Review on Statistical Analysis of a Chess Player using Data Science Pipeline

Yogen Ghodke[1], Mangesh Kauthale[2], Ajinkya Shewale[3],
Jaya Mane[4], Shobha Raskar[5]

[1] Student, BE Computer Engineering, MES College of Engineering, Pune
[2] Student, BE Computer Engineering, MES College of Engineering, Pune
[3] Student, BE Computer Engineering, MES College of Engineering, Pune
[4] Assistant Professor, BE Computer Engineering, MES College of Engineering, Pune
[5] Assistant Professor, BE Computer Engineering, MES College of Engineering, Pune

---***---

Abstract - Chess is a game which has been played by millions of people worldwide, since centuries. Although knowing the rules and legal moves of chess is enough knowledge to play it, a move strength feedback at every move played is the best way to climb the rating ladder quickly. Although there are third party tools available which help players to explore common openings and provide an overall score of the previous games played by them, there is a need of a move-by-move in-depth statistical analysis tool which will help chess players to get better at the board game. This is true at all levels, from amateur chess players to grandmasters. Our proposition is a system which will provide the chess players a live analysis of possible candidate moves, data graphs, game trees through the concepts of Data Science in a practice game.

Key Words: *Chess, Data Science, Analytics, Statistics, Decision Trees, Data Visualization*

1. INTRODUCTION

Chess is one of the most popular two-player strategy board games in the world. Players usually clean up their chess skills, which include strategy and tactics, by reading expert reviews and tutorial books. Given that computers have recently surpassed humans in playing chess, modern artificial intelligence (AI) has emerged as a helpful tool for analyzing the interaction of pieces and the evolution of a game. AI's thorough and deep searching enables the depiction of chess positions after more than 10 moves. Therefore, it is capable of analyzing long-term positional advantages of a game.

1.1 Current System Drawbacks

The current tools available in the market for Chess games analysis have huge limitations. The chess engines only provide analysis on individual games and not an aggregate analysis on multiple games played by the player in bulk. All the current chess tools assist users in analyzing games by providing an overall score of the move played. It gives a positive score if the move played is a good move, gives a negative score if the move is a bad move. As for the further progressing of the game, they simply describe a sequence of chess moves using algebraic chess notation. Hence, users have to figure out by themselves and sequentially investigate the game and try to figure out answers to questions such as "Why does the player give up the game?", "Is there a way to escape from king hunt?", and "How does the player turn defeat into victory?".[1]

1.2 Proposed System

We propose a system software where a player can feed all their previously played games into the system and get insights on what moves played were bad moves from a positional standpoint and give a reason why they were bad. Also, we want to provide the user an option to see for themselves, in an interactive setting what openings are their favorite, what typical blunders do they commit in those openings, while playing as black what defenses they play are the weakest and how they could improve on them.

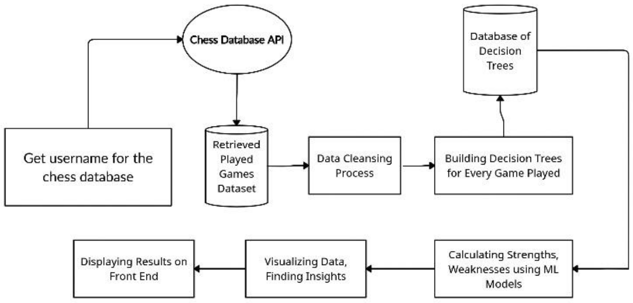
Chess Database API
Database of Decision Trees
Get username for the chess database
Retrieved Played Games Dataset
Data Cleansing Process
Building Decision Trees for Every Game Played
Displaying Results on Front End
Visualizing Data, Finding Insights
Calculating Strengths, Weaknesses using ML Models

Fig -1: Block Diagram of the Proposed System

2. PROPOSED SYSTEM FEATURES

The following are the functionalities that we look forward to see in the proposed system:

- Best Alternatives for Candidate Moves
- Opening Move Explorer
- Interactive Graphs for Wins / Losses / Draws
- Live Game Analysis
- Player Strengths and Weaknesses Analyzer

2.1 Best Alternatives for Candidate Moves

At any given board position in a training game where it is the users turn, the user can view the best possible candidate moves alternatives with reasons as to why they are best from a positional standpoint.

2.2 Opening Move Explorer

The user can load all of their previous games and see statistics about their gameplay in the form of the most frequently played openings as white, favorite defenses as black, most losses or wins in particular opening lines, blunders and weaknesses from those openings etc.

2.3 Interactive Graphs for Wins / Losses / Draws

In depth statistics and charts which the user can manipulate interactively using filters, different color coding and choosing only selective games. As an example, the user may not want stats about games which were played during their learning phase. So, they can selectively filter those games out using interaction.

2.4 Live Game Analysis

This feature can be quite helpful for commentators on a chess game who, usually just run a chess engine on the game which only gives the strength of the move played and the next strongest move in response to that. Our system will give insights about how isolated or doubled up pawns will be bad in the future endgame, how blocked pieces will slow the development, which player has more control of the center of the board, etc. This feature is also useful for a player playing a live unrated game against a real opponent looking forward to improve their game.

2.5 Player Strengths and Weaknesses Analyzer

Suppose the user is going to play a match against a known opponent the next day. The user can feed all of their opponent's games into the system. The system will then perform a complete analysis of the strengths and weaknesses of the opponent and find openings that will be the most punishing to them. The system will also find their development faults and advice the user on how to take advantage of those weaknesses. The user can also perform the same analysis on their own games and look forward to negate their own weaknesses.

3. IMPLEMENTATION

The implementation of this system can be done by dividing it into various stages. Since it is a Data Science project it will follow the typical cycle of any Data Science project in general. [as shown in Fig – 2] These steps are further explained in-depth in the coming subsections.

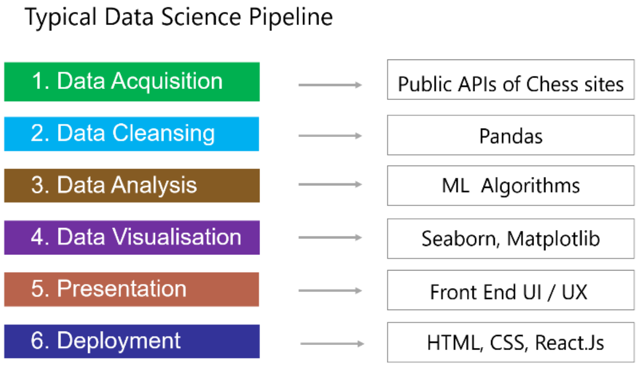

Fig - 2: Various steps involved in the implementation of the Project.

3.1 Data Acquisition

Chess games data is available in a very systematic digital format called the "Algebraic Notation". This data is stored in a file with extension ".PGN". These files contain each and every information about the game such as what Event was the game played in, the Date on which it was played, the moves made in the game, the result, who played the White pieces, who played the Black pieces, etc. This information can be easily parsed by computers. Some of the sources of acquiring these types of datasets are Lichess.org, Chess.com, ChessDB, etc.

3.2 Data Cleaning

Data Cleansing is done by removing the anomalies in the data. Many real-world datasets are incomplete due to factors such as data collection failures or misalignments between fused datasets.[2] This data can be refined by removing missing entries, incomplete data, unwanted excess information from the PGN files. Pandas library in Python is used to load up the moves into CSV tabular formats and remove all the unwanted annotations and comments linked with the moves to retain only the algebraic notations.

3.3 Data Analysis

In this phase, a Decision Tree will be constructed based on the moves that have been played in the game (as shown in Fig-2). Decision Trees can be used as discriminative classifiers. They can be used to choose the strongest next move from several good candidate moves. The classification algorithm tests each move against a depth of 'n' future moves and chooses one of the 'k' possible candidates. Each leaf node represents the current candidates.[3] Also, various Machine Learning Algorithms can be used to analyze given information and provide a richer understanding of the data contained in a specific context. Supervised Classification ML Algorithms can be used for making strong decisions and predicting opponent moves in Chess. [4]

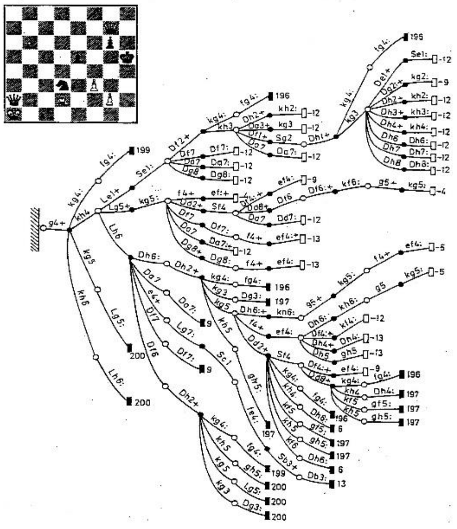

Fig - 3: Structure of the Decision Tree based on the moves played.

3.4 Data Visualization

The patterns and the insights found in the Analysis Phase will be turned into interactive visualizations in this phase. Interaction is conceptualized as a dialogue between the human user and the visualization system over a central object of interest – the data. Interaction plays an important role in Data Visualization. The visualization should take user input and alter the visuals according to the inputted queries for a deeper insight into the data.[5]

3.5 Presentation

In this phase, the front-end interface will be built for non-technical users to interact with the system. It will be in the form of interactive buttons, text fields, activities instead of command line prompt. This can be built using the Django or Flask Frameworks in Python.

3.6 Deployment

The project can then be deployed into the cloud to run on a remote server or hosted on a website domain. We can also choose to make it a desktop application or a web application.

4. CONCLUSION

With the current availability of the new found processing power, parallel computing and multi-tasking abilities of the latest lineup of computer processors, it is quite possible to design more systematic tools for an in-depth digital analysis of chess games played by amateurs and grandmasters regardless of rating levels for a much efficient ascend up the rating ladder. Our proposed system hopes to make it possible to enjoy and understand the logic behind chess games without the need of another human having to do the explaining behind engine analysis.

5. REFERENCES

1. Wei-Li Lu et al., “Chess Evolution Visualization”, IEEE Transactions on Visualization and Computer Graphics (Volume: 20, Issue: 5, May 2014)

2. Song and Szafir, “Where's My Data? Evaluating Visualizations with Missing Data.”, IEEE Transactions on Visualization and Computer Graphics (Volume: 25, Issue: 1, Jan. 2019)

3. Martine De Cock et al., “Efficient and Private Scoring of Decision Trees, Support Vector Machines and Logistic Regression Models Based on Pre-Computation”, IEEE Transactions on Dependable and Secure Computing (Volume: 16, Issue: 2, March-April 1 2019)

4. Yuri Nieto et al., “Usage of Machine Learning for Strategic Decision Making at Higher Educational Institutions.”, IEEE Access (Volume: 7, 27 May 2019)

5. Dimara and Perin, “What is Interaction for Data Visualization?”, IEEE Transactions on Visualization and Computer Graphics (Volume: 26, Issue: 1, Jan. 2020)