Overview

Game Playing In the AI field, Go is a very difficult problem. There are two reasons why AI is difficult to implement. First, the search space is very large. Second, it is very difficult to implement the evaluation function. Prior to alphago, Artificial Intelligence used the Monte Carlo Search Tree to implement an amateur level artificial intelligence Go. AlphaGo has achieved success in winning the world's best pros by incorporating deep learning into the existing Monte Carlo Search Tree.

Technology used in AlphaGo

Large games such as Go are not able to search end-game with bruth force. Therefore, it needs to reduce the search depth and breadth of the game tree. Alphago implemented a considerable level of evaluation function in the value network to reduce the search depth and implemented a policy network to reduce the breadth of the game tree to reduce the number of branching nodes. Monte Carlo Search Tree is applied to these two technologies to determine next move.

1) policy network

The policy network is a network that receives the current game status and outputs the probability values of the next moves. AlphaGo did not search all the legal moves in the game tree. Instead, AlphaGo runs a policy network in the current game state and limits the search breadth with only a few high-probability moves.

The authors of the paper studied the policy network in two stages. First, we learned about the next move in the supervised learning method through the notation of the pros. However, the supervised learning method alone can only mimic the next highest number of human beings. The next step is to enhance the performance of the policy network by self-learning through reinforcement learning.

2) value network

The value network is the network that receives the current game state and outputs the probability that the AlphaGo will win. The value network implemented in AlphaGo is remarkable as an evaluation function.

The authors used a policy network learned in reinforcement learning to learn the value network. They assume that the next move from the current state to the end-game is the next move in the policy network, and they set the score in the current state. They created training samples in the form of (state, score) in this way, and let AlphaGo learn the value network.

3) monte carlo tree search

The monte carlo tree search sets the scores of possible moves in the current state by repeating an experiment that randomizes the next moves up to end-game. Alphgo also uses MCTS. However, the authors of the paper did not completely randomize the next moves, but rather made it a more realistic experiment using a policy network.

4) searching with policy and value networks

Finally, alphago determines the next move by reflecting the next moves evaluated by the policy network and the next moves evaluated by the monte carlo tree search.

Performance of AlphaGo

After the paper was published, AlphaGo won 4: 1 against Lee Se-Dol, the world's best player.