Checkpoint for the Analysis of Chess Games and Chess Openings

Sahib Bajwa
University of Colorado Boulder
Boulder Colorado United States
sahib.bajwa@colorado.edu

Introduction

Chess is one of the most complicated games from a theoretical standpoint. The Shannon Number tells us that the conservative lower bound of the game-tree complexity of chess is 10^{120}.[1] To give this number some perspective, there are an estimated 10^{80} atoms in the universe[2]. 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.

The Oxford Companion to Chess states that there are 1327 named openings in chess[3], and each opening can have multiple variants. One example of this is the popular "Sicilian Defense". Based on a what situation the game is in, the "Sicilian Defense" can be modified into another opening. This includes another popular opening, the "Sicilian Defense: Bowdler Attack". Considering there are so many openings that exist and should be used in niche situations, helping players know and understand what opening to use would help them succeed in winning games.

I also believe that it would help lower rated players stay interested in the game as it would give them a strong starting point to build off. FM Steve Giddins said in 2008 that, "the average paler only needs to know a limited amount about the openings he plays. Providing he understands the main aims of the opening, a few typical plans and a handful of basic variations, that is enough." If a newer player learns the openings that are the best suited for them, they will have a much greater chance at becoming better at the game. Like I said before, I also think that once a newer or lower rated player sees the improvements in their play due to playing an opening suited for them, they will most likely choose to continue playing the game due to seeing improvement in their play.

Going along with this, it will not work to simply give a player an opening that is considered strong and tell them to use and learn it. This is because players at different ratings are better or worse at different openings. A complex opening that is highly successful may be a good option for a Grandmaster, but a newer/lower rated player may not understand the depth to the opening. Thus, at different ratings and skill levels, players should prefer and learn different openings.

Thus, I would like to take a look at and analyze large chess games data sets in order to determine how and when to use certain openings. Due to the reasons stated in previous paragraphs, I believe this project to be important and useful. There are a lot of intricacies in chess, and helping players determine a good opening will allow them to be better prepared for the game as well as make the game overall much easier to win.

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. Some of the most popular include: Stockfish, AlphaZero, Leela Chess Zero, and Komodo. A large list of chess engines and their ratings can be found at: https://ccrl.chessdom.com/ccrl/4040/

• 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. Due to the volume of books on the subject, it is easy to find resources on specific openings or theory about chess game states.

    • 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. One of the most recent examples of this that comes to mind is when Grandmaster, and considered best player in the world, Magnus Carlson was playing other Grandmaster Hikaru Nakamura at the Magnus Carlson Invitational. GM Hikaru Nakamura was able to theorize and practice an opening to the point where he played the game to an around 99% accuracy compared to what Stockfish AI believed would be the best moves for the game. Chess Theory is very researched and can be the deciding factor to who wins many high-level chess games.

[1] https://en.wikipedia.org/wiki/Shannon_number
[2] https://www.universetoday.com/36302/atoms-in-the-universe/#:~:text=At%20this%20level%2C%20it%20is,hundred%20thousand%20quadrillion%20vigintillion%20atoms.
[3] https://en.wikipedia.org/wiki/Chess_opening#:~:text=The%20Oxford%20Companion%20to%20Chess,the%20middlegame%20and%20the%20endgame.

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(s) Starting Color
- Opponent Opening Move
- Game Type
- Other Factors that may influence the player and their ability to succeed given an opening.

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.

The process for this project will include the following steps:

- Merging necessary datasets.
- Cleaning and removing non-useful data due to the sheer size and complexity of the datasets.
- Analyzing the data and creating subgroups of information and creating subgroups of openings based on rating.
- Creating a concrete way to recommend an opening to a player based on the above factors.
- Setting games with real players and/or computer AIs using the determined opening.

Evaluation

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

- A human's winning percentage when given an opening ~~versus~~ **based on** their previous winning percentages.
- The evolving game state variation when given an opening.
    - The ease of transitioning to said game state.
- ~~Overall player thoughts when given an opening they should practice and/or implement~~.
- AI winning percentages given different ratings and specific openings and ratings.

~~This will require a few human participants. Due to human error, I will also be using AIs that are consistent in their play in order to see the effectiveness of the openings at a specified rating/skill level.~~

Human participants have become difficult to find due to Covid-19, thus, all evaluation will take place under AI and manipulating situations to make them as close to human trial.

Milestones

The milestones will correspond to the Proposed work quite similarly. The currently laid out milestones are:

- Determining what datasets should be merged. - DONE
- Merging necessary datasets. - DONE
- Cleaning and removing non-useful data due to the sheer size and complexity of the datasets. – SEMI-DONE
- Analyzing the data and creating subgroups of information and creating subgroups of openings based on rating.
- Creating a concrete way to recommend an opening to a player based on the above factors.
- Setting games with real players and/or computer AIs using the determined opening.

I believe that some milestones will be added later, but that will only be known when they come to pass. The milestones will most likely have subtasks that I cannot think of currently.