

ANNOUNCEMENT

BACHELOR THESIS

KEYWORDS

- Empirical Analysis
- Prediction models
- Sports betting
- Bayesian Inference
- Probabilistic Programming

TOPIC: IS THERE A FOOTBALL GOD? - PREDICTING ASSOCIATION FOOTBALL GAME OUTCOMES

Association football (in the following just called “football”) is the most popular sport in the world and the related gambling market is known for its impressive growth rates (Constantinou et al. 2012). Thus, research as well as the industry developed diverse statistical models to predict tournament outcomes, league positions, match outcomes or even goals.

The ability to quantify beliefs in future sport game outcomes can also be very valuable from a marketer’s perspective. For example, this information can be used as a basis for estimating sponsorship costs and product demand or even for developing brand new merchandizing products for a team that is likely to win in the future.

Early work tried to predict individual goals based on the Poisson distribution. Especially, a popular paper of Maher (1982) laid the foundation for many following studies (Rue Salvessen 2000; Karlis & Ntzoufras, 2003). Another common approach focused on the prediction of match results (i.e. win, draw, or lose) using ordered probit regression models exploiting explanatory variables like the team quality, recent performance or the geographical location. Recently, also various machine learning techniques (e.g., Bayesian networks, neural networks, genetic programming and fuzzy models) were used to achieve better results and “beat the bookies”.

Another literature stream tries to predict match outcomes based on team quality ratings. During the FIFA World Cup, Andrew Gelman (2014) published a well-recognized blog post where he tried to predict the results. In this post, he explained how to effectively forecast world cup rankings by leveraging team quality ratings: The trick is to calculate score differentials for each game and use this information to update beliefs in future outcomes. To model such a

Bayesian setting with minimal afford, Gelman used a probabilistic programming language called Stan¹.

The goal of this bachelor thesis is to compare the performance of Gelman's model across different data sets (e.g., English Premier League, Spanish Primera División or German Bundesliga) and test possible model extensions to improve the predictive ability. For the analyses, we will provide a real-world data set comprising a wide range of diverse football league results over the last years.

To write this thesis, you will have to use probabilistic programming languages (e.g., Stan, pyMC) and statistic programming environments (e.g., R or Python). Thus, interest in quantitative methods and basic statistical knowledge is required.

LITERATURE:

- **Kharratzadeh, M. (2016):** Hierarchical Bayesian Modeling of the English Premier League. https://github.com/milkha/EPL/raw/master/epl_rev.pdf
- **Gelman (2014):** Stan goes to the World Cup. <http://andrewgelman.com/2014/07/13/stan-analyzes-world-cup-data/>
- **Constantinou A., Fenton N., & Neil M. (2012):** pi-football: A Bayesian network model for forecasting association football match outcomes. Knowledge-Based Systems, 36, 322-339.
- **Rue H., & Salvesen O. (2000):** Prediction and retrospective analysis of soccer matches in a league. The Statistician, 3, 339-418.
- **Karlis D., & Ntzoufras I. (2003):** Analysis of sports data by using bivariate Poisson models. The Statistician, 52:3, 381-393.
- **Maher, M. J. (1982):** Modelling association football scores. Statistica Neerlandica, 36:3, 109-11.

SUPERVISOR:

- Dr. Nils Wlömert: www.wu.ac.at/imsm/team/nils-wloemert/en/
- DI Christian Hotz-Behofsits www.wu.ac.at/imsm/jobs/team/christian-hotz-behofsits/

APPLICATIONS:

Applications with CV and transcript of records should be sent to Nils Wlömert (nils.wloemert@wu.ac.at).

¹ <http://mc-stan.org>