A Weighted-Goal-Score Approach to Measure Match Importance in the Malaysian Super League

Nor Hayati Abdul Hamid
e-mail: nhayati@tmsk.uitm.edu.my

Norazan Mohamed Ramli
e-mail: norazan@tmsk.uitm.edu.my

Universiti Teknologi MARA,
40450 Shah Alam, Selangor, MALAYSIA

Graham Kendall
e-mail: gxk@cs.nott.ac.uk

Automated Scheduling, Optimisation and Planning (ASAP) Group,
School of Computer Science, University of Nottingham,
Nottingham NG8 1BB, UK

Naimah Mohd Hussin
Universiti Teknologi MARA Perlis Kampus Arau,
02600 Arau, Perlis, MALAYSIA
e-mail: naimahmh@perlis.uitm.edu.my

1. Introduction

This work is a continuation of our previous work which aims to define the level of importance of each fixture in the Malaysian Super League. Malaysian football is witnessing a decrease in stadium attendance and scheduling fixtures into timeslots so as to maximise the number of supporters is critical to the league administrators. In our previous work (Abdul-Hamid et. al, (2010), we have applied AHP (Analytic Hierarchy Process) in order to calculate the priorities of each fixture.

The aim of this study is to investigate a different method to measure match importance. We propose to measure match importance using a statistical model.

The result will be used as an input to schedule the Super league fixtures with the aim of maximising stadium attendance.

2. Related Work

Wright (2009) recently reviewed 50 years of sports research, with forecasting being one of the categorized activities. Forecasting is usually associated with gambling but in Malaysia betting on matches is prohibited by law. Nonetheless, we consider forecasting as a tool to predict match importance, and thus enable us to schedule more effectively.

There are a number of papers that make use of statistical methods for forecasting. Maher (1982) and Dixon and Coles (1997) use independent Poisson distributions for the number of goals scored by the home and away teams. Dixon and Robinson (1998) developed a model to predict the results of football matches and updated their predictions during the course of a match. Koning (2000) also used a statistical model to assess the balance of a competition. In the work of Koning et al. (2003), they develop a simulation/probability model that identifies the team that is most likely to win a tournament based on a scoring intensity measure.

Min et al. (2008) developed a framework for sports prediction using Bayesian inference and rule-based reasoning, together with an in-game time-series approach. Based on the framework, they developed a football results predictor called FRES (Football Result Expert System).

Other than football, there has been work on forecasting for other sports. Boulier and Stekler (2003) evaluates power scores as predictors of the outcomes of the NFL (National Football League) (1994–2000 seasons). There has also been research on predicting success at the Olympic games. Condon et al. (1999) used linear regression and neural network models to predict a country's success during the Summer Olympic Games, and Heazelwood (2006) used mathematical models to predict elite performance in swimming and athletics at the Olympic games.

For additional information on sports research, the interested reader is referred to Kendall et al. (2010) which references about 160 papers going back over 40 years.

3. Proposed methodology

Based on previous research (Maher 1982, Dixon and Cole 1997, Dixon and Robinson 1998), we can model soccer matches using Poisson distributions, for example, the number of goals scored over a season. There are several factors to be considered when modeling data with a Poisson distribution. For example, competing teams should have comparable ability which in our case is reasonable as we are considering the Malaysian Super League. Other factors include home and away performance whereby (typically) teams perform better on their home ground. The composition of teams is another factor as players change based on their contracts.

We were initially going to adopt the independent Poisson model initially used by Maher (1982) and also applied by Dixon and Coles (1997) and Dixon and Robinson (1998). In this model, the home team's and away team's scores in any one match are independent Poisson variables. In a match between team i and j, let $X_{i,j}$ and $Y_{i,j}$ be the number of goals scored by the home and away sides respectively. We will assume that $X_{i,j}$ is Poisson with mean $\alpha_i \beta_j$, that $Y_{i,j}$ is also Poisson with mean $\gamma_i \delta_j$, and that $X_{i,j}$ and $Y_{i,j}$ are independent. We can then assume that α_i represents the strength of team i's attack when playing at home, β_j the weakness of team j's defence when playing away, γ_i the weakness of team i's defence at home and δ_j the strength of team j's attack away.

However, the Chi-Square Goodness of Fit test on our data showed that $X_{i,j}$ and $Y_{i,j}$ do not follow Poisson distribution. Thus we propose another statistical approach which we call weighted-goal-score, inspired by the work of Norazan et al. (2009) who proposed a weighted approach to downweight suspected outliers. We adapt the approach in the context of goal scores by finding the level of importance of each match using weighted scores. The motivation for using this method is to investigate if replacing the priorities obtained through AHP (the focus of our previous work) with a statistical model can provide comparable, or even superior, results.

The weighted-goal-score is modeled as follows :

Let :

X_i = goals scored for team i in all their home matches.

Y_j = goals scored for team j in all their away matches.

$$H = \sum_{i=1}^{n} X_i$$

$$A = \sum_{j=1}^{n} Y_j$$

where n is the number of teams.

$$w_i = \frac{X_i}{H}$$

$$w_j = \frac{Y_j}{A}$$

Therefore, the level of importance l, of a match between team i and team j is given by

$$l = w_i \,.\, w_j$$

We will use l, as a measure of match importance, in the mathematical model of our previous work (Abdul-Hamid et al., 2010) in order to generate schedules which are attractive to supporters.

We will report our results at the conference.

4. References

Abdul-Hamid N., S. M., Pais T., Kendall G., Subramaniam E.R , Mohd-Hussin N. (2010). A Combined Mathematical Modelling and AHP Approach to Solve Sport Scheduling Problem. Under review.

Boulier, B. L., & Stekler, H. O. (2003). Predicting the outcomes of National Football League games. *International Journal of Forecasting, 19*(2), 257-270.

Condon, E. M., Golden, B. L., & Wasil, E. A. (1999). Predicting the success of nations at the Summer Olympics using neural networks. *Computers & Operations Research, 26*(13), 1243-1265.

Dixon, M. J., & Coles, S. G. (1997). Modelling Association Football Scores and Inefficiencies in the Football Betting Market. *Journal of the Royal Statistical Society. Series C (Applied Statistics), 46*(2), 265-280.

Dixon, M. J., Robinson, M.E. (1998). A birth process model for association football matches. *Journal of the Royal Statistical Society Series D-the Statistician, 47*, 523-538.

Heazlewood, T. (2006). Prediction versus reality: The use of mathematical models to predict elite performance in swimming and athletics at the Olympic Games. *Journal of Sports Science and Medicine, 5*(4), 541-547.

Kendall, G., Knust, S., Ribeiro, C. C., & Urrutia, S. (2010). Scheduling in sports: An annotated bibliography. *Computers & Operations Research, 37*(1), 1-19.

Koning, R. H. (2000). Balance in competition in Dutch soccer. *Statistician, 49*, 419-431.

Koning, R. H., Koolhaas, M., Renes, G., & Ridder, G. (2003). A simulation model for football championships. *European Journal of Operational Research, 148*(2), 268-276.

Maher, M. J. (1982). Modelling association football scores. *Statistica Neerlandica, 36*(3), 109-118.

Min, B., Kim, J., Choe, C., Eom, H., & McKay, R. I. (2008). A compound framework for sports results prediction: A football case study. *Knowledge-Based Systems, 21*(7), 551-562.

Norazan, M. R., Habshah, M., & Imon, A. (2009). Weighted Bootstrap with Probability in Regression. In S. Y. Chen & Q. Li (Eds.), *Proceedings of the 8th Wseas International Conference on Applied Computer and Applied Computational Science - Applied Computer and Applied Computational Science* (pp. 135-141).

Wright, M. B. (2009). 50 years of OR in sport. *Journal of the Operational Research Society, 60*, S161-S168.