An Analysis of the Use of Centipawn Loss to Detect Cheating in Chess

James Horbury
University of Connecticut

October 10, 2022

Introduction

Recently there has been a very high profile case of alleged cheating in the world of professional chess where the current reigning world chess champion, GM Magnus Carelson, was defeated by GM Hans Niemann, a player that was ranked 42nd in the world and experineced massive growth in FIDE rating in the past two years. This has caused many skeptics in the chess community, including Carelson himself, to suspect cheating may have been at play. The ensuing comotion has caused many hobbiests to stop and consider how pervasive cheating really is in chess, more noteably in online venues, and what methods are employed to flag games with alleged engine-aided moves. This paper will surve as an analysis of the currently available methods of cheat detection and for which cases they are more or less robust.

Current research (Regan et al., 2011) has been done by Dr. Kenneth R. Regan, of the Computer Science and Engineering Department of the University at Buffalo, employing the use of a model that utilizes the Jenson-Shannon Divergence as a basis for detecting player vs. engine games. However, there is much scrutiny over the transparency of his methods and their effectiveness on different "intensities" of cheating. There are also other open-source options () that can provide an indepth analysis of move-for-move comparisons in specific games, an area where Dr. Regan's model appears to be weak.

Specific Aims

The most widespread methods of cheat detection rely on large amounts of data from the past games of specific players, such as graphing their average centipawn loss per game against their historic FIDE rating. What other alternative methods of cheat detection can be reliably used where the norm falls short?

The purpose of posing this question is to expose the inner-workings of the methodologies behind dectecting cheating in chess in the digital age and provide scientific reasoning to support accusations of cheating in instances where they are valid.

Data Description

I'll be utilizing a dataset of Data for 20,000+ chess games played on Lichess (Bhat, 2021) (including moves, victor, rating, opening details, etc.) to provide insight into how centipawn loss relates to FIDE score in the general population, then do the same using historical data for GM Magnus Carelson (Qureshi, 2022b) and GM Hans Niemann (Qureshi, 2022a) to observe any differences. These are user submitted examples through Kaggle, but I should be able to have the same access to the games through the Lichess website.

Research Design/Methods/Schedule

To reitterate, I plan to plot average centipawn loss per game vs. FIDE score for both the general population and for the GM's in question to illustrate the common view that this is a good indicator of cheating in chess. If other examples can be found that follow a growth pattern similar to Niemann's I'll present them also to offer a fair counterargument to the claim that his success is "unnatural". Then, using the opensource move-for-move comparison software mentioned in the Introduction I can go more in depth into how, if possible, Niemann could/could not have cheated in his game against Carlsen. Regarding specific metrics, I'll have to see what the PGN-Spy actually offers in terms of applicable output data and take my analysis from there, which will require more research than what was done for this proposal. Additionally some data cleaning/preparation will be necessary for each Lichess datasets to ensure accuracy.

Discussion

I expect to find that there may be reason to belive that the growth Niemann experience could be considered not outside the bounds of normal given the centipawn loss vs. FIDE score method and that a more in-depth analysis must be conducted using other methods of measurement. This would challenge the existing assumptions that centipawn loss is a "smoking gun" for cheating and that there are instances it is inaccurate.

Conclusion

In summary, the goal of this proposal is to analyze the use of comparing centipawn loss to FIDE rating as a metric to detect cheating in chess, and whether move-for-move analysis can support the claim that Niemann likely cheated in his game against Carlsen.

References

Bhat, M. A. (2021). Online chess games. Technical report, Kaggle. https://www.kaggle.com/datasets/mysarahmadbhat/online-chess-games.

Qureshi, Z. (2022a). Hans niemann chess games (1600+ in total). Technical report, Kaggle. https://www.kaggle.com/datasets/zq1200/hans-niemann-chess-games.

Qureshi, Z. (2022b). Magnus carlsen lichess games dataset. Technical report, Kaggle. https://www.kaggle.com/datasets/zq1200/magnus-carlsen-lichess-games-dataset.

Regan, K. W., B. Macieja, and G. M. Haworth (2011). Understanding distributions of chess performances. In *Advances in computer games*, pp. 230–243. Springer.