Research review of AlphaGo

The paper explains how they combine three different elements: Monte Carlo tree search, value networks and policy networks to create an agent that learns how to play Go.

Value Network

The team behind AlphaGo created a convolutional neural network, that takes as input the state of the game and outputs a single value that represents which player has the advantage in the given state. They didn't train on full games samples because this method caused overfitting, instead, they sampled 30 million states drawn from 30 million different games, in this way the network was much better at generalizing.

Policy Network

This is a neural network that given a state of the game, outputs the probability of each possible move being taken by a human player. They used two different training methods, first, supervised learning with a sample of 30 million positions from the KGS Go Server, this resulted in the network choosing the right action around 57% of the times. Then, in a second iteration they improved the accuracy of the policy network by using reinforcement learning, the agent would play against a random older version of itself, in order to prevent overfitting of the current policy.

Monte Carlo Tree Search

This method is used in combination with a quick evaluation function to explore relevant branches of the graph and use the previous networks in order to assess and chose what path follow and what moves to make.

Results

Alpha Go defeated consistently all other AI Go playing agents, it even defeated the human European Go Champion by 5 games to 0. Recently during late May 2017, it has keeping defeating world class human Go players at The Future of Go Summit in China.