Who's winning it? Forecasting sports tournaments

Karolina Guzauskaite
KTU Gymnasium, Lithuania

Mark Shepelev
EISP, Czechia

Leon Šopar
Croatia

Paja Stojkovič
Serbia

Vladimir Dzyuba DZYUBAV+S3@GMAIL.COM
KU Leuven, Belgium

Abstract

The purpose of this project was to highlight the importance of machine learning and the ability of the said learning to be adapted to almost any situation with a large amount of data. In this particular study the focus was on simulating a football tournament with usage of machines algorithms as forecasters. They would be imputed historical data and a prediction would come as a result.

1. Introduction

In 2016, Portugal won the European Football Championship. The goal of our project is to simulate the entire tournament to determine the probability of each team winning it at every stage of the tournament.

The methods that we use to determine winners of every stage of the tournament can also, given enough accurate data and by interchanging the parameters used in our calculations, be used to simulate virtually anything, ranging from wars, stock markets, epidemics, to something as small as protein synthesis, DNA replication, or mutations.

Machine learning is a process where the computer automatically teaches itself a program based on the data we feed it. Based on the data of all the national team matches since 1973 up until the start of the European Football Championship 2016 we have trained our program to calculate the probabilities of each of the participants taking home the trophy at each of the tournament stages (group stage, round of 16, quarter-finals, semi-finals, and the final).

Summer School of Science. July 19-27, 2017. Požega, Croatia.
The first four authors contributed equally to the work.

2. Related work

During the years, there were multiple attempts to predict the outcome of sports tournaments. We focus on association football, also known as soccer. Maher created a model where the goals scored by the home and away teams follow the Poisson distribution (Maher, 1982). Yet, it was unable to predict outcomes of future matches. Dixon and Coles used a similar approach and developed a model that generated match outcome probabilities (Dixon & Coles, 1997). Rue and Salvesen assumed that offense and defense parameters are time-varying, and used Bayesian methods to update parameter estimates (Rue & Salvesen, 2000).

Even though, to build a simulation various strategies can be used, finding a strategy which is well documented, performs well and is relatively easy to use may be challenging (Breiter & Carlin, 1997). In our case, it is very important to choose the right team ranking system. Lasek et al. aims to compare predictive capabilities of several ranking methods in association football (Lasek et al., 2013). They found that the FIFA rankings are not able to use the information on past results efficiently and quickly react to changes in team performance. As a result, it is not very accurate for tournament outcome predictions. In addition to the FIFA rankings, which are solely based on the position of team in table, Lasek et al. examined ranking systems based on rating points. They found the ELO rating system to be very competitive.

The ELO system originally was developed by Arpad Elo to estimate the strength of chess players (Elo, 1978). Nowadays it is applied to many various sports. It always shows the updated results based on the latest team performance. Considering this property, we decided to use the ELO ranking system for assigning the strength of the teams. Saraiva et al. introduced home field advantage into the ELO system (Saraiva et al., 2016). The main idea is that teams playing at home has a greater chance of winning the match. Therefore, in order to get reliable results, we took this property into account. In our work we used Poisson regression to estimate the effect of the strength of a given team and its opponent on the number of goals scored. We simulate goals scored with the predicted number of goals as the parameter of the Poisson distribution. This way, all the parameters were considered: attack and defense strengths, home field advantage. Finally, to make inferences we use iterative Monte Carlo simulation techniques to improve the accuracy of our model.

3. Method

Our goal was to simulate a tournament, The European Football Championship (EURO 2016) had two stages, the group stage and the knockout stage (see below). In order to simulate the two stages, we developed the ELO rating system and the goal simulation component.

3.1. ELO ranking

The ELO system is a system developed by Hungarian physicist and chessmaster Arpad Elo (Elo, 1978). It is most commonly used in chess. It has the task of assigning a number to each player, which is representative of their strength based on the matches that they played. In this project, we have expanded the parameters that go into calculating ELO, as football in itself does not use this system in its official rankings. Namely, all the participants start with the same ELO rating and that rating increases or decreases based on their performance. ELO points gained or lost are calculated using the formula:

$$R_1' = R_1 + (O - O_{exp} \times (R_1 - R_2)) \times K \times I \times N \times G \quad (1)$$

where R_1' is the updated rating of Team 1 after a match. R_1 is the rating of a team prior to the match and R_2 is the ELO rating of the opposing team before the match. O represents actual outcome of a match which can be a win, a draw, or a loss, and has the value of 1; 0.5; 0 respectively. K is a constant. The value O_{exp} (expected outcome) is a value between 0 and 1, which determines the probability of a team winning 0 for 0% chance and 1 for 100% chance. It is calculated through the formula:

$$O_{exp} = 1 \Big/ 1 + 10^{R_2 - (R_1 + H)/400} \quad (2)$$

The parameter H has been added to the formula to represent the home field advantage of a team. Namely, it gives an initial "artificial" bonus to the ELO rating of a team playing at home. The additional parameters added for the rules of football are I, which represents the importance of a game (Friendly, World Cup, World Qualifications, etc.), G, that depends on the goal difference between teams in a game and N, which depends on the actual goals scored by a team in a game. We set G and N manually. The values of K, I, and H are chosen experimentally, based on the predictive accuracy of the resulting ratings; see Section 4.

The ELO system makes defeats to higher rank teams less punishing and wins against those teams more rewarding. By the same logic, defeats against lower ranked teams bring a higher decrease in ranking while wins against lower ranked teams bring a smaller reward based on how much weaker the team is. However, there is a number of problems with the ELO system in football:

1. teams change and it takes a while for a team to fall/rise to their true rating after a change
2. teams on different continents do not play against one another, resulting in some teams having stricter competition that the others, leading to less accurate ratings.

3.2. Monte Carlo simulation

The Monte Carlo simulation is a type of simulation that requires the same event to play out many times, independently, to conclude the probability of all the outcomes as accurately as possible.

3.3. Goal prediction model

The aim of goal prediction model is to use available data to determine how many goals a team is likely to score and calculate probabilities for different match outcomes finding the most expected one. To achieve that we will use Poisson regression and Poisson distribution.

Poisson distribution is a mathematical concept that can be used to measure the probability of independent events occurring a certain number of times within a specific time interval. When the average number of times an event will happen is known, Poisson distribution can be used to calculate how likely other outcomes deviate from the average.

In our work, we want to determine the most likely outcome of the match. It turns out that goals scored by a team in a football match are approximately Poisson distributed: events can be counted in integer numbers, occurrences of the event are independent, and it is possible to count how many events have occurred. Therefore, we can use Poisson distribution to assign probabilities for different match outcomes and find the most-likely score line of the match. To use this method, we need to calculate the average number of goals each team is likely to score in that match. This is going to be our parameter of the Poisson distribution, λ. Note that it is very important to choose the right values of λ to get reliable results. For this purpose, we used Poisson regression model.

With Poisson regression we can easily estimate λ based on certain variables. What we want to consider are team strengths (as different teams are expected to score and concede different number of goals) and home field advantage as teams playing at home are likely to score more goals than away teams. Note that in Poisson regression each observation is assumed to be independent from the others and each match is included twice in our dataset.

3.4. EURO 2016 rules

Group stage In the Group Stage a total of 24 teams are given are a chance to advance to the knockout stage. All teams are split into 6 groups of 4 teams, which all play each other. A win gives 3 points, a loss gives 0 points and draw gives 1 point to both teams.

As such, teams can be sorted by points which awards them a 1-to-4 rank. Teams 1 and 2 advance automatically, whilst only 4 out of 6 third-placed teams advance to knockout.

In case there is a point tie within the group the following rules are used to resolve the tie:

1. Higher number of points obtained in the matches played between the teams in question;
2. Superior goal difference resulting from the matches played between the teams in question;
3. Higher number of goals scored in the matches played between the teams in question;
4. If, after having applied criteria 1 to 3, teams still had an equal ranking (e.g. if criteria 1 to 3 were applied to three teams that were level on points initially and these criteria separated one team from the other two who still have an equal ranking), criteria 1 to 3 would be reapplied exclusively to the matches between the teams who were still level to determine their final rankings. If this procedure did not lead to a decision, criteria 5 to 9 would apply;
5. Superior goal difference in all group matches;
6. Higher number of goals scored in all group matches;
7. If only two teams had the same number of points, and they were tied according to criteria 16 after having met in the last round of the group stage, their ranking would be determined by a penalty shoot-out. (This criterion would not be used if more than two teams had the same number of points.);
8. Fair play conduct (1 point for a single yellow card, 3 points for a red card as a consequence of two yellow cards, 3 points for a direct red card);
9. Position in the UEFA national team coefficient ranking system.

The following rules are used to determine the winners among the third-ranked teams:

1. Higher number of points obtained;
2. Superior goal difference;
3. Higher number of goals scored;
4. Fair play conduct;
5. Position in the UEFA national team coefficient ranking system

Due to practical reasons, we implemented the modified version of the rules, namely we only implemented the first seven tiebreakers in the first group and the first three tiebreakers in the second group, with criteria 2 and 3 swapped.

Knockout stage After the group stage comes the knockout stage. The 16 qualified teams are placed into 8 pairs based on official FIFA regulated rules and play one another. The defeated team is automatically eliminated from the tournament and the winner advances to the next stage, where it plays the winner from the neighbour bracket. This process is repeated 4 times, until there is only one team remaining in the tournament. That team is then pronounced the winner of the tournament.

Unlike the group stage, the knockout stage does not account for the probability of a draw between two teams, as each match has to have a definite winner to determine the team that advances and the team that is “knocked out”.

As the knockout stage has no regard for the numbers of goals scored in the match it does not use the goal simulation system. It instead uses a kind of a “biased coin toss”, which gives a higher win probability to a team with a higher ELO rating using the formula for the expected outcome.

4. Experiments

4.1. Experimental setup

Dataset The data used is concerned with football matches. The games played are between 30 November

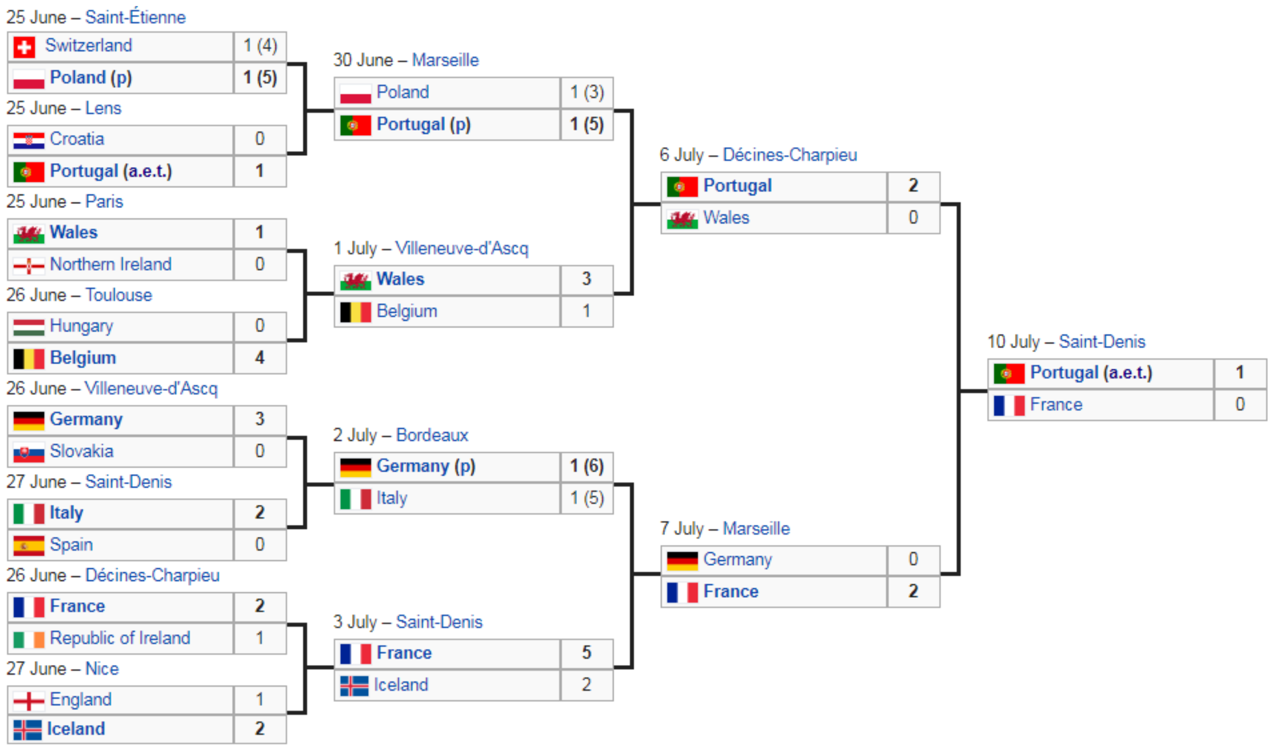

Figure 1. The knockout stage as it actually happened during EURO 2016.

1872 and 7 October 2016. For each match we have exact date, type of the game (Friendly, World Cup, etc.), playing teams and goals scored by each team.

Validation For ELO system parameters tuning we included validation set. Our validation set covers 2291 games between 1 January 2014 and 8 June 2016.

Training set Our training set covers 30 000 matches played between 1973 until the end of 2013. We used the training set to build the ratings and pick the parameter values, with which we had the lowest prediction error on the validation set. We evaluated prediction error using log loss function.

4.2. Results

ELO We estimated team strengths using ELO ranking system. To build the ratings we had to do pick the best parameter values with which we had the lowest prediction error on the validation set. According to our experiments, the best K is 50, the best starting date for the training set is 1 January 1974, the best home field advantage parameter H is 100. Our model showed that considering the importance of the game I (whether it is a friendly game, or World Cup, etc.) does not improve the performance of our system. The error of the forecasting system is 0.40, which is lower than 0.48, the error of random guessing. See Table 1 for a subset of the final ranking.

Goal prediction model To build the goal prediction model, we were mostly interested in the goals scored by each team during the time between 17 January 1974 until the end of 2013 and ELO ratings we have calculated previously. Using Poisson regression, we estimated the effect of the strength of a given team and its opponent on the number of goals scored. To put it simply, Poisson regression is a learning method, which uses historical data to determine best values of coefficients, which are used in calculating the best λ. For example, imagine we need to calculate λ for the match where Germany plays against Andorra. To do that, we can use this equation:

$$\begin{aligned} ln\left(\lambda_{\text{GER vs. AND}}\right) = 0.90 + 0.0020 \times \text{ELO}_{\text{GER}} \\ - 0.0024 \times \text{ELO}_{\text{AND}} \end{aligned} \quad (3)$$

Here 0.90, 0.0020 and 0.0024 are all coefficients learned by Poisson regression. 0.90 is intercept; 0.0020 and 0.0024 are specific for the team and its opponent. Therefore, using Poisson regression we can easily calculate the best value of lambda, which serves as a parameter for Poisson distribution which is used to predict the goals scored by Germany in the match. We observe that win probabilities resulting from Poisson simulations are very close to those directly predicted by the ELO ratings.

Table 1. ELO ratings according to our model. EURO 2016 participants are shown, along with top-10 teams in the world. The ranking generally matches the intuition, with stronger teams having considerably higher ratings.

1	Brazil	2404.7
2	Argentina	2380.4
3	Spain	2334.0
4	Germany	2281.7
5	Colombia	2278.2
6	Uruguay	2274.4
7	Mexico	2261.7
8	France	2260.7
9	Chile	2259.7
10	England	2253.3
11	Netherlands	2223.1
15	Belgium	2166.6
16	Portugal	2162.9
17	Croatia	2149.0
18	Italy	2147.1
23	Turkey	2109.6
24	Ukraine	2097.2
29	Switzerland	2062.0
30	Poland	2062.0
31	Russia	2058.6
32	Sweden	2057.9
33	Czechia	2052.3
34	Slovakia	2043.2
35	Ireland	2038.8
37	Romania	2033.4
47	Austria	2001.6
64	Hungary	1953.1
78	Wales	1900.3
79	Iceland	1898.7
90	Czechoslovakia	1863.6
91	Northern Ireland	1860.6
93	Albania	1855.5

Simulation Our simulation consists of two parts: the group stage (tiebreaking rules included) and knockout stage. We employed Monte Carlo simulation to increase accuracy of our model. During iterations results fluctuated until they started to level off, making the results as reliable as possible (Figure 2). After 10 000 iterations, Portugal chance of winning was around 3.8%, while Spain had the greatest probability to win.

5. Conclusions and future work

The aim of our project was to learn how to apply artificial intelligence algorithms to building a prediction model as the techniques applied can be widely used in various fields, such as biology, economy,etc. In our project the goal was to

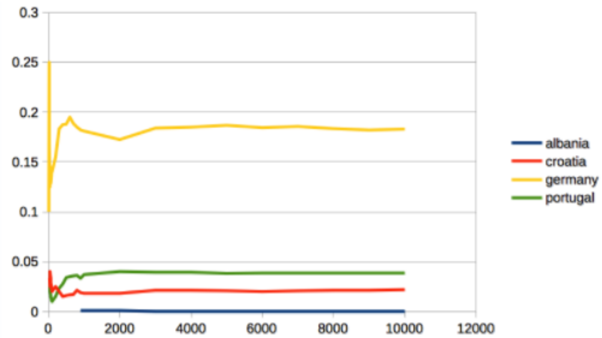

Figure 2. Probability estimates stabilize after 3 000 iterations. In general, the estimates match the intuition: for instance, Germany, one of the strongest teams, has one of the highest probabilities of winning, whereas Albania has one of the lowest.

simulate the entire tournament to determine the probability of each team winning it at every stage of the tournament. More specifically, we wanted to find out what probability each team had to win EURO 2016.

To develop our model we used historical data and considered such features as home field advantage, importance of the match, goals scored which all were meant to reflect the strength of the team. In our work we used the ELO ranking system to assign strengths to teams; Poisson regression, the approach to determine best value of parameter for Poisson distribution; Monte Carlo simulation to calculate the probability of all the outcomes as accurately as possible.

After 10 000 iterations our model showed that Portugal (the actual winner of EURO 2016) had 3.8% chance of winning. Even though, the probability of Portugal winning seems to be small, especially having in mind that it won after all, it is realistic. Additionally, the high chances of Spain or Germany winning the tournament is reasonable and seems to match intuition. Therefore, our model performs properly, though one should not rely on it completely.

The work we have done so far can be further improved in the future. Our prediction error turned out to be 0.40, which is 0.08 lower than random guessing error. To lower it we may have checked more different values of parameters to increase our chance of picking the ones which resulted in even lower error. Additionally, we may have considered more features, such as time-scaling, neighbors (Lasek et al., 2013) to take a more accurate approach on team strengths.

Acknowledgements

We would like to thank Sebastijan Dumančić and Nikolina Šostarić, the organizers of the Summer School of Science.

References

Breiter, David J and Carlin, Bradley P. How to play office pools if you must. *Chance*, 10(1):5–11, 1997.

Dixon, Mark J and Coles, Stuart G. 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, 1997.

Elo, Arpad E. *The rating of chessplayers, past and present*. Arco Pub., 1978.

Lasek, Jan, Szlávik, Zoltán, and Bhulai, Sandjai. The predictive power of ranking systems in association football. *International Journal of Applied Pattern Recognition*, 1 (1):27–46, 2013.

Maher, Michael J. Modelling association football scores. *Statistica Neerlandica*, 36(3):109–118, 1982.

Rue, Havard and Salvesen, Oyvind. Prediction and retrospective analysis of soccer matches in a league. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 49(3):399–418, 2000.

Saraiva, Erlandson F, Suzuki, Adriano K, Ciro Filho, AO, and Louzada, Francisco. Predicting football scores via poisson regression model: applications to the national football league. *Communications for Statistical Applications and Methods*, 23(4):297–319, 2016.