Methods for improving a team's position in the FIFA ranking

Jan Lasek
(joint work with Zoltán Szlávik and Sandjai Bhulai)

Institute of Computer Science
Polish Academy of Sciences

PhD Summer School, 2014

Outline

1. FIFA ranking for football teams
2. Scheduling of friendly games
3. Coalition of teams improving their ranks

Motivation

- FIFA ranking is used for scheduling competitions.
- The main motivation comes from the current Poland's position in the ranking:

FIFA World Ranking

67	Bolivia	483
68	El Salvador	481
69	Poland	474
70	Republic of Ireland	473
71	Trinidad and Tobago	470

FIFA ranking methodology

Key components of FIFA ranking:

- **Main idea:** Compute yearly averages of points gained in individual games and add them up with decreasing weights in time.
- It takes into account 4 years of play.
- The points for a single game are computed according to a formula

$$N = M \cdot I \cdot S \cdot C,$$

  where:
  - the outcome of the game (M points),
  - the importance of the game (I),
  - the strength of the opposing team (S) and the average of confederation strengths (C) of participating teams.
- Next, yearly averages of the number of points gained are computed.
- These four averages are added up with weights 1, 0.5 0.3 and 0.2 with more recent years being assigned higher weight.

Some remarks

We turn back to the FIFA ranking. At first, we note two things:

- The FIFA ranking methodology does not account for the home team advantage which is an important factor of the game - we should always invite the teams to play at our ground.
- Because of strategic reasons, a team should never play a match against an opponent for which the possible number of points to gain (in case of a win) is lower than the average points gained in current year.

Example

If Italy had avoided matches against Haiti and San Marino by the end of 2013 they presumably would have gained seeded rank in the FIFA World Cup 2014 draws, which was partially based on a team's FIFA rank. In this case they would have avoided being paired with Costa Rica, England and Uruguay, considered as a group of death. In the end, Italy was eliminated in the group stage of the tournament.

Scheduling of friendlies - setup

We turn to the problem of choosing opponents for friendly games.

Target

Our target is to advance at a given rank with the highest probability in the future (in particular: the next) FIFA ranking release.

Let us introduce some notation:

- Let A be the the set of ranked teams.
- Let $P_{(n)}$ denote a random variable of n-th order statistic (sorted in decreasing order, e.g., $P_{(1)}$ is the distribution of ranking points by the team ranked first).
- $P_{(n)}$ is dependent on a set of scheduled matches $\mathcal{S}$.

Scheduling of friendlies - setup

Let X_a denote the total number of a team's ranking points that it can gain when playing against an opponent $a \in A$. We can formulate our goal in term of the following decision problem

$$d = \underset{a}{\operatorname{argmax}}\, \mathbb{P}(X_a \geq P_{(n)} | \mathcal{S}).$$

We may also want to schedule several games in advance. We have that $\mathbf{a} = (a_1, a_2, \ldots, a_l) \in A^l$ is the sequence of teams to play against. Our objective function becomes

$$d = \underset{\mathbf{a}}{\operatorname{argmax}}\, \mathbb{P}(X_{\mathbf{a}} \geq P_{(n)} | \mathcal{S}).$$

We find the optimal decision by brute-force search.

Computing distribution of $P_{(n)}$ - Monte Carlo simulations

The first problem we encounter is to compute distribution of *n*-th order statistic $P_{(n)}$ given a set of scheduled games $\mathcal{S}$. If distribution of $P_{(n)}$ is influenced by the outcomes of 20 games, exact computation of it requires analysis of $3^{20} \approx 3.5$ billion game outcomes. Therefore in our experiments we decide to simulate match results to approximate distribution of $P_{(n)}$.

To restrict the number of necessary computations we determine two constants - lower and upper bound on $P_{(n)}$

$$K_{\min} \leq P_{(n)} \leq K_{\max}$$

and ignore all the games played by the teams with their (random) ratings points lying outside the interval $[K_{\min}, K_{\max}]$.

Prediction of football match outcomes

A basic model for football match prediction: we assume that the random number of goals scored by the teams follow Poisson distribution with mean λ and μ

$$\mathbb{P}(X_i = x, Y_j = y) = \frac{\lambda^x}{x!} exp(-\lambda) \cdot \frac{\mu^y}{y!} exp(-\mu),$$

where log-linear model for the goal scoring rates is assumed depending on offensive and defensive capabilities of the teams (corrected for the home team advantage factor h)

$$\log(\lambda) = c + h + o_i - d_j$$

$$\log(\mu) = c + o_j - d_i.$$

Other possible approaches: ordered logistic regression, machine learning models or just use betting odds.

Computing optimal choice of opponents

For solving the decision problem we used brute-force search.

- The search space $|A^l|$ can be large: with 200 teams as possible opponents and 5 games to schedule there are as many as $\binom{200}{5} > 2.5$ billion of possible choices.
- However, we may capitalize on particular structure of the set of possible actions A and dismiss dominated choices in advance:

$$a \succ \tilde{a} \iff N(W_a) \geq N(W_{\tilde{a}}) \ \wedge \ \mathbb{P}(W_a) \geq \mathbb{P}(W_{\tilde{a}}) \ \wedge \ \mathbb{P}(L_a) \leq \mathbb{P}(L_{\tilde{a}}),$$

  where $N(W_a)$, $\mathbb{P}(W_a)$ and $\mathbb{P}(L_a)$ denote the maximal number of points that we can obtain when playing against of team a, the probability of winning and losing against this team.
- After this restriction we are left with about 30 non-dominated choices.

Example - Poland

Let us focus on Polish national team by the end of 2013. Our goal is to climb up the December FIFA ranking release with the highest probability.

Table : Friendly games by Poland in November 2013 - January 2014.

Date	Match			Result
15/11/2013	Poland	–	Slovakia	0:2
19/11/2013	Poland	–	Ireland	0:0
18/01/2014	Poland	–	Norway	3:0
20/01/2014	Moldova	–	Poland	0:1

Question: Can we optimize the schedule for Poland?

Example - Poland

Table : One step optimal decision.

Rank *n*	**Switzerland**	**Iceland**	**Dominican Rep.**
64	0.0189	1e-04	0
65	**0.0864**	0.0038	0
66	0.1934	0.0488	0
67	0.2735	0.1898	0.0086
68	0.303	0.3926	0.0755
69	0.3072	**0.5355**	0.2774
70	0.3073	0.5871	0.5807
71	0.3073	0.595	**0.8106**
72	0.308	0.5954	0.8967

Example - Poland

Table : Two step optimal strategy of scheduling games.

Rank n	Iceland Romania	Algeria Iceland	Algeria Cuba	Cuba Dominican Rep.
62	0.1685	0.0148	0	0
63	**0.2169**	0.0697	0.0023	0
64	0.2335	0.1679	0.0292	0.0001
65	0.236	**0.2632**	0.1335	0.0049
66	0.2454	0.3225	0.2988	0.0634
67	0.2918	0.383	**0.4232**	0.2468
68	0.3778	0.4796	0.4735	0.5106
69	0.4545	0.5702	0.4984	**0.6973**
70	0.504	0.6196	0.5454	0.7732
71	0.5749	0.673	0.6312	0.8184

Example - Poland

5	Belgium	1175	44	Japan	634
6	Uruguay	1164	44	Wales	634
7	Switzerland	1138	46	Iceland	633
8	Netherlands	1136	47	Norway	632
8	Italy	1136	47	Tunisia	632

Figure : Places 4-8 and 44-47 in October 2013 FIFA ranking release.

Such strategies may be employed by, e.g., hosts of major competitions, which play a majority of friendly games prior to the tournament.

Coalitional improvement of ranks (1)

Two (or more) teams may agree to play many mutual games and, by winning half of the encounters, they might increase the number of accumulated points.

Let us assume that two low-ranked teams (lower than 150th place) employ such a strategy. According to the FIFA ranking methodology, each team obtains obtains $3 \cdot 50$ points and 0 points in case of a win and a loss respectively.

- In the limiting case each team will obtain an average of $\frac{1}{2}(3 \cdot 50 + 0) = 75$ points a year. Hence they can accumulate around $75 \cdot (1 + 0.5 + 0.3 + 0.2) = 150$ points over four years.
- **Question:** can they gain more? Denoting the rank of team as n and the associated number of points as $P_{(n)}$ in the long run limiting case it would hold that

$$\frac{3 \cdot \max(200 - n, 50) + 0}{2} \cdot (1 + 0.5 + 0.3 + 0.2) \approx P_{(n)}.$$

Coalitional improvement of ranks (2)

- Based on historical data we may find a value of n for which the above equation holds. It is most closely met (in mean absolute deviation sense) by the values of $n \approx 150$.
- We conclude that forming such coalitions would be profitable only for low ranked teams which would place them around 150th position. Higher ranked teams do not have incentive to implement such (dishonest) strategies.

Summary

- FIFA methodology often ranks the teams disproportionally to their strength.
- This creates opportunities for the teams to challenge highly rated opponents that they are likely to beat.
- Low ranked teams may artificially create points and advance in the ranking table, though the method is rather robust to such manipulation.
- An open question remains whether we can solve the decision problem more effectively than by brute force search.
- More real life scenario can be considered with, e.g., random teams' ranks, which were assumed to be known based on most recent FIFA ranking release (i.e., we may extend the horizon of scheduling).

EloRatings.net ranking method (1)

Unofficial football ranking is maintained on the website EloRatings.net.
It is based on the famous Elo rating system for chess players.
The model consists of following key components:

- **Main idea:** Adjust teams' ratings according to the discrepancy of the model expectations and the actual result of a match.
- Let two teams i and j be given and rated with r_i and r_j points.
- Prior to the game the expected outcome of the match is computed as

$$p_{ij} = \frac{1}{1 + e^{-a(r_i - r_j + h)}},$$

where $h > 0$ is a parameter accounting for *the home team advantage* ($h = 0$ if the game is played on neutral ground) and $a > 0$ is an appropriate scaling constant.

EloRatings.net ranking method (2)

- For a single game the teams are awarded the points according to the formula

$$u_i = k(goals, importance) \cdot (a_i - p_{ij}),$$

  where:
  - a_i stands for actual result of the game $a_i = 1, a_i = 0.5, a_i = 0$ in case of team i win, draw and loss accordingly,
  - k is a function of the number of goals scored and the importance of the game (friendly game, WC game, etc.).
- For team j we have $u_j = -u_i$.
- Team's i ratings are updated as follows

$$r_i' = r_i + u_i,$$

- FIFA ranking methodology for **women football** also is a modification of **Elo rating system**.

Questions?

Thank you for your attention slide

Thank you!

For Further Reading I

McHale, I. and Davies, S.
Statistical Analysis of the Effectiveness of the FIFA World Rankings.
Statistical Thinking in Sports, 2007.

Lasek, J., Szlavik, Z. and Bhulai, S.
The Predictive Power of Ranking Systems in Association Football
International Journal of Applied Pattern Recognition, Vol. 1, Is. 1, 2013.

FIFA.com
FIFA/Coca-Cola World Ranking Procedure
http://www.fifa.com/mm/document/fifafacts/ra-wr/52/00/97/fs-59010ewrpoints.pdf,
2014.

Eloratings.net
The World Football Elo Rating System
http://www.eloratings.net/system.html.

For Further Reading II

Wang, C. and Vandebroek, M.
A model based ranking system for soccer teams
Technical report, KU Leuven - Faculty of Economics and Business, 2013.