Research Review – AlphaGo

1. A brief summary of the paper's goals or techniques introduced

Goal

By using the new search algorithm that combines Monte Carlo simulation with value and policy networks and deep learning technique, AlphaGo is trained to easily win the game of Go, even facing top human experts.

Techniques

- Monte Carlo Tree Search(MCTS)

Monte Carlo tree search (MCTS) uses Monte Carlo rollouts to estimate the value of each state in a search tree. By simulation, it can improve the search tree to be larger and more accurate.

It will be combined with the latter networks by `Selection`, `Expansion`, `Evaluation` and `Backup`.

- Supervised Learning(SL) Policy Network

Here the paper use 2 SL policy network using convolution neural network.

The first p_σ is trained directly from expert human moves so it can be trained efficiently. It's used to predict human expert moves in a data set of positions.

The second p_π is called Roolout. It's smaller than the previous so it can be used to make a quick and rough prediction.

- Reinforcement Learning(RL)

Reinforcement learning is used to train the policy network p_ρ and the value network v_θ.

Initiated by the SL policy gradient, the RL policy network is train by playing Go with another policy network with the SL policy gradient.

The value network is trained by regression to predict the expected outcome.

2. A brief summary of the paper's results

AlphaGo have achieved a 99.8% winning rate against other Go programs and defeated the human European Go champion by 5 games to 0.