G1 Ranking Sports Teams

G1.1 Massey’s Original Ranking Idea

The following is the idea for ranking put forward in Massey’s undergraduate thesis. (He has more advanced methods now.) By ranking we mean a numerical value; from this one can produce ordinal values, if desired.

The method is based on the idea: that in a perfect world the teams have numerical ratings such that, in every game, the resultant net score is the difference between the ratings of the teams. That is, there are ratings $r_i$ such that the margin $g_{ij}$ of team $i$ over team $j$ is given by

$$g_{ij} = r_i - r_j$$

So the task is, given the actual margins $g_{ij}$, to determine the $r_i$. There are of course many more equations than variables; so this system of linear equations almost surely has no solution. Instead, the goal is to choose the vector $r$ to minimize the errors $g_{ij} - (r_i - r_j)$. The standard approach is to minimize the sum of the squares of the errors.

When the dust settles, one needs to solve the system of equations where for each team the net difference of the ranks equals the net score difference. This is a matrix equation

$$Sr = N$$

where $S$ is an $m \times m$ matrix with each diagonal entry the number of games that team played and off the diagonal the value is $-1$ if the corresponding teams played, and 0 otherwise. And vector $N$ is the net score for each team: points-for minus points-against.

Note that $S$ is never invertible. (Why?) So the solution is to discard one row of $S$ (since it is redundant) and add a row that normalizes the rankings. For example, add the equation that the rankings sum to 0.
G1.2 Ranking: The Elo Method

The Elo rating system is used in chess and elsewhere. It was designed by Árpád Élő. The idea here is that the teams have a rating such that the difference in the ratings exactly determines not the margin but the probability of one team winning. Further a player’s ability is assumed to change slowly over time.

Specifically, ratings are continually updated. A player is rewarded if their result is better than expected and their opponent penalized by the same amount. The question is what the adjustment curve should look like.

In chess one can have draws. But that is handled by replacing the probability of team $i$ beating team $j$ with the expected value of $g_{ij}$, where $g_{ij}$ is a random variable that is either 1, 0, or $\frac{1}{2}$, depending on whether $i$ wins, $j$ wins, or the game is a draw.

A basic assumption might be that a chess player’s performance has a normal distribution. A related idea is used; namely the S-shaped logistic curve. Specifically, chess ratings are based on the assumption that the expected score for a player with rating $r_i$ playing against a player with rating $r_j$ is:

$$E_{ij} = \frac{1}{1 + 10^{(r_i - r_j)/400}}$$

For example, if the ratings are equal, the expected score is $1/2$ (as it should be); if the ratings are 400 apart, then the weaker player has an expected score of $1/11$ (that is, they should win 1 point in 11 games—for example, 2 draws). (Aside: this formula appears in the movie “The Social Network” but incorrectly written).

The players are then rewarded or penalized by being moved towards the rating that reflects their actual performance. Specifically, if $E_{ij}$ is the expected result, and $g_{ij}$ their actual result, then player $i$’s rating is updated by

$$r'_i = r_i + K \ (g_i - E_i)$$

where $K$ is some weighting factor.

There are many practical issues. What if a player has no rating? How do we start? Well, one can work backwards based on their first tournament. Assuming the opponents of player $i$ do have ratings, then one can compute that value of $r_i$ such that the total $g_{ij}$ equals the total $E_{ij}$. A minor hiccup is that if Wayne loses every game at his first tournament, he would start with a rating of $-\infty$. So, the US chess federation gives such a player an initial rating of 400 below the lowest-rated player they played.
On can also incorporate actual game scores in say football; for example, one might let \( g_{ij} \) be the proportion of points that team \( i \) scored.

### G1.3 Markov Methods and PageRank

Here you create a network out of your elements. The standard is websites, but could be sports teams again. The idea is that, at first approximation, your importance is determined by how many sites point to you. But of course, the importance of the websites pointing to you is also important. In a sporting context, you could point to the teams that you lost to as sharing your vote for the best.

Then use what is called the Markov method. The idea is that one takes a random walk through the network. The rating of a site is the steady-state average time spent at that vertex.

Take an \( n \times n \) matrix where entry \( s_{ij} \) is 1 if team \( i \) lost to team \( j \) and 0 otherwise. We immediately have a problem if a team is undefeated, since that is an absorbing vertex in the random walk. So one fix is to make that row all 1’s. In any event, normalize each row so that it sums to 1; call the result \( S \).

We want the stationary distribution. That is, \( S^t r = r \). Software can calculate that for us. In this case, there is a simple numerical idea called the power method. Start with all entries in \( r \) being \( 1/n \). Repeatedly calculate \( S^t r \) and normalize each time, until converges.

This is equivalent to the idea that the weight of a team’s vote should be based on how good they are. There are mathematical theorems about convergence et cetera. Computational issues. For football, one can scale the links by the margin of victory. PageRank and adjustments are discussed at website whydomath.org

### References


Wikipedia