Research review of the Deepmind paper about AlphaGo

Abstract

Deepmind has created a very successful game of Go playing agent called AlphaGo, that plays at superhuman level.
The program uses a combination of supervised learning and deep reinforcement learning with Monte Carlo Search Tree (MCTS).

Earlier implementations of a Go playing agent has used the MCTS algorithm and a linear combination of input features.
In this paper, Deepmind states that by using deep neural networks in a pipeline and combine them with MCTS, a playing agent is able to beat all previous implementations of a Go playing agents with a 99.8% winning rate.

The techniques used will summarized in this short resumé.

Techniques

AlphaGo has three deep convolutional networks. There are two policy networks, one Supervised Learning policy network (SL policy) and a Reinforcement Learning policy (RL policy). Then there is a Reinforcement Learning Value network that estimates a value function for position evaluation. The neural networks is created with 13 layers.

The SL policy network is trained with 30 million positions from the KGS Go server and then the weights are used to initialize the RL policy network which has the same architecture. Then the RL policy network plays against itself with current RL policy network playing against a randomly picked previous version of the network. The role of the RL policy network is to produce a probability distribution of the current legal moves.
The value network has a similar architecture as the policy network but outputs only a single prediction and not a probability distribution. It is trained by creating a new dataset with 30 million distinct positions, that is, they are not interdependent, by playing against the RL policy network.

When playing, an image of 19X19X48 image stack consisting of 48 feature planes is sent to the SL policy network which sends a policy gradient to the RL policy network. This sends a probability distribution to the value network that predicts the expected outcome. All this is run in a MCTS search for every node. At the end of the search AlphaGo selects the action with maximum visit count.

Results

The paper concludes that combining a MCTS with neural networks are able to achieve a performance that is much stronger than seen before with a Go playing agent.
It is the first Go playing agent to beat a professional human Go player in a formal 5 game match. AlphaGo won 5-0.