Analysis of cheating in human patterns using Stockfish suggestions

João Batista Andrade[1], Emanuel Ávila[1], Caio Oliveira[1]
Vitor Manuel Gomes[1], José Gilvan R. Maia[1], Artur Franco[1]

[1]Federal University of Ceará (UFC) CE – Brazil

{batistajoaoguns,cfoviana,vitor2215,emanuel.pires}@alu.ufc.br
gilvanmaia@virtual.ufc.br, arturfranco@ufc.br

Abstract. *Online games are growing, and cheating is becoming more common. Chess, considered a sport by many, often faces cheating in tournaments. The website chess.com daily blocks hundreds of accounts accused of cheating, often using engines like Stockfish. This study analyzes human patterns in thousands of matches using Stockfish to identify cheating. Results show that Stockfish analysis helps identify cheating by comparing players' Elo ratings and move sequences. However, this method alone is insufficient for definitive accusations, necessitating the combination with other robust methods to increase the accuracy of accusations, providing a more comprehensive approach to combating cheating.*
Keywords *Chess, Chess cheating, Cheating, Chess robots, Online Games*

1. Introduction

In digital games, various strategies have emerged for cheating, including the use of game injection tools, which are considered a more secure means of achieving victory [Martinson e Rangel 2023]. However, these strategies can be monitored through process state examination and packet analysis. Other cheating methods exist that violate established norms of human interaction within games. These include the use of machines or artificial intelligence programs that simulate the presence of a player. This practice can be observed in board games, such as chess, as well as real-time strategy games, facilitated by advancements and widespread adoption of deep learning and computer vision.

Chess is considered an intellectual sport by a segment of the chess community, characterized by its strategic complexity and demand for strong cognitive abilities [Aciego et al. 2012, Kobiela 2018]. Over the centuries, chess has evolved from an enjoyable board game into a competitive discipline, with a rich tradition dating back over five centuries.

Cheating in chess is a recurring phenomenon that has significantly increased since the popularization of online games [Hoque 2021]. With the advent of virtual gaming platforms and the increased access to them, players face an unprecedented competitive digital ecosystem. However, this new era of interactive entertainment also presents a major challenge: artificial intelligence. A significant event in human versus machine competition occurred when a chess engine defeated the top world champions [Iliescu 2020], demonstrating the machines' ability to rival and even surpass the best human players. However, this same technology has also brought ethical challenges. Chess engines, with their ability to quickly analyze millions of positions and suggest the best moves, are constantly used to seek unfair advantages. A site well-known to the chess community, “chess.com” [Chess.com 2020] , discusses this issue and explains the forms of cheating and how the use of engines leads to over 500 user accounts being banned daily.

The rise of chess as an official sport is evidenced by the existence of national and international associations dedicated to its development and regulation. These organizations, such as FIDE [Federation 2024], play a crucial role in organizing competitions, establishing rules, and recognizing the best players. In particular, FIDE it plays a central role in promoting the sport worldwide, encouraging participation, ensuring fair play standards, and fostering cultural exchange through chess.

The study analyzes sequences of moves similar to those suggested by the engine used, aiming to identify movement patterns across various matches for each player. The focus is on the mid-game and endgame moves, where advantages can be obtained. It is important to highlight that this method alone is not sufficient to detect cheating. Therefore, this study aims to add a new identification mechanism, capable of assisting in the detection of cheaters in conjunction with other methods already used on online sites.

2. Related Work

In Barnes, 2015, the authors utilize analysis methods based on the researcher [Regan et al. 2011][regan 2014], who is a pioneer in the analysis of chess engine moves and cheats. Regan et al. 2011, developed several analysis methods, and [Barnes e Hernandez-Castro 2015] uses some of these methods for their analysis. Both works have made significant contributions to the theme studied by the current article. The most obvious method, according to Barnes, is to calculate the percentage of human moves identical to the engine. This method is named MM by Regan. Barnes introduces the CV, which, unlike MM, compares the proportion of non-opening moves by a player with the same evaluation as the move chosen by an engine. However, Barnes et al, 2015, states that this is a very rudimentary method of cheating assertion, as occasionally it may yield false positives. This article differs from Barnes’ study by using a more updated engine, a dataset with numerous complete games, and multiple games under the same ID, resulting in more than one game per player, increasing the likelihood of cheating. Additionally, this study aimed to analyze sequences of moves coinciding with the engine. This sequence is defined from the beginning of the coincidences to where they end. If the player is not coincident with the engine and subsequently executes 5 consecutive moves, it results in an advantage over the opponent.

3. Methodology

The dataset contains 20,058 games in PGN format and 15,635 different player IDs, with an average of 50 moves per game. This is a collection of chess games extracted from Lichess [lichess.org 2024], an online chess platform that allows users to play games against other players over the internet. There are games with only 1 move and games with over 300 moves. Players are identified by their IDs and the colors of their pieces. Additionally, the dataset provides all the moves of each game, openings, winner (if any), and their ELO rating

Available since 2017, the dataset has been downloaded over 43,000 times. It will be used to test the utilization of the chess engine, thus generating a new dataset that will allow for detailed data analysis.

The initially created dataset contains all the moves used, followed by the moves indicated by the Stockfish engine, along with the result: `True` or `False`, as well as the color of the piece. The player of the white pieces will always be identified in even-numbered lines, while the player of the black pieces will always be identified in odd-numbered lines, based on the fact that the dataset starts from line 0. When a game ends, a blank line is added to separate them, and the board is cleared so that a new game can begin. Each process in the analysis using Stockfish consumes an average of 100 MB of RAM and uses over 80% of a core, allowing parallel work, distributing tasks among multiple CPU cores to speed up the analysis process.

The experiment was conducted in Python 3.10, with Python-chess library allows for creating and validating moves, in this case, using the UCI protocol to communicate with the Stockfish version 16 chess engine, enabling moves to be made in parallel with the dataset.

4. Experiments

The experiment was carried out with a group of players who stood out for their sudden variation between victories. The time required to conduct the experiment was based on the engine's search depth multiplied by the number of moves in the entire dataset, resulting in more than 10 days of processing. To reduce this time, the processing was divide 8 parallel running groups, powered by Tilix, of 2,507 each. With the analysis engine set to a search depth of 1 second per move, each terminal took an average of 42 hours to process the games.

The second experiment arose from the idea of evaluating sequences of moves coinciding with the engine that indicate the tool's usage. For instance, a player with more than five games who start with poor performance, with many "False" outcomes identified by the engine, but ends the games with many "True" outcomes coinciding with the engine, providing an advantage over the opponent or even victory.

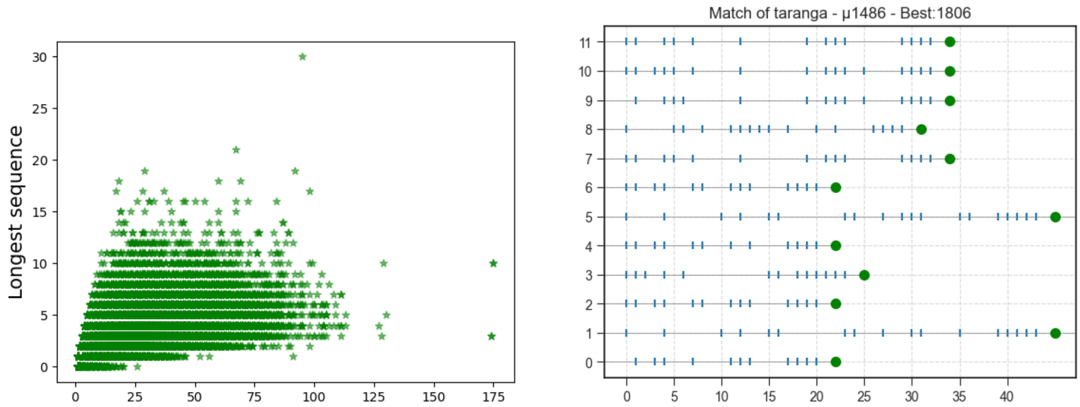
To do this, the second dataset was created, and from it, a Python script was developed to analyze each line from a specific column. In this experiment, moves with more than ten moves and more than two games per player were filtered. However, this time, the final moves were calculated in sequences of at least four. For example: "1,0,0,1,1,1,1". As shown in the example, the number one repeats four times in a row at the end of the game, reflecting the "True"or "False"values from the first dataset in binary.

In addition to this experiment, another one was conducted, similar to the previous one, but using a binary analysis model. In this case, sequences of 3 "True" that repeat at least three times were filtered. For example: "1,1,1,0,1,1,1,0,1,1,1". This changes the type of sequence and provides a new analytical perspective on the moves.

5. Results and analysis

Next, the results obtained after the experiments will be analyzed, addressing visualization analyses of the best games that show signs of cheating, focusing on players who exhibit this characteristic.

Figure 1(a) illustrates the largest sequences obtained for each game of each individual player. The y-axis represents the largest sequence of consecutive coincidences,



(a) Longest sequence of coincident movements. (b) player suspected of cheating.

Figura 1. Results of the cheating suspect's sequence and games.

while the x-axis represents the number of moves in the games. By crossing the axes, a game with 100 moves, of which 30 are consecutive coincidences with the engine, can be identified. Initially, it can be stated that, compared to the total number of moves, the sequence is relatively smaller, representing about one-third of the total. However, depending on where this sequence starts, it may be that the player used this strategy to bypass an MM/CV identification mechanism and used an engine for assistance at the end of the game, aiming to win. This strategy is openly known and adopted by many cheaters. However, if this strategy is used too frequently, i.e., in several games, this cheating method becomes more exposed to investigations. Obviously, depending on the player's ELO rating and history, accusing them of cheating is less likely.

As observed in Figure 1(b), the graph shows the total number of games on the y-axis and the number of moves on the x-axis. Each game is interpreted as follows: if the vertical symbol exists, it means the move coincides with the engine (`True`); if there is a blank space, it means the move was different (`False`). Taranga has 12 filtered games, all victorious, indicated by the green dot. Combining this with the opponent's lower Elo and the high ratio of moves coincident with the engine, it is inferred that Taranga used tools to secure victory. This inference is based on the observation that Taranga often has a poor start, not coincident with the engine, which is unusual for well-ranked players who typically follow the moves from the ECO. Additionally, he has many good moves in sequence towards the end, leading to victory. Although Taranga has many good moves, there are some blank spaces, but no losses, unlike many higher Elo players who played better. This suggests a high probability that Taranga used means to avoid detection of cheating, such as using the engine at opportune moments.

As shown in Tables 1 and 2, Taranga has a lower Elo rating than his opponents and also a lower coincidence ratio. Based on the fact that players with higher Elo and better accuracy should be the champions of the game and yet lost, Taranga can be accused of using an engine at opportune moments. However, as shown in Table 1, resulting from the final coincident sequence filtering, this player's name only appears once, showing that the method was not used more than once, making it a false positive if accused solely by this methodology. Since all players in this table appear only once, except for *peer1966*, all of them can be excluded from accusation, as according to Barnes, accusing a player of

Tabela 1. Sequence with coincident final.

Color	Player	Rating	Movement Ratio	Oppenent	Rating	Movement Ratio	Winner
black	oldpaths	1568	0.45	porquepepe	1618	0.50	black
black	taranga	1289	0.42	striker123	1806	0.45	black
white	tfeng	1579	0.48	ironboy	1623	0.51	white
black	per1966	1670	0.33	saviter	1689	0.52	black
black	per1966	1670	0.38	saviter	1689	0.52	black
white	saviter	1709	0.45	leps000	1716	0.49	white
white	vfhfnvfhfn22	1602	0.39	natyulpan	1671	0.41	white
black	masterzyd	1272	0.30	pradeeprajjjj	1311	0.43	black

cheating based on only one suspicious game is impractical.

Tabela 2. Sequence of 3 followed during a match.

Color	Player	Rating	Movement Ratio	Oppenent	Rating	Movement Ratio	Winner
white	taranga	1280	0.39	moon50	1485	0.41	white
white	taranga	1280	0.35	moon50	1485	0.39	white
white	taranga	1280	0.40	moon50	1485	0.46	white
white	kardsalan	1937	0.46	mrhorta50	2090	0.50	white
white	skulll1	1710	0.47	narcad	1712	0.48	white
white	butterlandz	1278	0.36	ki_f	1463	0.45	white
white	crax_is_bax	1637	0.35	sebasfei89	1661	0.43	white

In the case of *peer1966*, the Elo difference is very small, making the game fair for any participant, but it does not exclude the possibility of cheating in the game.

In Table 2, resulting from the filtering of 3 repeated coincident sequences, Taranga appears 3 times, one of the few players repeating in this filter. However, the Elo difference is very small, making the game balanced. If we compare the two tables mentioned in this section, we can observe that finishing with many coincident moves is rarer than hitting sequences of three during the game. This adds strength to the analysis of the end of games as a viable methodology, as sequences of three in a row are more common among players. However, it is important to note that this method cannot be used alone, as even the strongest methods present false positives in this context.

6. Conclusion

The analysis proved effective in some cases, particularly when a higher Elo-rated player loses to a lower Elo-rated player who made poorer moves but showed significant similarity to Stockfish at the end of the game, gaining an advantage and winning. This demonstrated that concrete evidence of cheating is unlikely, even when examining movement patterns and Elo information of each player, though false positives can still occur. However, integrating this model with others in the field may increase confidence when penalizing cheaters.

As for future work, the intention is to include additional analytical variables, such as the response time for each move. Furthermore, expanding the datasets to include more games and players, as well as increasing the decision time of the Stockfish engine and utilizing datasets where cheating is confirmed and where it is not, such as matches where the engine plays against another engine, are planned.

Referências

[Aciego et al. 2012] Aciego, R., García, L., e Betancort, M. (2012). The benefits of chess for the intellectual and social-emotional enrichment in schoolchildren. *The Spanish journal of psychology*, 15(2):551–559.

[Barnes e Hernandez-Castro 2015] Barnes, D. J. e Hernandez-Castro, J. (2015). On the limits of engine analysis for cheating detection in chess. *Computers & Security*, 48:58–73.

[Chess.com 2020] Chess.com (2020). Sobre trapaça no xadrez online. https://www.chess.com/pt/article/view/trapaca-xadrez-online/. Acessed on July 27, 2024.

[Federation 2024] Federation, I. C. (2024). International chess federation. https://www.fide.com/. Acessed on July 27, 2024.

[Hoque 2021] Hoque, M. (2021). *Classification of Chess Games: An Exploration of Classifiers for Anomaly Detection in Chess*. Minnesota State University, Mankato.

[Iliescu 2020] Iliescu, D. M. D. (2020). The impact of artificial intelligence on the chess world. *JMIR serious games*, 8(4):e24049.

[Kobiela 2018] Kobiela, F. (2018). Should chess and other mind sports be regarded as sports? *Journal of the Philosophy of Sport*, 45(3):279–295.

[lichess.org 2024] lichess.org (2024). Lichess webpage. https://lichess.org/. Acessed on July 27, 2024.

[Martinson e Rangel 2023] Martinson, F. e Rangel, D. (2023). A comprehensive analysis of game hacking through injectors: Exploits, defenses and beyond. *International Journal of Computer Applications*, 185:8887.

[regan 2014] regan (2014). Kenneth w. regan's chess page. https://cse.buffalo.edu/ regan/chess/. Acessed on July 27, 2024.

[Regan et al. 2011] Regan, K. W., Macieja, B., e Haworth, G. M. (2011). Understanding distributions of chess performances. In *Advances in computer games*, pages 230–243. Springer.