Summary of Mastering the game of Go with deep neural networks and tree search

This article mainly focused on introducing a new approach to computer Go, which uses 'value networks' to evaluate board positions and 'policy networks' to select moves. The game of Go is the most challenging of classic games for artificial intelligence, considering its enormous search space and the difficulty of evaluating board positions and moves. These deep neural networks of 'value networks' and 'policy networks' were built by a new combination of supervised learning from human expert games, and reinforcement learning from games of self-play. Instead of lookahead search, the neural networks play Go at the level of stateof-th-art Monte Carlo tree look for programs that simulate plenty of random games of self-play. In addition, this article also introduced a new search algorithm that integrates Monte Carlo simulation with value and policy networks.

This article used an architecture for the game of Go, which built multiple layers of neurons, each arranged in overlapping tiles, to construct increasingly abstract, localized representations of an image. It pass in board position as a 19 x 19 image and use convolutional layers to construct a representation of the position, which reduced the effective depth and breadth of the search tree (evaluating positions using a value network, and sampling actions using a policy network).

Firstly, the author coached the neural networks using a pipeline consisting of several stages of machine learning, which began with a supervised learning policy network directly from expert human moves. Subsequently, the author also trained a fast policy $p\pi$ that can rapidly sample actions during rollouts. And then, they trained a reinforcement learning policy network that improves the supervised learning policy network by optimizing the final outcome of games of self-play, which adjusted the policy towards the correct goal of winning games, rather than maximizing predictive accuracy. At the end, they train a value network that predicts the winner of games played by the reinforcement learning policy network against itself.

In the section of the evaluation of AlphaGo, the author ran an internal tournament among variants of AlphaGo and several other Go programs. The results of the tournament suggest that single-machine AlphaGo is many dan ranks stronger than any previous Go program, winning 494 out of 495 games (99.8%) against other Go programs. They also assessed variants of AlphaGo that evaluated positions using just the value network or just rollouts. Finally, they evaluated the distributed version of AlphaGo against a humen professional player, who is the winner of the 2013, 2014 and 2015 European Go championships.

As a result of the search algorithm described above, the program AlphaGo successfully made a 99.9% wining rate against other Go programs, and completely beat the human European Go champion by 5 games to 0. Before this happened, there was no computer program that has beaten a human professional player in the full-sized game of Go, which was considered as an achievement that was previously thought to be at least a decade away.