Mastering the game of Go with deep neural networks and tree search.

July 17, 2017

1 Introduction

AlphaGo is the strongest Go player so far which defeat the professional Go player. AlphaGo defeat 3 times European champion, Mr. Fan Hui in october 2015 with 5-0 win. Also it compete with legendary Go player Mr. Lee Sedol, a winner of 18 world titles and won by 4-1.
AlphaGo combines the advanced tree search with deep neural networks[1].

Neural Network

Neural network are model adapted as like human brain where multiple neuron have connection to get complete model. The weights update to get better results by minimizing the error between output and the desired output of the network. There are two parts in neural network; policy network and value network. Policy network select the next move to play while value network predicts the winner of the game. This process of training by comparing with desired output is called supervised learning. The AlphaGo did train the policy network by taking 30 million position data from KGS Go Server in two different way; faster one with low accuracy and a slower one with high accuracy. The fast policy network achieved 24.2 % accuracy while slow achieved 57.0 % accuracy.

Reinforcement learning

The second step is learning from mistakes and increment improving itself until it become strong enough is called reinforcement learning(RL). AlphaGo applied RL which it played policy network several times and learn from mistakes. After training with RL, policy network won 80 % of the games against SL policy network.
In general, value network calculate the goodness of the strategy that were sampled by the policy network. Thus the value network controls the depth and the policy network the breadth of the search tree.

Monte Carlo Tree Search(MCTS) is the search algorithm for AlphaGo, which uses the an asynchronous version of monte carlo tree search technique. On this tree search moves are selected for expansion based on the supervised learning policy network. The resulting positions are then evaluated two ways; one with the value network and another using

a random rollout using fast rollout policy. The two values are conbined using a mixing parameter.

References

[1] D. Silver et al., Mastering the game of Go with deep neural networks and tree search, Nature, vol. 529, no. 7587, pp. 484489, 2016.