AlphaGo: When Intuition meets Reflection

Ju Yang

The game of Go has long been a challenging classic game problem for artificial intelligence due to its search complexity and lack of well-defined evaluation function. In the past, researchers have used Monte Carlo Tree Search algorithms to select winning moves through a large number of simulations. Such tree search combined with carefully designed rules from expert knowledge has achieved strong amateur play. In the Nature paper 2016 [1], DeepMind team led by Demis Hassabis developed a new approach using neural networks, and combined neural networks with Monte Carlo simulation. The goal of the new algorithm, called AlphaGo, is to outperform human professional players.

AlphaGo represented current game state of the whole board as an image with additional features, and trained 3 convolutional neural networks: 2 policy networks and 1 value network. First, AlphaGo trained a supervised learning policy network from 30 million positions of human moves. The network predicted expert moves with an accuracy of 57%. A fast policy but less accurate rollout policy was also trained. Second, AlphaGo trained a reinforcement learning policy network by optimizing the final outcome of the game through self-play, using the outcome of these games as a reward signal. At each move, AlphaGo chose the move with the highest likelihood of winning predicted by the policy network. Using no search at all, the reinforcement learning policy network won 85% of the games against Pachi, a sophisticated Monte Carlo search algorithm. This suggests that intuition is very critical in the game of Go. Third and the final stage of the training pipeline is to train a value network to predict the likelihood of a win given current game state. AlphaGo used a mixture of the output of the value network and the result of a self-play simulation of the policy network. While the value network provides intuition, the simulation tree search provides deep reflection [2]. Both are critical for the game of Go.

Combining expert knowledge, neural network, and tree search, AlphaGo achieved 99.8% winning rate against other Go programs, and defeated a human professional Go player in the full-sized game of Go.

Reference

[1] Silver, D., et al., ***Mastering the game of Go with deep neural networks and tree search.*** Nature, 2016. **529**(7587): p. 484-489.

[2] Christopher Burger, ***Google DeepMind's AlphaGo: How it works***, https://www.tastehit.com/blog/google-deepmind-alphago-how-it-works/