

Predicting Chess Opening Through Modelling Of Chess Opponents

Debarpan Bose Chowdhury^[1] ; Banashree Sen^[2]

^[1] Department of Computer Science and Engineering, Heritage Institute of Technology, WB, India.

^[2] Department of Applied Mathematics, Maulana Abul Kalam Azad University of Technology, WB, India.

Abstract: A chess opening is the preliminary stage of a chess game which typically consists of moves from formerly analysed openings. Opening strategy plays a crucial role in the entire game and decides the destiny of the middlegame and endgame. Here in this article, we attempted to introduce a method to predict the opening moves of a specific opponent. The technique analyses the past games played by a specific player to discover the most probable opening that the player is going to play in the subsequent games. The overall performance of this method is analysed and demonstrated by taking several factors such as transience, average turn and opponent's response in a given situation into consideration. This comparative study enables us to acquire the knowledge about the opening preferences of an opponent, which gives a strategic advantage to a chess player and helps them to develop a game strategy from the beginning of the game. In this article, we attempted to give an overview of how to predict the most favoured opening of a player and how we can utilise it for our benefit^[10].

Keywords: Chess Opening, Opponent, middlegame, endgame.

1. Introduction

Chess is a board game played between two players. A board in chess consists of sixty-four squares, and every square is uniquely identified by a letter from 'a' to 'h' together with a number from '1' to '8'. Each player has sixteen pieces to play with. These pieces include one king, one queen, two rooks, two knights, and eight pawns with different point value assigned to each of them based on their importance. The main aim of this game is to checkmate the opponent's king.

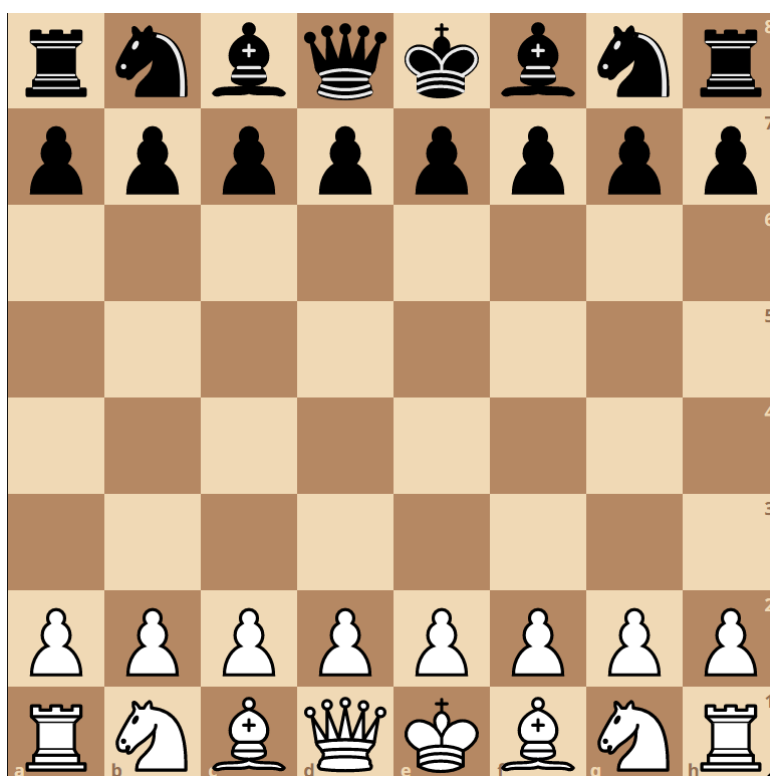


Fig.1. Chess Board ^[1]

A chess game is composed specifically of three parts - Opening, Middle game, and endgame ^[2]. All the games might not consist of all the three parts. But the opening is an integral part of a chess game and leads to the development of the middle game. It is the initial stage of all the games wherein players particularly focus on the development of pieces, taking control of the centre and moving the king to safety ^[3]. In chess, the middle game is the part of the game that occurs between the opening and the endgame, though there may be no clear distinction between the opening and the middle game or between the middle game and the endgame. The middle game begins when each player has finished the development of most of their pieces and the king has been brought to relative safety. ^[4] The transition from middle game to endgame is frequently ambiguous, and it can occur gradually or abruptly with the exchange of a few pairs of pieces. The endgame, on the other hand, has distinct characteristics from the middle game, and the players' strategic concerns differ accordingly.

The major characteristic of a good chess opening is proper development and mobilisation of the pieces to govern the critical squares on the board. It additionally focuses on developing positions in which the player is more comfortable than the opponent via transposition ^[5]. But keeping this key feature of chess in mind, each player has his or her preferences regarding chess opening. Thus, we will analyse a person's game to get the flavour of his or her preferences. But the sad part of the tale is that we can predict a player's move up to a certain limit, after which they turn out to be quite inconsistent with their move. This is what makes the game of chess a lot interesting and further studies can be conducted to overcome this limitation.

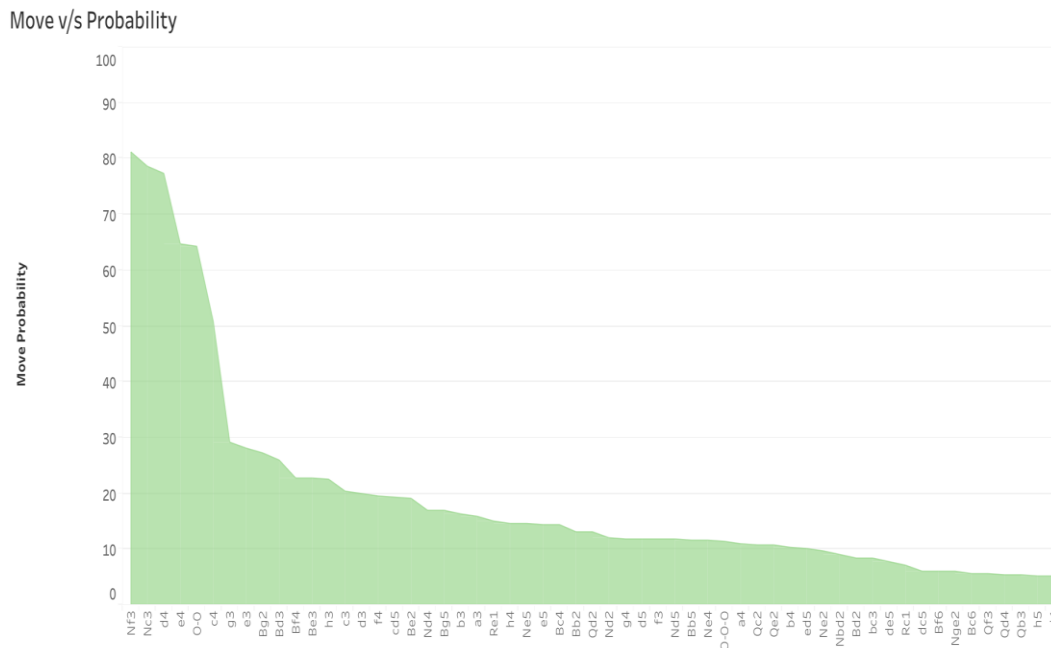


Fig. 2. A plot showing the probability of a particular move being played in a particular game as white (Source: Lichess Account: DrNykterstein)

The above diagram was plotted after analysing the games played by World Champion Magnus Carlsen. This shows that only a few moves appear consistently in all the games. These moves with a high confidence level can be predicted to determine the opening moves of the opponent. The number of moves with a high confidence level may vary from player to player, which would influence the accuracy of the prediction.

2. Factors Affecting a Chess Move

The number of moves in a current position can grow exponentially and due to this, the number of games of chess may be exceptionally large in number. Claude Shannon found a lower bound of the game-tree complexity of chess, resulting in about 10^{120} chess games [6]. He additionally confirmed that the possible number of chess positions may be roughly 10^{43} which is a massive number. The graph plotted below shows how the number of possible games increases exponentially with the number of plies.

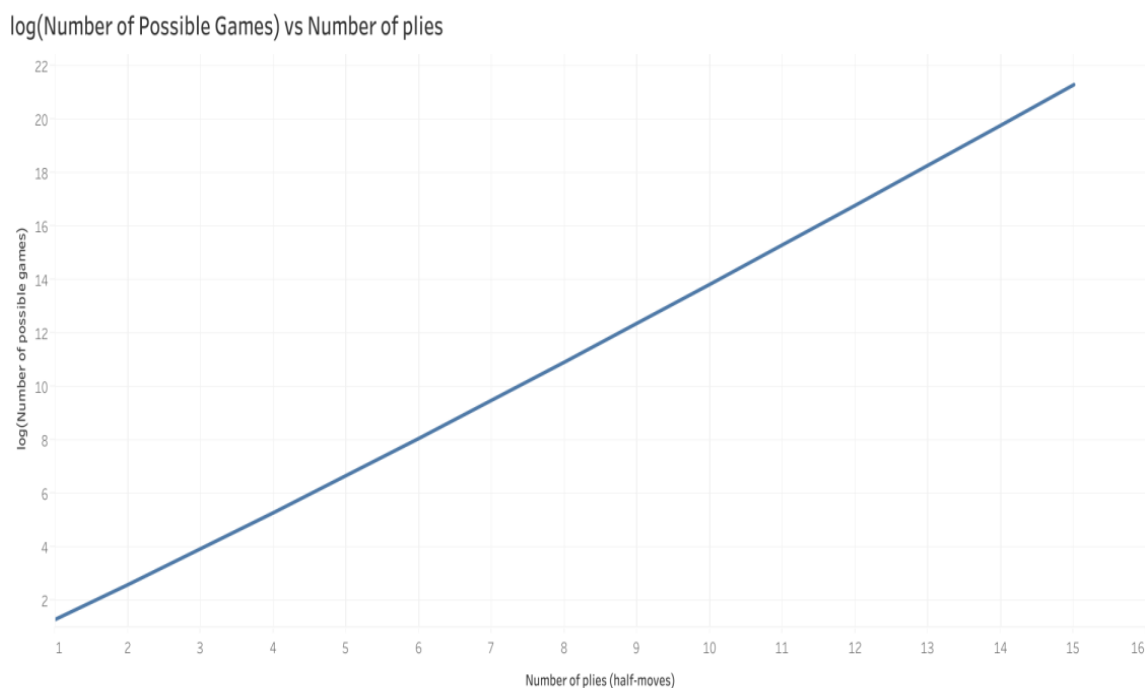


Fig. 3. A plot to show how the number of possible games increases exponentially with the number of plies

But legitimate and correct moves rely on numerous factors. Thus, decreasing the number of possible moves in a given position. However, we cannot use any deterministic approach to forecast the player's current move because it would demand enormous computing capacity. As a result, we choose to develop a probabilistic system to forecast a player's next action.

2.1 Current Position

It is quite a broad term and in turn, be affected by many factors. But here we are considering only the opening moves. Thus, a position in the opening with high probability is already analysed and thus we know what to play next.

2.2 Personal Preference

All chess players have their favourite openings and often like to stick to those openings. For example, some of the favourite openings of Magnus Carlsen are Ruy Lopez- Berlin Defence, Sicilian Defence, Ruy Lopez – Open Variation etc. The above-mentioned variations have some of the common moves for white such as e4 and nf3 and from Fig. 2. We found that the probability of Carlsen playing those moves is quite high.

2.3 Average Turn

A move may appear multiple times in a game. But as we are predicting only the opening moves, we can safely assume that a particular move will not repeat itself as it is advised by the grandmasters not to repeat a chess move in the opening. So, we use the average as the measure of the central tendency of the turn of a move.

Probability v/s Average Turn

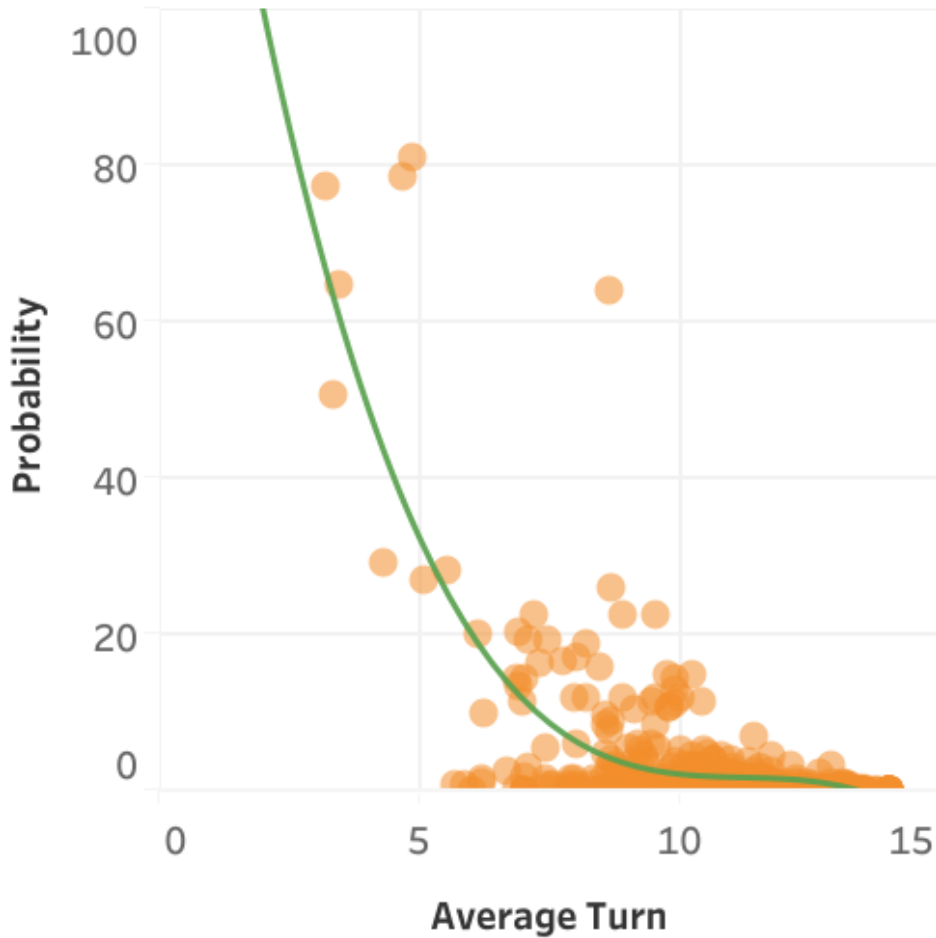


Fig. 4. Shows us the relationship between the Probability of the move and the Average Turn (R squared: 0.573187 and p value: < 0.0001) (Source: Lichess Account: DrNykterstein)

From the above graph, we found that a move with high average turn is less likely to be played. Thus, our earlier hypothesis which states that only the first few moves of the game are consistent is true.

2.4. Memory Retention

A chess player might try to change his or her style of play. This might happen due to a change of preference of move or due to some other cause. Suppose a player has played the move e4 in the first 50 games of his career but then he slowly shifts his preference from e4 to d4 and played 50 games using d4 as the first move. Thus, if we are asked which move will that player would start his game with then our answer should be d4. This phenomenon can be considered using Ebbinghaus' Forgetting Curve^[7].



Fig. 5. Forgetting Curve (Amount of data or memory retained with time)

2.5 Transposition

In chess, a transposition is a sequence of moves that results in a position which may also be reached by another, more frequent sequence of moves. Transpositions are particularly common in the opening, where a given position may be reached by different sequences of moves.^[8] Players sometimes use transpositions deliberately, to avoid variations they dislike, to lure opponents into unfamiliar or uncomfortable territory or simply to worry opponents.^[5]

For example,

- | | | |
|------------|-------------------------------------|------------|
| 1. d4 d5 | | 1. c4 e6 |
| 2. c4 e6 | Led to the same position as that of | 2. Nc3 Nf6 |
| 3. Nc3 Nf6 | | 3. d4 d5 |

3. Devising an Opening Predictor

3.1 Dataset Used: Data is fetched from the lichess database using lichess API^[9]. We fetch the games of a particular player and choose the colour we are interested in.

3.2 Algorithm

- Find the probability of a move appearing in a game (P1).
- Find the percentage of retention of each move using Ebbinghaus' Forgetting Curve (P2).
- Both above probabilities are independent of average turn. So, let us assume that the turn value of a move is normally distributed with average turn as its mean and the standard deviation can be calculated easily.
- Now calculate P1' and P2' using the distribution curve and P1 and P2, respectively.

- But the moves predicted using only these probabilities are independent of the current position so to make the prediction dependent on the current state we find out the move played by the opponent to predict our move.
- Use P1` and P2` to predict the moves

4. Mathematical Inference of the Model

4.1 Finding Probability of a Move (P1)

$$P1 = \frac{\text{Total number of games in which a particular move was made}}{\text{Total Number Of Games Analysed}}$$

For example, suppose that we have analysed 2000 games out of which 505 games have the move d4. Then the probability of the move $P1 = 505/2000 = 0.2525$. That is the move d4 appears in 25.25% of the games analysed.

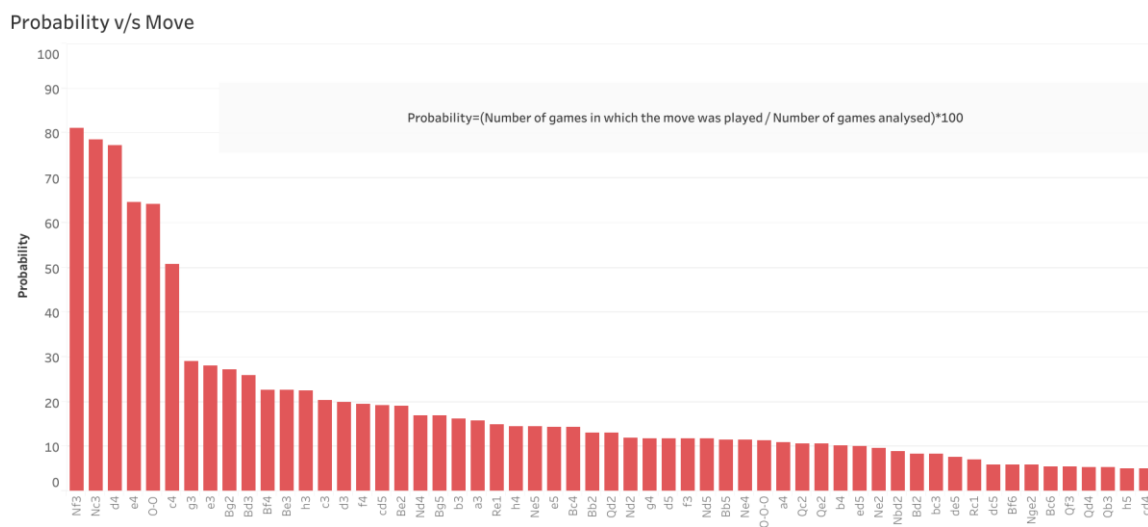


Fig.6. The Probability of each move analysed

4.2 Calculating Percentage of Retention

Hermann Ebbinghaus in his publication known as *Über das Gedächtnis* has provided an equation to approximate his forgetting curve.^[7]

$$b = \frac{100k}{(\log(t))^c + k}$$

Where b = percentage of retention and t = time in minutes and c = 1.25, k = 1.84. Here, we use this equation to calculate the probability of a move to be played after a period t.

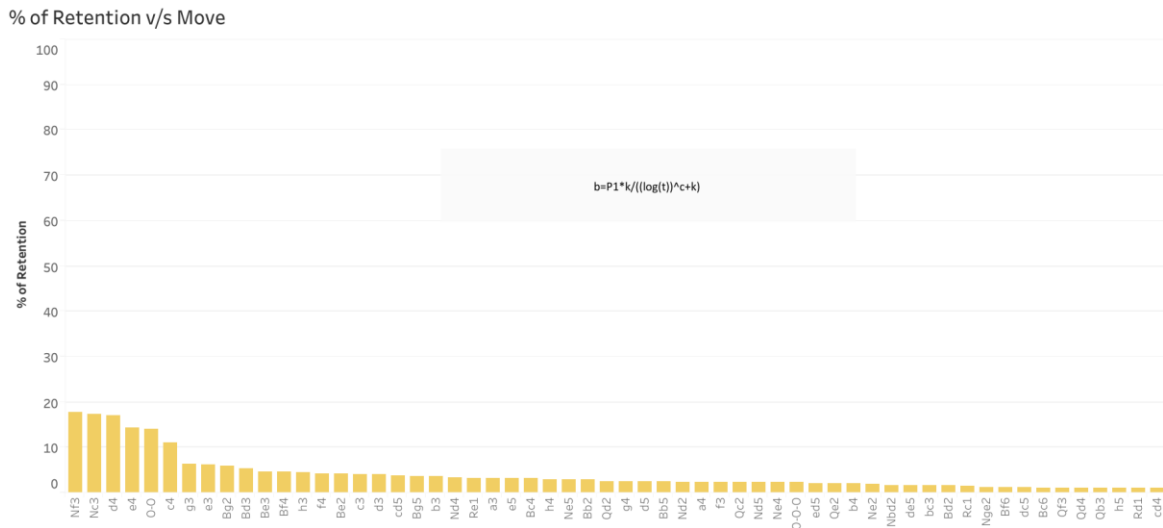


Fig.7. Percentage of Retention v/s Moves

This graph stands for the modified probability of a move being played. As discussed in section 2.4 that there might be situations where a player used to play a particular set of moves with a remarkably high probability but now, he prefers to play some separate set of moves. So, if we make the probability of a move time-dependent then we can handle the effect of change of style of play of a particular player over time.

4.3 Relating Average Turn to Probability of the Move

Here in this section, we introduce a method to relate average turn to the probability of the move. We can assume that the turn variable is normally distributed with mean $\mu = \text{Average Turn}$ and standard deviation σ can be calculated from the data. So, we can calculate the probability of a move at a given turn T by

$$P1' = P1 \cdot \frac{1}{\sigma \cdot (2 \cdot \pi)} \cdot \int_{k-\theta}^{k+\theta} e^{\frac{-1}{2} \cdot \frac{(x-\mu)^2}{\sigma^2}} \cdot dx$$

Where θ is a small value and k being the current turn.

As we know that probability at a point supposed to be k in continuous probability distribution curve is zero. So, we consider a small range around k and calculate the probability of a particular move depending on the current turn.

But the above equation is quite tough to calculate and we can approximate it to be

$$P1' \approx P1 \cdot \frac{1}{\sigma \cdot (2 \cdot \pi)} \cdot e^{\frac{-1}{2} \cdot \frac{(T-\mu)^2}{\sigma^2}} \cdot 2 \cdot \theta$$

As the value of θ is small.

Therefore, the value of

$$P2' = P2 \cdot \frac{1}{\sigma \cdot (2 \cdot \pi)} \cdot \int_{k-\theta}^{k+\theta} e^{\frac{-1}{2} \cdot \frac{(x-\mu)^2}{\sigma^2}} \cdot dx$$

$$P2' \approx P2 \cdot \frac{1}{\sigma \cdot (2 \cdot \pi)} \cdot e^{-\frac{1}{2} \cdot \frac{(T-\mu)^2}{\sigma^2}} \cdot 2 \cdot \Theta$$

5. Statistical Analysis of the Model

When we try to predict the first 5 moves of a game using only the factor of how frequently a move appears in a game (P1), we acquire an accuracy of about 54%. The accuracy will decrease as the number of moves to be predicted increases.

EXPECTED MOVE	PREDICTED MOVE
d4	d4
e4	Nc3
Nc3	e4
Nf3	Nf3
Bg5	Bg5

Table 1. Expected move v/s Predicted move for a game (Source: Lichess Account: pIlaba)

But one of the important things to note is that the moves in the Predicted set are the same as that of the Expected set. So, we need to maintain the order in which the move appears. This can be done by considering the current state or position and the turn variable. To do this we consider the move played by the opponent to predict the current move. This improves the accuracy to about 80%.

But the above two cases do not consider the factor of transience. Thus, considering that factor might increase the accuracy for some cases where there is a change in the style of game play of the player. For Example, if we consider a lichess account named theGladiator123 we would notice that there is a change in the style of play or preference of the player. He shifts his preference from e4 to d4 as the starting move for white. Thus, considering Ebbinghaus' Forgetting Curve will help us to increase the accuracy in this case.

6. Drawback of the Model

To improve the accuracy of our model, we only consider the previous move played by the player to predict the current move; nevertheless, in real life, a move is reliant on a series of moves played before it. However, considering a sequence of moves will enhance the complexity of our model. The above-mentioned model also fails to handle transposition correctly. However, if we could handle transposition correctly, the accuracy of our model would improve dramatically.

7. Summary and Conclusion

Other than focusing on our game play, analysing opponents' game play is a vital component of chess. Being able to predict one's opponent makes one an excellent chess player. This model will assist chess players in evaluating their opponent and determining what style of opening their opponent favours. Thus, this would give them an upper hand in the tournament. This also

decreases the effort required by players to recall a massive collection of openings by substantially lowering the size of that collection.

7.1 Future Scope of Study

As previously stated, the model presented in the study can accurately predict just the first few moves. Further research may be undertaken to see what additional elements may influence a player's opening move and how we may offer a better and more accurate model for predicting a specific player's opening move. We can also create an algorithm to effectively deal with the effect of transposition, which could also improve the accuracy of our model.

Acknowledgement

I would want to convey my heartfelt gratitude to Dr. Banashree Sen, whose assistance, stimulating suggestions, and encouragement aided me throughout the project. I would also want to express my thanks to Mr Rahil Miraj, BMath, ISI, and Bangalore for several fruitful suggestions. Also I would like to express my gratitude to Maulana Abul Kalam Azad University of Technology and Heritage Institute of Technology for pushing me to undertake this sort of project work.

Declarations

Ethics approval is not applicable whereas we use data from 'Lichess account'^[9], freely accessed from internet. Chess game information has been fetched by API^[9] calling for a particular individual which is license free and if needed for further review the statistics can be included anytime. The authors, DBS and BS declare that they have no competing interest and no funding as well.

References

1. Duplessis T (2010) lichess.org
2. Soltis AE (1999) Development of theory. Britannica
3. Fine R (2012) Ideas behind the chess openings. Ishi Press, Tokyo
4. Contributors W (2022) Chess middlegame. Wikipedia
5. Soltis A (2007) Transpo tricks in chess: Finesse your chess moves and win. Batsford, London
6. Shannon CE (1988) Programming a computer for playing chess. Computer Chess Compendium 2–13
7. Ebbinghaus H (2013) Memory: A contribution to experimental psychology. Annals of Neurosciences. doi: 10.5214/ans.0972.7531.200408
8. (2022) Transposition (chess). In: Wikipedia. [https://en.wikipedia.org/wiki/Transposition_\(chess\)](https://en.wikipedia.org/wiki/Transposition_(chess)). Accessed 22 Jun 2022
9. Lichess.org API Reference (2.0.0). In: Lichess.org API reference. <https://lichess.org/api>. Accessed 16 Jun 2022
10. Online Prediction of Chess Match Result, [10.1007/978-3-319-18038-0_41](https://doi.org/10.1007/978-3-319-18038-0_41), Conference: Pacific-Asia Conference on Knowledge Discovery and Data Mining (2015)