Research Review

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

Michael Richardson

Overview

AlphaGo has used a combination of deep neural networks to achieve a 99.8% winning rate and beat the European Go champion. AlphaGo uses value networks to evaluate the board positions and a policy networks to select moves. The deep neural networks are trained using supervised learning and reinforcement learning.

Implementation

The supervised learning stage consists of a 13 layer Policy Network. It has been trained on randomly sampled state-action pairs from 30 million positions from the KGS Go server.

The reinforcement learning stage improved the policy network by using policy gradient reinforcement learning. Games were played between the current policy network and a randomly selected previous iteration of the policy network. The randomization from a pool of opponents stabilizes the training prevents overfitting of the current policy.

The reinforcement learning for the value network focuses on position evaluation, estimating the value function that predicts the outcome from a position of games played by using a policy for both players. This neural network outputs a single prediction instead of a probability distribution.

The last stage combines the policy and value networks in an MCTS algorithm that selects action by a lookahead search.

Results

The results from the tournament show that the single machine AlphaGo is many dan ranks stronger than any previous Go program, winning 494 out of 495 games. AlphaGo also won 77%, 86% and 99% of handicap games against Crazy Stone, Zen and Pachi. The distributed version of AlphaGo was even better, winning 77% of games against the single machine AlphaGo and 100% of games against other machines. AlphaGo beat the European Go champion 5 games to 0. This was the first time a Go program had beaten a professional human Go player.
AlphaGo demonstrated that using value networks is a viable alternative for Monte Carlo evaluation in Go.