

Alpha Go Paper Review

Alpha Go is a narrow AI Software developed by Google Deepmind Team. It doesn't use traditional AI methods to play because those methods simply don't have a change against the complexity of the game of go. New search and learning methods combined with a huge amount of training were applied to crack Go.

Monte Carlo Tree Search

This method uses random sampling of the search space to deduce what are the most promising moves. Alpha Go traversed the tree selecting the edge with the best value, calculated using the max action value plus a probability stored on that edge.

Deep Neural Networks

Alpha Go uses at least 12 different layers, each of these networks receive a board description as an input. This information is processed using two different types of networks. One is the **Policy Network**, which decides the next move, and the other is the **Value Network**, which determines the winner of the game.

Training

At first, Alpha Go trained with data from games played by top human players, being able to predict the human moves 57% of the time, then it started to play against multiple instances of itself, this was because researchers wanted Alpha Go to develop its own strategies instead of just “copying” humans. Some of those instances were a single computer and others were distributed computers. This learning process was done using **Supervised Learning** and **Reinforcement Learning**. RL is a technique inspired on behaviorist psychology that basically learns by trial-and-error solely from rewards or punishments.

Policy Network

In the first stages of the training process this network was applied using supervised learning. In the second stage of the training process reinforcement learning was applied. Its structure is identical to the SL policy network, however results were 80% better in games against the SL policy network.

Value Network

The last stages of the training process focused on estimating a value function that predicts the outcome from position of games played. Although this network is similar to the PN, it doesn't output a probability distribution. Instead it returns a single prediction value, generated computing the outcome of games played with a policy p for both players.

Results

First, Alpha Go reached a 99.8% winning rate against other Go programs, then went to play with the European Champion Fan Hui, winning by 5 – 0. At last Alpha Go was able to beat top go player Lee Sedol by 4 – 1 becoming the first AI program that beat a human go player without any handicap.