Prediction Model Analysis for Ping Pong

Jasper Chen

November 2014

Question

Game prediction is popular. If you look on the web for game prediction, you'll find online game predictors, leaderboards with user rankings and speculation about the next major game. For predicting games, most use the ELO rating system, where the probability of a player's predicted outcome is determined by the ranking of him and his opponent. The ELO rating system has been used to predict chess, football, hockey and soccer.

However, there is very not much information about prediction models for ping pong games. The purpose of this project is to determine if the ELO model makes sense for predicting ping pong games.

Methodology

In order to begin analyzing predictive models, I collected and formatted the ping pong results of recreational ping pong players at Olin College and Babson College. The dataset contains the results of 151 ping pong games between 18 players. A player can only win a game if they are ahead of the opponent by 2 points and have at least 11 points.

ELO Model

I wrote a script that parses the dataset for a player's outcome and assigns a default ELO ranking or updates the ELO ranking. For every game played, a player's ranking increases or decreases depending on the player's expected outcome, his actual outcome and a scaling factor, which adjusted based on the player's skill and experience. In this case, the scaling factor is set to 30, because by the ELO's methodology, there aren't enough games to result in a predominantly large difference of skill or experience between players.

Poisson Model

In hopes to compare the ELO Model, I've used a Poisson Model. The Poisson Model assumes that events can occur at any time with equal probability. The rate at which the events occur, μ and ς is determined by taking the mean and standard deviation of the average scores amongst the 18 players, respectively.

Given a set of previous matches between a player-opponent pair, the script updates the prior hypothesis for each player, ν, and returns their corresponding posterior hypotheses. These posterior hypotheses is used to predict a player's outcome against the opponent.

Result

The 18 ELO rankings have a range between 1069 and 1308 with $\mu = 1192$ and $\varsigma = 61.5$. The μ and σ of the average of each player's ping pong scores were 9.72 and 1.61, respectively.

A test was conducted to evaluate the accuracy of the ELO prediction model and the Poisson prediction model. To increase the amount of sample size in our test, I choose a pair who have played the most amount. For instance, Person A had played 33 games of ping pong with Person B.

After Person A's 20th game with Person B, Person A has an ELO Score of 1090 while Person B has an ELO Score of 1303. It should be also noted that Person A has won 6 games out of 20 games against Person B, equivalent to a 30% win rate.

$E_A = \frac{1}{1+10^((R_B - R_A)/400)}$

$E_A = \frac{1}{1+10^((1303-1090)/400)}$ = 22.69% chance that Player A wins.

Given the 20 datapoints, the Poisson Process estimates that Player A would win 35.85% of the remaining games.

However, in actuality, Player A had won 2 of the 13 remaining games, so Player A won 15.38% of the remaining games.

Interpretation

At the project's current state, it would be naive to suggest a predictive model over another. There clearly needs to be more data to test each model's predictive accuracy. For instance, in the case of Player A's outcome prediction against Player B, each win or loss would result in a difference of $1/13 = 7.7\%$ actual rate. A difference of 1% and as a result, 100 games per player would provide a more telling picture.

In addition, players need to play with their opponents more equally. For example, Player C and Player D have had 23 games between each other and very little games with other opponents. Player C and D had a final ELO score of 1308 and 1069, respectively. In one game, when Player C had played Player B (with an ELO score of 1303), Player C had lost 0-11 (and shutouts are very rare). An unequal distribution of games between players results in skewed ELO scores relative to other players. As a result, the ELO scores become more unreliable as a predictive measure.

However, with more data, suggesting a predictive model over another does seem possible.