

Analysis of Chess Games

Overview

Clearly, Chess is a very complex game with a rich history of being utilized as a proving grounds for the latest and greatest in artificial intelligence. For our final project, we have decided to look into a large set of chess games in order to predict their outcomes based on a variety of game features, including opening, time in game and a given set of moves.

We are aiming to utilize several machine learning tactics in order to develop a useful and accurate classification system for chess games, and will hopefully be able to look at a board position and determine the winner of the game. We're hoping that this will be a fun and interesting project where we will be able to utilize a wide variety of data science methods.

Datasets

Fortunately, there are a wide variety of datasets focused on chess games and their outcomes, with varying features for each.

Some of the datasets we are looking to use are:

- LiChess 20k game archive with opening names and moves:
<https://www.kaggle.com/datasnaek/chess#games.csv>
- 35 million chess games annotated: https://www.kaggle.com/milesh1/35-million-chess-games#all_with_filtered_annotations.zip
- A number of chess datasets exist on the UCI machine learning repository:
 - [https://archive.ics.uci.edu/ml/datasets/Chess+\(Domain+Theories\)](https://archive.ics.uci.edu/ml/datasets/Chess+(Domain+Theories))
 - [https://archive.ics.uci.edu/ml/datasets/Chess+\(King-Rook+vs.+King-Knight\)](https://archive.ics.uci.edu/ml/datasets/Chess+(King-Rook+vs.+King-Knight))
 - <https://archive.ics.uci.edu/ml/datasets/Chess+%28King-Rook+vs.+King%29>
- LiChess also provides monthly game records: <https://database.lichess.org/>

Ideally, we will be able to combine these datasets using some processing/cleaning tactics

Possible Feature Generation

Use of Pieces

Use the move set to identify the frequency with which each player uses different pieces

Methods of Analysis

Semantic Analysis

The games are annotated with commentary. This can be analyzed for sentiment and summed up to get an overall feel for the moves of each player.

Classification of Win-Loss

Use opening, moves, and ratings to predict win or loss

Prediction of Moves to Win

Use opening, moves, and ratings to predict the number of moves in the game

Prediction of Win Condition

Predict the method by which the game ends.