Review of "*Mastering the game of Go with deep neural networks and tree search*"

Goal

The goal of this publication is to develop an artificial intelligence framework that can play Go. Go is a classic game that has a board of 19-by-19 which means a very large search space. Therefore, an efficient way to prune branches and evaluate board state is necessary.

Techniques Used

In general, this paper combines Monte Carlo tree search(MCTS) and deep neural networks to next move in the Go game. In order to reduce the depth and breadth of the search tree, this paper trained a value network to evaluating various positions and a policy network to sample actions.

The policy network is trained to predict expert moves: given current state of the board, output of policy network is a softmax layer that represent a probability distribution over all legal next moves. This probability distribution is used as a prior distribution when evaluate a potential move.

This policy network is improved by a reinforcement learning. By repeatedly play the game, the reinforcement learning updates the network parameters by stochastic gradient ascent towards the direction that maximizes expected outcome. This reinforcement learning process also generate big amount of new self-play data that can be used to train the value network.

The value network is trained on game states together with their outcome under a certain policy network. The output of the value network is just single value representing the likelihood one player can win given current state.

Both policy and value network are used in MCTS to help evaluate a certain move. MCTS randomly pick up moves according to the action value, visit count, and prior probability of each move. For each randomly picked step, AlphaGo evaluate it combination of value network and outcome of random rollout play supported by a fast policy network. The ultimately selected move is the most visited one.

Results

AlphaGo have two versions, one standard version and one distributed version. The standard version dominated all other Go programs with a 99.8% winning rate. It also won 77%, 86%, and 99% against Crazy Stone, Zen, and Pachi in a four handicap stones.

The distributed version of AlphaGo won the match 5 games to 0 against Fan Hui, a professional 2 dan Go player. This is the first time a computer program defeat a human professional player.