

Project Proposal: Chess Predictor

Inwon Kang

Dataset

Chess Game Dataset:

This dataset contains almost 20,000 games collected from a selection of users on the site Lichess.org.

The fields in this dataset include:

- Game start/end time
- Player ids
- Player skill ratings
- Winner/Loser of the game

Problem

- Can we predict upsets (lower rating victory over higher rating)
- Regular victories

Ideas:

For each user, we could keep a set of users who beat them and users who lost to them.

Then we can compare the common neighbors in each set for each match.

Using the two sets from both players in a match, we can calculate values such as Adamic- Adar Index or Jaccard Index.

For example, if player A lost to B, and player B lost to C, even if C has a higher rating, there is a chance that C may beat A.

Testing

Similar to homework 2, I plan to use the first 25% of the data to build a predictor and run the rest of the games to check my accuracy.

In the rest of the data that I do not use to build predictors, I will identify 'upset' games.

The goal is to be able to predict over 50% of the upset games.

Reference:

I found this dataset on kaggle while searching for game-related datasets.

<https://www.kaggle.com/datasnaek/chess/data>