A Machine Learning Approach to Building a Chess Engine

Daniel Cloutier   Sagar Patel

December 14, 2020

Contents

Abstract

In the current era, high volumes of data are being collected at an incredible velocity. Much of this data is embedded with valuable knowledge. [1] One example of the types of data that is collected is chess data. In fact, lichess.org, one of many websites where chess can be played online against opponents from around the world, contains a data dump of over one and a half billion games, seventy eight million of

them being played in the month of November 2020 alone [3]. We intend to present an automatic learning method for chess, utilizing deep neural networks. We created our deep neural network using a combination of both unsupervised and supervised training. Employing unsupervised training, our engine pre-trains to identify and extract high-level features that exist in a dataset of chess board positions. The supervised training compares two chess positions and evaluates the board which is more favourable, from the white player's perspective. Therefore, we create a deep neural network that can understand chess without incorporating the rules of the game and using no prior manually extracted features. Instead, the system is trained from end to end on a large dataset of chess positions. The resulting deep neural network should allow for simulations that extend the possibilities of chess strategies, highly valuable to chess players at all levels.

1   Introduction

The game of chess is a very popular board game and has become a well liked study for computer enthusiasts, especially in the domain of Artificial Intelligence, from the creation of Deep Blue in 1996 which defeated the world chess champion at the time, Garry Kasparov, up to modern data analysis techniques being tested on chess games as a show for it is possible uses. Chess has become a good starting point for both more advanced software developers attempting to test something innovative or for newcomers trying to increase their knowledge.

One of the many reasons for the game's popularity is that chess data is collected at an incredibly high volume and velocity. Not only this, but the data collected is often readily available to the public. Databases such as chessbase and lichess contain thousands, if not millions of games in a standardized notation format for chess called Portable Game Notation. This makes the game of chess a perfect starting point for new tools in Data Science to be tested.

Another of the many domains of study that has been recently using chess as a means of improving itself is that of Machine Learning. Namely, with the advent of algorithms

such as AlphaZero [5], some new ideas have risen in an attempt to create algorithms that not only play chess better than humans, but even some that try and mimic human play [4].

Keeping all of this in mind, we wanted to take advantage of the very large amount of data stored by lichess in order to implement and perhaps try and reinvent some already existing ideas using this data that, as far as we could tell, has not been used much in recently published material.

2   Review of Research and Ideas

2.1   Chess Game Categorization

2.2   Predicting Game Results

2.3   Creating a Chess Engine

3   Problem Statement

In this era, there is a massive amount of data collected in a large variety of different topics, of which lots of information can be extracted. The issue here comes with the fact that we cannot simply extract the information by visual inspection. There is currently too much data.

One such domain that contains a large amount of data is the game of chess. Many players like to study chess games to try and improve their play, but it's very hard to find games that are relevant or interesting to learn from amidst the plethora of games to choose from. Although there are many famous known games such as the Opera game of Paul Morphy vs Duke Karl or game 6 from the 1972 world championship match between Robert Fischer vs Boris Spassky [2], these are usually kept as beginner studies. Also, the fact that they are already well known doesn't necessarily help us discover anything new.

On top of these numerous known games, surely there are many more games that you could learn from. In fact, the website lichess.org logged 68,027,862 games at all levels

of play in September of 2020 alone [3]. Some of these could be interesting to study, irrespective of the players ratings not being at the level of Grandmaster. All of this to say, there must be support from algorithms or machine learning techniques to be able to group similar games of interest together, or show games that lie far outside the norm, which could be interesting studies.

4 Our Approach/Model

5 Algorithm and Discussion

6 Implementation

7 Conclusions

References

[1] J. A. Brown et al. "A Machine Learning Tool for Supporting Advanced Knowledge Discovery from Chess Game Data". In: *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*. 2017, pp. 649–654. DOI: `10.1109/ICMLA.2017.00-87`.

[2] Timothy Glenn Forney. *Best Chess Games of All Time*. URL: `https://www.chessgames.com/perl/chesscollection?cid=1001601`.

[3] *lichess.org game database*. URL: `https://database.lichess.org/`. (accessed: 22.10.2020).

[4] Reid McIlroy-Young et al. "Aligning Superhuman AI with Human Behavior: Chess as a Model System". In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '20. Virtual Event, CA, USA: Association for Computing Machinery, 2020, pp. 1677–1687. ISBN: 9781450379984. DOI: `10.1145/3394486.3403219`. URL: `https://doi.org/10.1145/3394486.3403219`.

[5] David Silver et al. *Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm*. 2017. arXiv: `1712.01815 [cs.AI]`.