A Brief Summary of “Mastering the game of Go with deep neural networks and tree search”

Techniques

In this paper,the authors employed 3 deep convolutional neural networks (CNN) for the game of GO: 2 policy networks and 1 value network, all of which takes as input the current game state, presented as an image:

1. train a supervised learning (SL) policy network p_sigma directly from expert human moves. – predict expert moves in the game of GO – the SL policy network alternates between convolutional layers with weights sigma and rectifier nonlinearilities, a fnal softmax layer outputs a prob distribution overall legal moves a. The state s to the policy network is a simple representation of the board state. The network is trained to maximize the likelihood of the human move a selected in state s

2. train a reinforcement learning (RL) policy network p_rho that improved the SL policy network by optimizing the final outcome of the games of self-play. The RL policy network is identical in structure to the SL policy network, and the weights rho are initialized to the same value sigma.

3. train a value network nv_theta that predict the outcome from position s of games played by using policy p for both players, i.e. the value network is to predict the likelihood of a win, given the current game state. This is similar to the evaluation function, but instead it is learned, not designed.

AlphaGo combines the 2 policy and 1 value networks in an MCTS algorithm that select actions by lookahead search. Each edge (s, a) of the search tree stores an action value Q(s,a) visit count N(s,a) and prior probability P(s,a), at each time step of the simulation, the action is selected

Results

Based on the Elo rating system, Distributed AlphaGo has the score of 3140 which is much greater than other high-perfromance MCTS algorithms (Highest being CrazyStone with a score of 1929). The tournament suggests that single machine AlphaGO is many dan ranks stronger than any previous Go program, with the winning rate 99.8%.

Unlike other Go program, AlphaGo combined the Deep learning with the search tree, which is the first time to include a learning components. This can be used in many other domains; for example, general game-playing, classical planning, partially observed planning, scheduling and constraint satisfaction.