A review of the AlphaGo game-playing agent

Here I introduce a short summary of the paper of AlphaGo game-playing agent that has been developed by the DeepMind team. AlphaGo is a computer program that has an only purpose – play the board game Go. It has been first introduced in 2015 and beaten the best player of Go game in 2016.

Goals & Techniques

The main goal, of course, was a development of a Go game-playing agent that beats **professional** Go players and existing Go game agents, having reasonable response time for the next move.

Since traditional approaches such as Monte Carlo Tree Search cannot give sufficient results on a **professional** level of Go, the DeepMind team wanted to prove that deep learning approach could solve that problem.

AlphaGo uses a **value network** that valuates the current position and a **policy network** for selecting a next move. Both networks uses convolutional layers and consider a board as 19x19 image. First, DeepMind team trained a policy network in 2 steps:

- Supervised learning (SL) using 30 millions positions from the KGS Go Server to predict expert moves;
- Reinforcement learning (RL) where the current policy network played with a randomly selected previous iteration of the network, i.e. the agent played against himself.

The last preparation step was training of a value network using RL. To avoid overfitting they generated a data set of 30 million positions sampled from a different game. Each game was played between RL policy network and itself.

At the end they combined both network in Monte Carlo Tree Search algorithm, where valuation network and Monte Carlo rollouts are used together for a position valuation.

Results

For evaluation of AlphaGo, the DeepMind team ran an internal tournament among variants of AlphaGo and strongest commercial and open source Go programs. All programs were allowed 5 s. of computation time per move.

Single-machine AlphaGo was much stronger than other Go programs, winning 494 of 495 (**99,8%**) of games. The **distributed version of AlphaGo** was much stronger, winning **77%** of games against single-machine AlphaGo and **100%** of games against other Go programs.

Another interesting output was that with **only using a value network** for a position valuation AlphaGo exceed performance of all other Go programs, though the mixed valuation still performed better.

The last result – **distributed version of AlphaGo** won 5 of 5 games against Fan Hui, the winner of the 2013, 2014 and 2015 European Go championships.