CS 234
Project Proposal

A General RL-Based Approach for Board Games

Siyu Liu(siyuliu3)
Ace Hu(acehu)
Yanpei Tian(yanpeit)

Project Mentor:Rohan Badlani

Motivation and Introduction

In October 2017, AlphaGo Zero was introduced by DeepMind [1], becoming the best Go player in the world. Later on, the algorithm in the original program was generalized to play the game Chess and Shogi [2], indicating the generalizability of the reinforcement learning algorithm used in the original version of AlphaGo Zero. In this project, we want to explore two specific settings of Board Games, namely Gomoku and Chinese Chess, by using a reinforcement learning algorithm based on self-play.

Both of the game Gomoku and Chinese Chess fall into the general category of Board Game but they have their specific rules. By generalizing a single reinforcement learning framework into these two different settings, we can demonstrate a unified way to solve Board Games.

Problem Setup

In the first part of the project, we plan to reproduce the result obtained in AlphaGo Zero. After getting a good understanding of the model and algorithm used in the original Go player, we want to generalize the general framework into other board games, potentially Gomoku and Chinese Chess.

Gomoku, also called Five in a Row, is an abstract strategy board game. It is traditionally played with Go pieces (black and white stones) on a Go board. It can be played using the 15×15 board or the 19×19 board. [3]

Chinese chess, is a strategy board game for two players. The game represents a battle between the two armies, with the object of capturing the enemy's general. It is one of the most popular board games in China, and is in the same family as International Chess [4].

Literature Review

We went over previous research on AlphaGo, AlphaGo Zero[1] and AlphaZero[2]. In general, AlphaGo made use of supervised learning from human expert moves and reinforcement learning from self-play to train the neural network. In the game, Monte Carlo Tree Search (MCTS) is implemented to search for the optimal action on the next move. While AlphaGo Zero got rid of restriction from supervised learning by only learning from self-play, AlphaZero is a more generalized version which has predetermined search hyperparameter. As a result, AlphaZero is able to adapt to play more games.

Methodology and Algorithms

- Training

Traditional board game engines usually rely heavily on rules and heuristics handcrafted by human players. However, such human guidance can put constraints on the model. To create a creative engine unconstrained by the norms of human play, we plan to use neural network and do self-play from scratch.

Starting from an untrained neural network, the agent will play randomly against itself and learn the mechanism of the game gradually by adjusting its parameters according to the outcomes of each movement. During the training, the neural network will be used to guide a search algorithm to select the most promising moves according to the current setup of the game.

- Simulation

During training, we can monitor the performance change of our model by playing against an "earlier self". For example, after 1M iterations of training, we can save the model as a reference. Performance improvement due to further training can be gauged by playing the latest version against the saved model.

Evaluation Metrics

Qualitatively, we can evaluate our model via human-computer plays or self-plays simply by computing the win rates and tabulating the results. Quantitatively, we could evaluate the performance of our agent/model using Elo rating[5] using Whole-History Rating(WHR) algorithm[6]. Specifically, we can use Bayeselo program[7] to evaluate our model's elo score. Bayeselo may be considered as a special case of WHR, with $w^2 = 0$, or a special case of decayed history, with an infinitely long decay.

References

[1]“AlphaGo Zero: Starting from Scratch.” *Deepmind*, deepmind.com/blog/article/alphago-zero-starting-scratch.

[2] “AlphaZero: Shedding New Light on the Grand Games of Chess, Shogi and Go.” *Deepmind*, deepmind.com/blog/article/alphazero-shedding-new-light-grand-games-chess-shogi-and-go.

[3] “Gomoku.” *Wikipedia*, Wikimedia Foundation, 30 Jan. 2020, en.wikipedia.org/wiki/Gomoku.

[4] “Xiangqi.” *Wikipedia*, Wikimedia Foundation, 8 Jan. 2020, en.wikipedia.org/wiki/Xiangqi.

[5]“Elo Rating System.” Wikipedia, Wikimedia Foundation, 2 Feb. 2020, en.wikipedia.org/wiki/Elo_rating_system.

[6] Coulom, Rémi. "Whole-history rating: A Bayesian rating system for players of time-varying strength." International Conference on Computers and Games. Springer, Berlin, Heidelberg, 2008.

[7] Rémi Coulom - Bayesian Elo Rating https://www.remi-coulom.fr/Bayesian-Elo/