Research Review: AlphaGo

Dave Feltenberger, July 2017

In 2016, Google DeepMind held a high-publicity event and beat a top-ranked Go player using a new AI system they call AlphaGo. This was remarkable because most experts thought it would be at least 10 more years before an AI could challenge even expert level players, let alone the world's best. But it happened. Below is a discussion of the techniques and results of this remarkable feat.

Goals

The goal of the researchers at DeepMind was to build upon their work with Atari games and novel use of reinforcement learning (for self-play) to be the best Go AI in the world. They had previous success beating any Atari game by simply looking at the pixels and feeding the score as its loss function.

Techniques Introduced

AlphaGo uses two types of neural networks – one for learning how to evaluate board positions (called "value network") and one for learning how to best select moves ("policy network"). The policy network uses supervised learning to build 13-layer convolutional network (and ReLu activations), and interestingly uses gradient *ascent* to achieve a high loss for the human opponent player. This is the first stage of the training pipeline.

The most novel approach is the introduction of reinforcement learning (RL) for policy networks. This is the second stage of the training pipeline. Using the weights from the output of the first stage, the network plays against a randomly-selected older output from stage 1. It then builds a reinforcement policy network by giving positive reward for plays that result in a win and negative reward for plays that lead to a loss, and zero for everything else.

Finally, after building this, they combined MC tree search algorithm on the RL policy nodes to search ahead.

Results

Using this hybrid approach, with supervised neural networks and reinforcement learning, DeepMind was able to achieve state-of-the-art and beat the European (and eventually World) Champion Go player. Using RL alone beat state-of-the-art supervised network 80% of the time. Using no move search at all, the RL system alone beat the previous world's-best AI Go program 85% of the time! Combining supervised convolutional neural networks, RL, and tree search on the RL policy edges results in the best Go-playing intelligence (artificial or human!) ever created.