A blitz talk on ELO rating

Naoki Hiratani

Tea Talk

2018-11-22

World chess championship is happening here in London, right NOW.

- Every one, two, or three year(s). This year in London.
- 12 matches between the champion (Magnus Carlsen) vs. the challenger (Fabio Caruana)
- 9 draws so far…
- How strong are they? (compared to us, or AlphaZero)

ELO rating

- Score of relative strength, developed by Arpad Elo around 1960.
- Since 2018, employed for the calculation of FIFA ranking too.
- The idea: assign rate r_i for all the players in a way that for any player A and B,

$$\Pr[A\ beats\ B] \approx \frac{1}{1+10\text{^}[-(rA-rB)/400]}.$$

- Representative rate:

  Beginner: 600-800

  Expert: ~2000

  World champion (human): ~2800

  World champion (computer): ~3500

  AlphaZero (AI): 3750

How to calculate ELO rating?

Given data *D*, the Bayesian estimation is

$$\log p(\boldsymbol{r}|\, D) = \sum_{a,b}\left[d_{a,b}\log \rho_{a,b} + d_{b,a}\log \rho_{b,a}\right]$$

where $\rho_{a,b} \equiv \frac{1}{1+10\text{^}[-(ra-rb)/400]}$ and $d_{a,b} \equiv \#[a\ beat\ b]$

However,

- The rate should be updated online (after each game).
- The update should be local (to the two players fought the game).
- The update rule should be easy to calculate (closed-form solution).

The approximated update rule is

$$r_a^{t+1} = r_a^t + K_a\left[s_{a,b}^t - \rho_{a,b}\right] \text{ where } s_{a,b}^t \equiv [\text{if } a \text{ beat } b \text{ at game } t]$$

$$K_a = \begin{cases} 40\ (\text{if } a \text{ is a new player}) \\ 20\ (\text{if } a \text{ is a weak player}) \\ 10\ (\text{if } a \text{ is old and strong}) \end{cases}$$

Is ELO rate inflating or deflating?

The total/average rates are not conserved.

- Influx/outflux of players
- The increments for two players fought a game are not balanced when $K_a \neq K_b$

Arguments for inflation:

- Empirical observation (only 1 player rate 2700+ in 1979, but 46 players in 2018)
- Rates of new players are typically underestimated. ($K_{new} > K_{old}$)

Arguments for deflation:

- Most players start at a low rate, and retire at a high rate.
- Larger influx than outflux.