Network based ELO Ranking

Owen Lu

Electroimpact Inc.

owenl@electroimpact.com

*Abstract*—Calculation of relative ranking scores in many systems use some variation of Elo’s original method. The main principle being that each player samples a distribution for a score and the one with the higher score wins. Adjustment of Elo scores can be done with very simple calculations as it is often done in chess. However, the methods are unlikely to generate accurate global scores when people are weakly connected in a network. This network based method looks to use a network method to adjust scores periodically as new games are added.

# Introduction

To demonstrate the problem with the original Elo’s method suppose there are 4 players numbered 1 to 4. Suppose the pair of players 1 and 2 are of equal skill level, and pair 3 and 4 are also of equal skill level. Using the traditional rating of 1000 as the starting score, this would mean that all players have a rating of 1000. However, suppose that players 1 and 3 compete and player 1 wins consistently, generating a score higher than 1000. We see that players 2 and 4 both have Elo scores of 1000 assuming they have not played with others, but it is likely that player 2 is a better player than player 4. Ideally, player 2’s score would be adjusted to be equal to player 1’s score since they are approximately equal in skill rating, even with no further games being played by player 2.

Central to the idea of global skill rating is that paths between players in a network affect one’s skill rating even if no direct games have been played. This information would allow one to better estimate the probability of winning between two players based on their skill ratings before any games have been played between them if there was a graph connection between them. In the end however, it seems a good skill rating whether one number, or a tuple of numbers would allow one to easily and accurately estimate the probability of players winning.

# Equation Formulation

Using a logistic curve, which is traditionally used in Elo formulations, the probability of the first player winning can be estimated. The equations following use the standard Elo formulation.

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

Probability is the complement of

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

Where denotes the Elo score of the ith player.

Suppose that we have an estimate of from actual games played. Then the difference in Elo score is solved as such.

If we consider the network to be only comprised of two players, we have the additional constraint that the average Elo must be 1000. Another way to say this is that the number of points in the system is constant for a set number of players.

So the system generated is a square matrix, and in this case .

Which can be solved explicitly for and . For example, suppose .

In an player system, the matrix and can be formulated below.

is an matrix and is an . Since and matrices are formulated using the probability of player 1 winning, this system represents the Elo scores in Player 1’s own network.

In such a network, the total Elo is conserved and although players that have played Player 1 have not played each other (according to the network) they still have relative scores. We can think of an implicit inequality occurring. For example, if Player 2 beats Player 1 more often than Player 3 beats Player 1, we would assume that Player 2 has a higher Elo than Player 3 even if they have never played. We can simulate such a network as another example.

Suppose that:

Player 1 beats Player 2 with 60% probability

Player 1 beats Player 3 with 70% probability

Generate and

Solve for

As expected . Then using the and we can estimate the probability of beating .

Of course, because Player 2 and 3 have never played in this network representation, so this is just a naïve estimate.

In order to extend this idea to a global network consider the following problem first.

If we consider all the games that Player 1 has played as we did before, we can generate a relative ranking within Player 1’s own network which is consistent. However, Player 2 could have played with Player 1 and others as well which means Player 2 has his/her own network. This means that if we solve the equation for given player 1’s network, and then solve in player 2’s network, we will get different Elo scores in the most likely case. This gives rise to the possibility of contradictions.

For example, one network might say , another might say . This gives rise to the notion that everyone’s **Elo is referenced to a network** or that Elo is much like a vector position, there is no absolute value and it depends on one’s frame.

A way to reconcile the difference in Elo is to calculate each player’s Elo based on **first order** connections and average them. Perhaps a more sophisticated way to do this would be to weight dependent on binomial probability distribution variance, but this will be left out for now.

Let denote the Elo of the ith player as calculated by the jth player’s network. The Elo is simply an equally weighted mean average.

Where:

is the Elo of Player

is the number of players in Player ’s network

This can also be thought of as players in a network voting on the Elo of an individual based on their own experience. At the end of computing the network, the Elo’s should be normalized to an average of 1000.

# Variance and Wilson’s Score

It is obvious that for small number of games, the probability a player winning is unlikely to be accurately estimated by wins divided by the total number of games played. The problem is exacerbated when most people in the network will play certain players a lot, but most players only a few times during events. This brings up the concept of variance, and confidence intervals in binomial distributions.

Luckily Wilson score interval gives a quick way to estimate the confidence interval as a function of games observed , z-score and probability observed  .

Clearly, this equation is symmetric, so we can simply use the mean to estimate the probability of winning. For example, in the extreme case where 2 players play 1 rack, one player will win with 100% observed probability.

Therefore in this scenario:

Then it is up to the user to decide the score. Suppose we use which basically means the probability is 68% likely to be bounded between the interval calculated by the Wilson’s score. However, for simplicity we take the mean.

An interesting point of using Wilson’s score to estimate the confidence interval is that this can be translated into a Elo bound as well, further investigation is required to make this of practical use.

# Conclusion

This is the first formulation of a global ranking system that does not try to penalize abuse of the system. It also does not take into account how recently the games have been played as there is no time factor in the calculation. More study is needed to make the system robust to certain factors.