COMP 579 W2025 Project Proposal: Replicating the AlphaGo Zero Algorithm for Application to Connect Four

Geneviève Larose-de Billy **Isaac Simoneau** **Melissa Katz**
Group 45

1 Motivation

In one of the most notable artificial intelligence displays of all time, AlphaGo (1), a computer program developed by DeepMind Technologies in 2015, defeated Lee Sedol, the Go world champion. Its successor, AlphaGo Zero, was both more powerful and advanced, learning entirely from self-play without human game data. The goal of our project is to understand what made AlphaGo Zero so successful by replicating its architecture on a smaller scale. We aim to build a program capable of playing Connect Four, trained solely by playing against itself. Since Connect Four has only 7 possible moves at each state, we are aware that Monte Carlo Tree Search (MCTS) may not be the best approach. However, we aim to accurately replicate the move-selection techniques used in AlphaGo Zero.

2 Method

We will use Python to build a simple Connect Four simulator to display the 7x6 board, where each square can be blank, red, or black. Just like AlphaGo Zero, our agent will use a CNN with the current board representation of the game (42 slots) as input, which outputs a policy vector over all possible next moves, as well as the current predicted outcome. In addition, it will use the MCTS for action selection. Each node of the tree represents a board state, and contains the action-values, the prior function and the actions' visit counts at the given state. These values are updated as the agent learns, and are leveraged for its decision-making. In order to perform self-play successfully, we will train the model on multiple older versions of itself to prevent it from staying stuck on suboptimal policies. We will be running the experiments on our personal computers.

3 Evaluation

We plan to assess our model's performance first by having it play against an opponent acting randomly. At the end of the training, we will measure its success rate against human players. To go beyond the reproduction of the algorithm, we will also explore the effect of Monte Carlo search depth and CNN architecture on the models by pitting fully trained models with different hyperparameter values against each other.

References

[1] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, and D. Hassabis, "Mastering the game of go without human knowledge," *Nature*, vol. 550, no. 7676, pp. 354–359, Oct 2017. [Online]. Available: https://doi.org/10.1038/nature24270