Udacity AIND Project II - Research Review

Rongyu Lin

Selected Paper: Mastering the game of Go with deep neural networks and tree search

In this paper, DeepMind team (DMT) determines to create a new approach that evaluates much faster and has higher winning rate against both other Go programs and human Go champion. Search sapces of games like Go are mainly effected by 2 factors – depth and breadth. In order to reduce both of them, DMT uses deep convolutional neural networks, in which value network is used to evaluate positions and policy network is used to sample actions. At last, their program, the famous AlphaGo, combines policy and value networks with Monte Carlo tree search (MCTS).

DMT train the neural networks using a pipeline consisting of 3 stages of machine learning. In the first stage, DMT build on prior work on predicting expert moves in Go using supervised learning. They trained a 13-layer policy network, called SL policy network. In the second stage, DMT aim at improving SL policy network by policy gradient reinforcement learning. Outcome of this stage is called RL policy network. In final stage, DMT focus on position evaluation. Neural network in this stage (value network) has similar architecture to policy network, yet outputs a single prediction instead of a probability distribution.

At last, DMT combine MCTS with deep neural networks mentioned above in AlphaGo, which uses an asynchronous multi-threaded search that executes simulations on CPUs and computes policy and value networks in parallel on GPUs. AlphaGo can also be implemented in a distributed version.

To evaluate the playing strength of AlphaGo, DMT first ran an internal tournament among variants of AlphaGo and other Go programs including Crazy Stone, Zen, Pachi , GnuGo and Fuego. In the tournament, single-machine AlphaGo won 494 out of 495 games against other Go programs, and even won 77%, 86% and 99% of handicap games against Crazy Stone, Zen and Pachi. The distributed version of AlphaGo was even much stronger, winning 77% of games against single-machine AlphaGo and 100% against other programs. Finally, distributed version of AlphaGo was evaluated against Fan Hui, European Go Champion. During the match, AlphaGo evaluated thousands of times fewer positions than Deep Blue did in its chess match against Kasparov, and won the match 5 games to 0, which was the first time that a computer Go program had defeated a human professional player.