Summary on AlphaGo

AlphaGo by DeepMind team

Chansung, Park - 29 November 2017

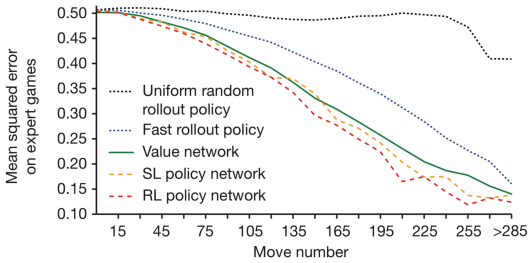
Goal

All games of perfect information may be solved by recursively computing the optimal value function in a search tree. The search tree represents b^d possible moves where b is number of legal moves per position and d is its game length. AlphaGo is built to solve the Go game efficiently which the exhaustive search is infeasible because possible number of b and d is 250 and 150 respectively. The Go game has been considered unbeatable for computer against human player.

Techniques

In this paper, the DeepMine team employed deep convolutional neural networks to reduce the effective depth and breadth of the search tree after passing board position as a 19x19 image. The neural networks was trained by using a pipeline consisting of several stages of machine learning. The first stage is supervised learning policy network which the training is directly done from expert human moves. The second stage uses reinforcement learning policy network to improve the first stage by optimizing the final outcome of games of self-play. The final stage employs value network which predicts the winner of games played in the second stage against itself.

Result



The figure on the left shows how SL/RL policy networks and value network performs better than other rollout policies.

The figure one the right depicts successfully achieved the professional level. Fan Hui is a professional human go player. The normal AlphaGo had similar performance with him, and the distributed version outperformed over him.