INFO 6205

Final Project

Ranking System

**Introduction**

**What is Ranking System?**

Statistical Football prediction is a method used in sports betting, to predict the outcome of football matches by means of statistical tools. The goal of statistical match prediction is to outperform the predictions of bookmakers, who use them to set odds on the outcome of football matches.

The most widely used statistical approach to prediction is ranking. Football ranking systems assign a rank to each team based on their past game results, so that the highest rank is assigned to the strongest team. The outcome of the match can be predicted by comparing the opponents’ ranks. Several different football ranking systems exist, for example some widely known are the FIFA World Rankings or the World Football Elo Ratings.

There are three main drawbacks to football match predictions that are based on ranking systems:

1. Ranks assigned to the teams do not differentiate between their attacking and defensive strengths.
2. Ranks are accumulated averages which do not account for skill changes in football teams.
3. The main goal of a ranking system is not to predict the results of football games, but to sort the teams according to their average strength.

**Aim of the project**

Predicting the final standing of the EPL if all the remaining matches were also played by applying the ranking system. The input to our system will be a set of prior encounters with a result. With the objective to develop a ranking system capable of evaluation the expression P(xi,xj) where xi and xj are competing team and P is the probability of x beating y in a head to head matchup. The probability should be given as a probability density function of resultant match scores.

**Dataset used for this project:** In the experimental evaluation, I have used a dataset of historical football match results. The dataset contains match results for 10 seasons of English Premier League. The dataset dates back to the 2009 season and contains nearly 3800 match results.

**Project Description**

Create a ranking system with Object Oriented Programming and Data Structures.

* Using Map structure to store the dataset of historical football match results, with key as the team name and value as an Array of game results.
* Using a RankTable class to hold the home rank and away rank of all teams based on prior encounters
* Building a probability distribution function which denotes the probability that team A will beat team B in a head-to-head matchup

**Implementation**

To evaluate the model, we split the match results into two sets: a training set and a test set. We use the training set to learn the parameters of the model and the test set to assess the performance of the model. We predict the outcomes of the matches in the test set in an iterative way. We use the model learned on the training set to predict the outcomes of matches. We do not update the parameters of the model using the expected outcomes and prediction for current season is determined with previous season’s results only. We repeat this procedure until all matches have been predicted.

To account for the fact that performances tend to vary. We predict a match outcome by first computing the probability distribution over the possible outcomes and then sampling an outcome from this distribution

Rating of the team will be , where is the average skill (or mean) and is the degree of uncertainty (standard Deviation)

A close up of a plane

Description automatically generated

We calculate these bell curves for each team as it plays with all other team played at home and away. And add each of these resulting curves by these formulas.

A strong team will have a thin and tall bell curve like this, resulting in ranking would be calculated with

A close up of a logo

Description automatically generated

Whereas weaker team would have short and wide bell curve like this

A screenshot of a cell phone

Description automatically generated

**Models**

Implementing the Split Elo-rating model (SEM): This is a modification implementation of the ELO rating system, which represents each team by two ratings. The first rating represents the strength at home, while the second rating represents the strength of team on away matches.

Rating of a team is calculated using

A picture containing clock

Description automatically generated

Where Rcur and Rnew denote the rating of a team before and after a match. The parameter I is a positive number that denotes the importance of the match and G denotes the importance of the goal difference (which in our calculation are both 1). The parameters Ract and Rexp represent the actual and expected output of the match respectively. The value of Rexp is calculated using the following formula, where Raway and Rhome are denoting the ratings of home and away team respectively.

A picture containing object, clock

Description automatically generated

**Output**

* Screenshot of the output of the program

A close up of text on a black background

Description automatically generated

A screenshot of a cell phone

Description automatically generated

**Conclusion**

Using this Ranking System, I have tried to predict the EPL standing of the previous season. Where the top five teams were predicted with 60% correct and last three teams in the standings were predicted with 100% accuracy. Keeping these results in mind, I believe this ranking system will predict this season’s standing with about 75% accuracy.

**Reference**

1. Data Source

<https://datahub.io/sports-data/english-premier-league#data>

1. Wiki about statistical Association football Predictions

<https://en.wikipedia.org/wiki/Statistical_association_football_predictions>

1. Understanding ranking system like true skill

<https://www.youtube.com/watch?v=VnOVLBbYlU0>