Proposal for the Analysis of Chess Games and Chess Openings

Team: Sahib Bajwa

Introduction

The Shannon Number tells us that the conservative lower bound of the game-tree complexity of chess is 10^{120}. To give this number some perspective, there are an estimated 10^{80} atoms in the universe. It is very difficult to determine the best way to play the game, even given the use of AI (Stockfish, AlphaZero, Leela Chess Zero, etc.). Having a strong opening is one of the most important moments of a chess game as it sets the pace and structure for the rest of the game. Given that humans do not have the computational ability of a strong AI, finding out what the best opening is for a player would help them increase their chances of winning a large amount.

Related Work

- Chess AI:

  - There are Chess AI that are built around specific types of playstyles that prefer specific types of game states. Thus, they start a specific way in order to push the game towards that game state.

- Chess Theory:

  - There has been a lot of theory crafting about how to play in specific game states. One great example of this is the book *FCO: Fundamental Chess Openings* by Paul van der Sterren.

  - There is also lots of chess theory that is done on the individual level by high level chess players before specific games. This is usually not seen in/by lower level players.

Proposed Work

I plan to determine the best openings for players to use dependent on multiple factors. Some of these factors include:

Player(s) Rating
Player Color
Opponent
Opening Move
Game Type
Many more factors

This will be done mostly by analyzing previously played chess games and then taking in the stated variables. There are many factors going into this analysis, so the output for the best opening will vary greatly.

Evaluation

- There will be a few ways to measure success in the opening evaluation. This will include:

  - Human winning percentage when given an opening.

  - Game state variation when given an opening.

    - Ease of transitioning from opening to later game state.

  - Overall player thought when given an opening they should practice/implement.

  - AI winning percentages given different ratings and specific openings.

Milestones

- No concrete dates, but there is a project timeline:
  - Determine if datasets should be merged.
  - Clean data
    - Determine what data should be used.
  - Specified analysis of openings for all players.
    - Specified analysis of openings by player rating.
    - Specified analysis of openings by opponent opening.
  - Setting games with openings vs each other to determine viability of openings.
  - Maybe more if thought of later.

Checkpoint Slides

- No concrete dates, but there is a project timeline:

  - Determine if datasets should be merged. --- Done, current data being used is from the Lichess data set.

  - Clean data --- Semi-done

    - Determine what data should be used. --- Semi-done, data called by specified cells.

  - Specified analysis of openings for all players. --- Done, statistics of frequency and graph shown.

    - Specified analysis of openings by player rating. --- Not done

    - Specified analysis of openings by opponent opening. --- Not done

  - Setting games with openings vs each other to determine viability of openings. --- Semi-done, analyzing player vs analyzing player developed.