

Research Review

based on [Mastering the game of Go with deep neural networks and tree search](#) by the DeepMind Team

Paper's goals or techniques

The goal of this paper is to design a Go-playing agent that can even defeat a human professional player in the full-sized game of Go.

All games can be solved by two steps recursively. First search and then evaluate states. However, Go's states are too large to search and its rules are too simple but hard to evaluate where a move is good or not. This paper tries to solve these two steps by introducing a new approach - A Monte Carlo tree search method combines value and policy networks for evaluation.

This paper uses 'Monte Carlo tree search' to search.

This method rollouts the search tree but doesn't branch all actions, it only samples actions from a probability distribution over possible moves in the position which are provided by the a set of policies and averages the samples in order to select an effective move. By using these pre-trained policies, the agent no longer need to branch all and these policies does narrow the search. But the performance of the Go agent will be determined by the performance of the policies. Because of the unprecedented development of deep neural networks, these policies can be trained better and gain performance better by using deep convolutional neural networks.

So, this paper uses two deep convolutional neural networks to evaluate current state.

1.'Value netowrks' for evaluating board positions and predicting the winner.

2.'Policy networks' for selecting moves.

These networks are trained by using a pipeline mainly consisting two machine learning methods - supervised learning by expert human moves and reinforcement learning by self-play. The 'policy network' first is trained by directly from expert human moves under supervised learning method and then is optimized by reinforcement learning method. The second step is important because it adjusts the network to win rather than being correct as expert player

does. Finally, the 'value network' is trained by reinforcement learning policy network against itself.

Paper's results

The overall performance of AlphaGo is remarkable. This agent has finally reached a professional level in Go. This paper evaluates the relative strength of computer Go programs by running an internal tournament and measuring the Elo rating of each program. The tournament games were scored using Chinese rules. Under these conditions, a single-machine AlphaGo achieved a 99.8% winning rate against other Go programs including Crazy Stone, Zen and Pachi which all are based on high-performance MCTS algorithms. Further on, the distributed version of AlphaGo is expressively stronger and it has even defeated a human professional player Fan Hui the European champion of Go. This is really a feat that was previously believed to be at least a decade away.