

Research review: Mastering the game of Go with deep neural networks and tree search

For people who are not familiar with machine learning or latest technology, Google's AlphaGo is just like a superstar and it occupied the headline of almost every media in this world. It won all the games when playing with human players, who are world's best Go players. The game of Go has been considered as the most difficult classic games to develop an efficient artificial intelligence agent. As we know in a board game, in order to computing the optimal evaluation functions, it has approximately b^d possible sequences of moves. Specifically in game of Go, b and d are about 250 and 150. The computation is very large and developing an efficient algorithm has been challenging researchers for a long time.

In order to solve the problem, the researchers from this paper has trained the neural networks using a pipeline consisting of several stages of machine learning. In the first stage, they trained a supervised learning policy network by studying the moves from human experts. A 13-layer policy network was trained on randomly sampled state-action pairs, using stochastic gradient ascent to maximize the likelihood of human move. In the second stage, the researchers improved the policy network by policy gradient reinforcement learning. In the reinforcement learning, the structure is identical to supervised learning but used a different strategy to calculate the weights by maximizing the expected outcome. In the final stage, they focused on position valuation by estimating the value function for their strongest policy, using the reinforcement learning policy network. They trained a value network by regression and

used stochastic gradient descent to minimize the mean squared error between predicted value and corresponding outcome. At the end, AlphaGo combines the policy and value networks in a Monte Carlo tree search algorithm. The tree is traversed from the root state, and the traversal maybe expanded or not at a leaf node. They also developed a method to evaluate the visit count and action value.

When talking about results of AlphaGo, the most impressive thing is how it beats the world's best human players. It is so powerful that even by playing games with four handicap stones, it still has a large win rate. The AlphaGo is an art with all the hard work from the researchers and it is a sign of how powerful machine learning techniques can be. I believe in the future the machine learning techniques will be much more stronger and bring us new supervises.