

Machine Learning for Soccer Analytics

Research group: Declarative Languages and Artificial Intelligence

Promoter: professor Jesse Davis

Daily supervisors: Jan Van Haaren and Guy Van den Broeck

Information: jan.vanhaaren@cs.kuleuven.be

1 Context

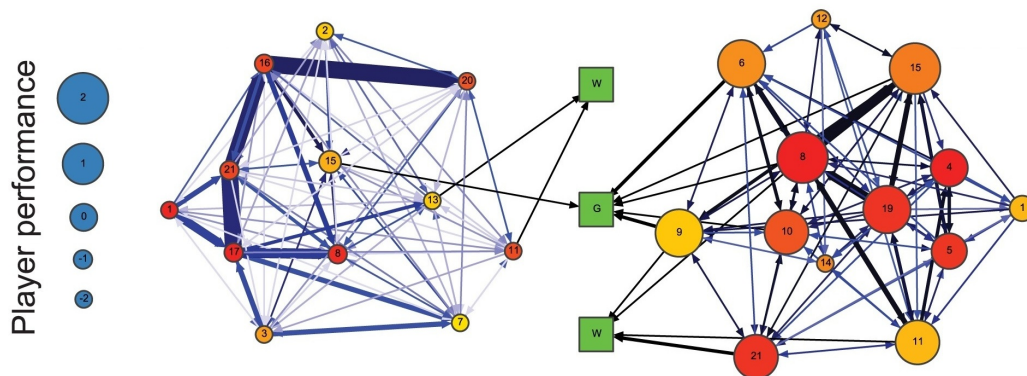
The field of performance analytics in soccer has been held back for many years due to a lack of publicly accessible data. Over the past two decades, performance analytics has flourished in major US sports such as baseball, basketball and American football, which have objective analysis within their fabric. Much of the analytics work now done in these sports has stemmed from the excellent pioneering work of sports enthusiasts. Making performance data publicly available drove these sports forward as a collective.

Manchester City recently set up an extensive soccer analytics programme to support the soccer analytics community. Manchester City's main priorities are position specific metrics to accurately measure player performance and compare players, seasonal player profiling and data visualization. In cooperation with data provider OPTA, Manchester City made an extensive dataset for the 2011-2012 English Premier League season publicly available, which contains detailed statistics for all players involved in each match.

2 Objective

The objective of this thesis is to develop metrics to measure the performance of both individual players and teams. We will develop a position specific metric for each position and combine these metrics into a team performance metric. We will investigate what the impact of a certain player on the team performance is. If time permits, we will also investigate more advanced performance metrics that account for player interactions.

The objective of this thesis is very flexible. Please, feel free to contact us and propose an alternative learning task that might be valuable to the soccer analytics community.



Player performances during the Germany vs. Spain match at EURO 2008.

Source: <http://www.axonpotential.com/the-new-soccer-metric-flow-centrality/>

3 Approach

The proposed approach of this thesis is as follows:

1. Perform a literature study on performance analysis in soccer (e.g., Capello Index).
2. Develop position specific performance metrics and combine these metrics into a team performance metric. We will employ Weka Toolbox, which is a collection of popular data mining and machine learning algorithms.
3. Extend the position specific performance metrics to account for player interactions. For example, two players might be complementary and perform better as a result. We will employ a relational learning language such as Markov Logic.

4 Literature

Please contact us if you are interested in working on this topic or have further questions. We will provide you with appropriate reading material given your prior knowledge about machine learning, data mining and soccer analytics. The references listed below provide more information on the Weka Toolbox, Markov Logic, an existing performance metric in soccer and Manchester City's soccer analytics programme.

- **Weka Toolbox**

M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I.H. Witten (2009). *The WEKA Data Mining Software: An Update*, SIGKDD Explorations, volume 11.

- **Markov Logic**

M. Richardson and P. Domingos (2006). *Markov Logic Networks*, Machine Learning Journal, volume 62, pp. 107-136.

- **Capello Index**

http://en.wikipedia.org/wiki/Capello_Index

- **MCFC Analytics web page**

<http://www.mcfc.co.uk/The-Club/MCFC-Analytics>

5 Profile

A motivated student with interest in machine learning and data mining who has good implementation skills. Experience with the Weka Toolbox and/or Markov Logic as well as an understanding of soccer is a significant advantage but is not required.