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

Ryan Shrott

The seminal paper introduces a novel approach to using value networks to evaluate board positions and policy networks to select moves. The neural networks are trained using human expert input and reinforcement learning algorithms. The search algorithm used combines value networks and policy networks with MC simulation. It wins 99.8% of games against other Go bots.

Optimal value functions are computationally expensive in complex games. In Go, there are 250^150 nodes on the search tree. We can reduce the depth of the tree by truncating prematurely and using an approximation to the value function. The breadth of the search can be reduced by using MC simulation on a policy p(a|s). Averaging over the rollouts gives an effective position evaluation. A value network is trained to predict the winner of games played by the RL policy network. The program efficiently combines policy and value networks with MCTS.

The trained policy network outputs a probability distribution over all legal moves. The value function must be estimated using the RL policy network and by taking an average over all sampled outcomes. Each edge of the search tree stores an action value, a visit count and a prior probability. Once the search tree is complete, the algorithm chooses the most visited move from the root position.

Evaluating policy and value networks requires a much higher level of computation than traditional search tree schemes. In order to efficiently combine MCTS with deep neural networks, AlphaGo uses a multithreaded search on a CPU and computes policy and value networks on a GPU.

AlphaGo was evaluated against many state of the art Go algorithms. It won 99.8% of the games played. Even with a handicap it won 77% of the games. A mixed valuation function performed the best. It also beat Fan Hui 5-0 in a series. He is considered the strongest human Go player in the world. This is a considerable achievement in AI research history.

This paper has achieved one of the grandest challenges in AI. The move selection process based on deep neural networks were trained on a novel combination of supervised and reinforcement learning. The search algorithm is able to effectively combine neural networks with Monte Carlo simulation. It evaluates much less moves than Deep Blue (IBM) and the software architecture is much closer to the way that a human would think. The paper provides hope that human level performance is achievable for AI in seemingly intractable areas with the use of deep neural networks.