

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

The goal of the paper was to create a Go playing agent that exceeded previous attempts using a novel approach that combined value networks to evaluate board positions and policy networks to select moves. Rather than try to examine an exhaustive number of moves, which is infeasible in Go, AlphaGo instead reduces the effective search space by selecting positions more intelligently. For example, during the match against Fan Hui, AlphaGo evaluated thousands of times fewer positions than Deep Blue did in its chess match against Kasparov.

To accomplish their goals, the AlphaGo's creators implemented a new search algorithm that combines neural network evaluations with Monte Carlo rollouts. During the course of this research, it was also discovered that a supervised learning policy network performed better than a reinforced learning policy network. This is possibly because humans select many promising moves, whereas reinforced learning picks a single best move. Along with a mixed implementation, there were variants of AlphaGo assessed that contained just a value network or just rollouts. It was determined that even without rollouts, the implementation of AlphaGo exceeded the performance of all other Go programs.

The implementation of AlphaGo was also interesting, it uses an asynchronous multi-threaded search that executes simulations on CPUs and computes policy and value networks in parallel on GPUs. The final version ended up using 40 search threads on 48 CPUs and 8 GPUs. There was also a distributed version of AlphaGo implemented which utilized even more resources; 1,202 CPUs and 176 GPUs.

As a result of these techniques, the researchers were able to develop a Go program, based on a combination of deep neural networks and tree search, that plays at the level of the strongest human players and was able to win against previous Go playing programs. AlphaGo is many dan ranks stronger than any previous Go program, winning 494 out of 495 games against other Go programs. AlphaGo's neural networks are trained directly from gameplay, as opposed to using a handcrafted evaluation function like Deep Blue.