Research Review of AlphaGo by the DeepMind Team

Natalie Young

This article introduced us a new approach to computer Go that uses ‘value networks’ to evaluate board positions and ‘policy networks’ to select moves, and also a new search algorithm that combines Monte Carlo simulation with value and policy networks. These deep neural networks are trained by a novel combination of supervised learning from human expert games, and reinforcement learning from games of self-paly. First stage of the training pipeline is to build on prior work on predicting experts’ moves in the game of Go using supervised learning. Second stage is to improve the policy network by policy gradient reinforcement learning. The final stage is position evaluation, estimating a value function that predicts the outcome from a certain game state by using policy for both players. Then AlphaGo combines the policy and value networks in an MCTS algorithm that selects actions by lookahead search.

To evaluate, internal tournaments was ran among variants of AlphaGo and several other Go programs. The results indicate that single machine AlphaGo is much stronger than any previous Go program. The mixed evaluation ($\lambda=0.5$) performed best, winning ≥95% of games against other variants. AlphaGo has reached a professional level in Go by combining tree search with policy and value networks.