Relational Learning for Football-Related Predictions

Jan Van Haaren JAN.VANHAAREN@CS.KULEUVEN.BE
Guy Van den Broeck GUY.VANDENBROECK@CS.KULEUVEN.BE
KU Leuven, Celestijnenlaan 200A, 3001 Heverlee, Belgium

Keywords: statistical relational learning, application

Association football is becoming increasingly competitive and the financial stakes involved are causing football clubs and football leagues to become more professional. Over the past 25 years, club budgets have grown enormously due to ticket sale revenues, broadcasting revenues, merchandising, and prize money. Recently, player tracking systems were introduced and are producing overwhelming amounts of data which are being used by experts to analyze matches.

State-of-the-art approaches for predicting football match results fail to leverage the full range of rich data that is currently available. These approaches are mostly extensions of well-known statistical methods that learn models with limited expressivity. Two important reasons lead to these limitations. First, until recently match statistics were usually not publicly available. Second, it is not obvious how to derive meaningful statistics from football matches.

We propose using machine learning and, more specifically, relational learning to address the shortcomings of current prediction techniques (Van Haaren & Van den Broeck, 2011). Relational learning is particularly well suited and overcomes the two most important challenges. First, it considers a variety of aspects that influence a match result. Second, it reasons about time-dependent and positional information.

A relational model offers a lot of flexibility to represent the data since its parameter set is not fixed but varies according to the events that happen during a football match. Consequently, a relational model is able to stress rare, but important, events such as a red card. Furthermore, it can represent complexly structured data such as team lineups using relations among objects (e.g., football players). All of this is much harder or impossible in a propositional representation.

Due to their expressivity, relational models can tackle many interesting learning tasks such as regression and classification. However, more complex learning tasks such as collective regression (e.g., jointly predicting player statistics), collective classification (e.g., predicting a team's starting lineup) and link prediction (e.g., predicting who passes the ball to whom) take full advantage of a relational model's capabilities to handle rich structured data.

Experiments with kLog (Frasconi et al., 2011) show that our relational approach yields competitive results when compared to propositional models. We used match statistics from the 2010-2011 English Premier League season to predict match results and goal differences (see Figure 1), and classify matches according to their outcome. The experimental results for relational and propositional approaches are very similar, but we expect relational learners to outperform their propositional counterparts once we manage to exploit the relational structure of the data better.

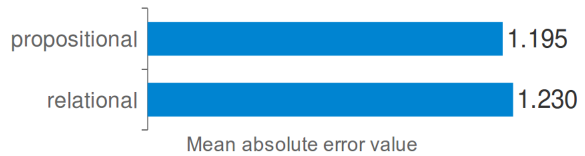

Figure 1. Our relational approach yields mean absolute error values for goal difference predictions that are similar to these of a propositional approach. Lower is better.

Acknowledgments

Guy Van den Broeck is supported by the Research Foundation-Flanders (FWO-Vlaanderen).

References

Frasconi, P., Costa, F., De Raedt, L., & De Grave, K. (2011). *kLog - A Language for Logic-Based Relational Learning with Kernels* (Technical Report).

Van Haaren, J., & Van den Broeck, G. (2011). Relational Learning for Football-Related Predictions. *Preliminary Papers ILP 2011* (pp. 1–6).