A Needle in a Data Haystack Introduction to Data Science - 67978 - Final Project

Chess Games Analyzing System

To access the project on GitHub use the following link: https://github.com/avihuxp/FinalAssignmentNeedle-

Video with the code running of the project: **https://youtu.be/UEmQcQigfwo**

Important note: We received approval from Prof. Dafna Shahaf to include a link for submission instead of submitting a code in the model.

Team members info (name, email, TZ, CS id)

- Itay Boiangou, itay.boiangou@mail.huji.ac.il, 322274697, Itay.boiangou
- Daniel Gibor (Bogatyrevich), daniel.gibor@mail.huji.ac.il, 334065281, danielgibor
- Amir Kelman, amir.kelman@mail.huji.ac.il, 208288811, Amir.kelman
- Avihu Almog, avihu.almog@mail.huji.ac.il, 315709980, avihuxp

I. The Problem Description

The system we developed is a **Chess Games Analyzing System**, is designed to achieve three primary objectives:

Assessing Correlation Between Chess Player Communities and Their Ratings

The system aims to reveal insights into the community structure of chess players in the site. We created a graph representing players who play against each other and examined whether we identified communities. We wanted to see if there is a link between the chess players ratings as nodes in the communities and if the size of the community affects this question.

Identification of the Most Influential Chess Players

The system aims to identify the most significant chess players. The criteria chosen to define a player as the most influential are by addressing the number of games played by the player, as well as the nature and complexity of the game, meaning the length of the game and the number of moves and reason for termination of the games.

Analyzing the Relationship Between Chess Game Volatility and Elo Ratings

The system examines the questions whether lower rated players have more chaotic games then higher rated players. Meaning lower rated players have less control over time on their games.

II. The Project Data

The data published on the lichess.org open database website will be used for the project. We will analyze 10,000 to 30,000 (about 60MB) chess games played on the website lichess.org, in January 2017. The file is encoded in PGN format. While it includes each game information: date and time of the start of the game; players names, rating and outcome (who won the game;) analyzed data by the game: which opening was identified by the chess engine, the time control for the game, termination reason and game information, such as the moves played by each player.

III. The Project Solution and Evaluation

Assessing Correlation Between Chess Player Communities and Their Ratings

Our Solution:

Our approach to this solution was first to create an undetected graph of players based on their chess games. Each player represents a node in the graph and a weighted edge between the players who played the same game. The weights are evaluated by the number of games the players played against each other [Figure 1]. On this graph, we ran the Louvain **algorithm**; In order to try and reveal communities of players who played each other on linchess site [Figure 2].

In the second stage, after viewing the created communities, we calculated the average standard deviation of elos as a function of the number of communities or community size the ratings [Figure 3]. We took, the average of the elos of players in the same community is computed. Then, we calculated **Manhattan distances** from the average of each elo in the community, and finally, the mean of the Manhattan distances is computed.

Setup:

The data available to us behaves as a logarithmic function [Figure 4]. This means that the number of players who have played with each other decreases as a function of the number of games. To validate our findings statistically, we chose the **threshold** of minimum played games by a player. Additionally, we reviewed all communities, varying thresholds to improve the qualities of the communities based on their modularity.

Evaluation Criteria:

To validate our results of our communities, we wanted to review the number of players who actually played on the site multiple times. The result could indicate the protentional size of communities – low number of returning players could mean much weaker and smaller communities, and high number of players could point to the exact opposite - based on our definitions of this graph. After reviewing the results We found that most of the players played very few games, which fits our results for the communities formed by Louvain [Figure 5].

By comparing the nature of the the logarithmic behaviour of the data on the number of games played by different pairs of chess players, and between the resulting distribution of the standard deviation of the player's ratings as a function of their amount, we are convinced of the non - randomness of the results obtained [Figure 6].

Results:

We found that the modularity naturally correlates with the size of communities: as the communities grow, there is a tendency for the modularity to decrease [Figure 5]. Based on the histogram in [Figure 6], we found that players within the same community differ in their ratings, and therefore, it can be concluded that players do not prefer to choose opponents with the same rating. Additionally, the rating values are not related to the size of the Louvain community.

Identifying Influential Chess Players using PageRank Algorithm

Our Solution:

We decided to identify the most influential players on the site by implementing a **PageRank-based player influence model.** First, we built a directed graph representing the chess player network, where: nodes represent individual players and edges represent games played between players [Figure 8].

The main idea lies in the edge weight calculation:

Primary Factor – Number of games each player played.

The base weight of an edge is determined by the number of games played between two players. This serves as the primary indicator of player interaction and potential influence on the site.

Secondary Factors: game length, outcome, ending.

We added more factors to wight in the edges in order to "quality" the player's game.

Game Outcome: Wins slightly increased the edge weight[1], Losses slightly decreased the edge weight.

Game Length: Longer games (more moves on the game) were given slightly more weight[2].

Checkmate Endings: Games ending in checkmate received a small boost[3].

The final edge weight was calculated by combining these factors, with the game count serving as the base and the secondary factors acting as modifiers.

To ensure the relevance of our analysis, we implemented a minimum game threshold. Players with fewer than 4 games were excluded from the final graph. With them out of the way, we could focus the analysis on players with a significant presence.

Setup:

[1] Wins increase edges weight d by multiplier of 1.1 and losses by a multiplier of 0.9. longer games
[2] We used a formula: min(number of moves / 40, 1.2) to cap the influence of extremely long games
[3] (multiplier of 1.1)

After constructing our weighted graph of chess players, we applied the PageRank algorithm to identify the most influential players. Our setup for the PageRank calculation was as follows:

Weight Consideration: The PageRank calculation took into account our custom edge weights, which primarily reflected the number of games played between players, with minor adjustments for game outcomes and quality.

Normalization: The resulting PageRank scores were normalized to sum to 1, providing a probability distribution of influence across all players in the network.

Convergence: The PageRank algorithm was set to run until convergence[4], in order to ensures a high degree of accuracy in the final influence scores.

Evaluation Criteria:

PageRank Score Distribution:

We examined the distribution of PageRank scores across all players. The wide range of scores indicates a clear differentiation between more and less influential players. We looked for a power-law distribution that indicates a natural hierarchy of influence [Figure 9].

Correlation with Game Count:

We calculated the correlation between a player's PageRank score and their total number of games played. The positive correlation validates that our model rewards active players, which aligns with the concept of influence in the chess community by the method we chose to measure their influence. We also looked for deviations from this correlation, which might indicate players who are influential despite playing fewer games (could be more skilled players.

Results:

The top 12 most influential players were successfully identified, with a strong correlation between the PageRank scores and players who frequently played on the site with a favor to stronger players (more wins and higher quality games). The influence network visually demonstrated the relationships between top players and their rivals, showcasing the most connected and influential players.

Correlation between PageRank and game count: 0.6096375725775064

P-value: 1.7027227021224194e-30

Analyzing the Relationship Between Chess Game Volatility and Elo Ratings

Our Solution:

Our approach to this solution was first to try and find a meaningful evaluation method to quantify game "chaos" or volatility. The core analysis of the chaos calculation is by making use of the Stockfish chess engine[5]. The engine is used to evaluate positions throughout each game and to come up with these factors to provide a view of game dynamics:

1. **Volatility**: We measured significant swings in position evaluation, using a threshold that varied based on the players' Elo ratings – in order to take into account that significant swings may differ between professional and casual players.
2. **Mistake Frequency**: We categorized moves as blunders, mistakes, or inaccuracies based on the magnitude of evaluation changes, each weighted differently in the calculation.

Integration: The final chaos score combined these two factors into one holistic score that includes a calculation based on these factors up to 20 moves for each game.

[4] with a tolerance of 1e-6

[5] In this project we use the **version** chess engine Stockfish version 17.

Setup:

To validate our findings statistically, we collected the data to ensure fair comparison for applying these followed up techniques .We normalized the values of the elos and chaos scores and also made sure we took only games[6] that differ by a maximum of 50 Elo points from the average points of the game.

After calculating the chaos score and average elo for each game, we proceeded to apply **K-means clustering** to identify patterns in the relationship between Elo ratings and game volatility. To determine the optimal number of clusters for our K-means algorithm, we employed the elbow method for a range of k values, from 1 to 10 in our case [Figure 10].

Evaluating results:

Throughout our analysis, we created dual-plot representations that showcased both normalized and original data distributions. These visualizations were made to better analyze the clustering results and understand the overall trends in the dataset [Figure 11].

We decided to create similarity matrices [Figure 12] to review and assess the quality of our clustering results:

Cluster Similarity Matrix: We created a matrix showing which data points were assigned to the same cluster, providing a clear visualization of the cluster structure.

Distance Similarity Matrix: We computed pairwise distances between data points in the feature space, converting these distances to similarities. This matrix showed the actual proximity of data points based on their Elo ratings and chaos scores.

Agreement Matrix: By comparing the cluster similarity and distance similarity matrices, we created an agreement matrix. This highlighted areas where the clustering aligned well with the actual data structure and where discrepancies existed.

With this in mind, we were able to assess the coherence of our clusters by examining how closely related points within the same cluster were - and we found that they were not.

Results:

Ultimately, our statistical validation indicated that the differences, after close examination of the plots, matrices, and elbow method results, in chaos scores across Elo groups were not statistically significant. The elbow method provided insights into the optimal number of clusters, but the lack of a clear "elbow" in the plot further supported **our conclusion that there is not a strong, distinct relationship between Elo ratings and our defined chaos metric.**

IV. The Project Impediments

1. **Incomplete player data:** Handling missing, or incomplete player data was challenging, as some games had incomplete player names or missing ELO ratings and many without evaluations for the data.

   Solution: These games were filtered out, and only those with complete information were included in the analysis. We used a separate engine to evaluate the players moves on the board – the Stockfish engine.

2. **Limited Number of Games:** Some players had too few recorded games, which could distort their influence scores in the PageRank algorithm and the overall understanding of communities using the Louvain – based algorithm. In the case of Louvain base algorithm, the logarithmic distribution of player interactions [Figure 4] suggested that as the number of games increased, the number of players playing against each other decreased. This sparsity made it challenging to form robust communities.

   Solution: In an identifying inflectional chess players using PageRank – based algorithm, a threshold was set (min 4 games) to remove players with too few matches to contribute meaningful influence. Threshold was also used in the Louvain – based communities' solution to establish the minimum games parameter.

[6] Each position is evaluated based on the advantage in centipawns, where 100 centipawns equal to one paw advantage in game.

3. **Computational Limitations:** Stockfish engine evaluations can take up an extremely long time for a much accurate evaluation. It reviews at all possible moves from the current position into a really large depth (can be up dozens of moves).

   Solution: Creating an upper bound for time for the evaluation to be calculated of 0.1 seconds per move, this could have distorted some of the information but while maintaining a somewhat more reasonable time to evaluate the results.

4. **Community Detection Accuracy**: When running the Louvain algorithm, we faced difficulties in accurately identifying distinct communities due to the complex and overlapping nature of player interactions.

   Solution: We varied the thresholds for minimum played games to evaluate the quality of detected communities based on their modularity. This iterative approach allowed us to refine our community detection process and assess different scenarios to optimize community formation.

V. The Future Work Prospects

Additional Metrics: Incorporating more sophisticated player performance metrics like opening strategies, time control variations, can increase our accuracy metric of the influence of the chess players.

Digging Deeper into Game Dynamics: We believe that we have only scratched the surface of what makes a chess game "chaotic". A deeper and more breadth look at chaotic game factors like time pressure, opening choices, or even player styles could help and refine the results.

Aiding players to better assess their games: we hope that future development could integrate this use machine learning tools to accurately evaluate the chaos for chess players game. In this way, players could assess how chaotic their games were and could maybe learn how to improve their control over the game over time.

Impact of Opening Strategies: Analyzing how different opening strategies correlate with player ratings and community membership. This could help us reveal trends in popular openings within specific communities and their effectiveness against different player types.

Advanced Graph Metrics: Beyond modularity, exploring additional graph metrics could have help us to provide further insights into player interactions. For instance, analyzing which players are central to multiple communities might highlight key influencers in the chess community.

VI. Conclusions

Assessing Correlation Between Chess Player Communities and Their Ratings: on this part of the system, we managed to identify distinct communities and examined the relationship between community characteristics and player ratings. Our findings indicate a clear correlation between community size and modularity, with larger communities tending to exhibit lower modularity. Additionally, the analysis revealed that players within the same community typically have diverse ratings, suggesting that there is no preference for selecting opponents with similar Elo ratings. Which was a surprising insight; To the best of our knowledge, the matching players algorithm on the site matches players randomly based on their elo. This idea could suggest that the formed communities by Louvain could indicate friend groups that exists outside of the linchess site.

Identifying Inflectional Chess Players using PageRank Algorithm: on this part, we successfully identified the most influential players on the chess site using a PageRank-based influence model. We believe that the PageRank technique was a valuable tool the review the influence of the players. By integrating it with other tools, and reviewing the influence over time, we could reveal have an even more accurate representation of influential players.

Analyzing the Relationship Between Chess Game Volatility and Elo: our surprising result on the relationship between game chaos and elo rating challenges the stigma of lower elo players having much more chaotic games. We believe that by expanding our data to many more games and specifically to take a deep dive into much lower elo games (that were not found on this particular data set and are known to be rare on this site) could give us a more accurate on this relationship and to give more specific recommendation for players on how to improve their games(through gaining more control over them over time).