +3.1% and +1.5% for respective seasons beginning 1999 and 2000, and gross returns of +8% for respective seasons ending 1999 and 2000.