Chris Morse
August 2, 2017

Research Review

Summary of AlphaGo

Goals

The goal of the team at DeepMind was to use the latest in artificial intelligence to master the game of Go. Go is one of the most challenging of the "games of perfect information" for artificial intelligence because of its complexity. It is significantly more complex then Chess because of the number of legal moves per position and its game length. Exhaustive search is not feasible because of this complexity.

Implementation

The general principles to deal with the complexity revolve around reducing the depth of the search and by reducing the breadth of the search. Deep convolutional neural networks are used to reduce the effective depth and breadth of the search tree.

AlphaGo used Monte Carlo tree search (MCTS) combined with policy and value networks. 3 stages were used to train the neural networks. 1) A supervised learning of expert human moves is used to train the policy network. 2) This network is then improved by using a reinforcement learning policy network. 3) They train a value network to predict the winner of games played against itself. The MCTS efficiently combines the policy and value networks.

Supervised learning

The first stage in the training pipeline used supervised learning. A 13 layer policy network and 30 million positions was used to predict human expert moves. The output is a probability distribution over all legal moves.

Reinforcement learning

The second stage in the training pipeline used reinforcement learning to improve the results of the first stage. Games are played between the current policy network and a random previous selected policy network.

Value networks

The final stage in the training pipeline focuses on estimating a value function to predict the outcome. This is estimating the probability that the current move leads to a win. They estimate the value function for the stronger policy network. This neural network has a similar structure to the policy network, but outputs a single prediction instead of a probability distribution.

MCTS

AlphaGo combines the policy and value networks in an MCTS algorithm that uses lookahead search to select actions.

Results

The results were very positive. Using 5 seconds per move, the single machine AlphaGo achieved a 99.8% winning rate against other Go programs. The distributed version of AlphaGo was significantly stronger than the single-machine version. The distributed version defeated the human European Go champion by 5 games to 0. The is the first time that a computer has beaten a professional in Go. Most people thought this would be at least a decade away.