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

Research Review:

The game of Go has long been viewed as the most challenging of classic games for artificial intelligence owing to its enormous search space and the difficulty of evaluating board positions and moves. This paper introduces a new approach to computer Go that uses ‘value networks’ to evaluate board positions and ‘policy networks’ to select moves. These deep neural networks are trained by a novel combination of supervised learning from human expert games, and reinforcement learning from games of self-play. Without any lookahead search, the neural networks play Go at the level of stateof-the-art Monte Carlo tree search programs that simulate thousands of random games of self-play. This paper also introduces a new search algorithm that combines Monte Carlo simulation with value and policy networks. Using this search algorithm, AlphaGo program achieved a 99.8% winning rate against other Go programs, and defeated the human European Go champion by 5 games to 0. This is the first time that a computer program has defeated a human professional player in the full-sized game of Go, a feat previously thought to be at least a decade away.

They also tested against the strongest open-source Go program, Pachi, a sophisticated Monte Carlo search program, ranked at 2 amateur *dan* on KGS, that executes 100,000 simulations per move. Using no search at all, the RL policy network won 85% of games against Pachi. In comparison, the previous state-of-the-art, based only on supervised learning of convolutional networks, won 11% of games against Pachi and 12% against a slightly weaker program, Fuego.In addition, they included the open source program GnuGo, a Go program using state-of-the-art search methods that preceded MCTS.

In this work they have developed a Go program, based on a combination of deep neural networks and tree search, that plays at the level of the strongest human players, thereby achieving one of artificial intelligence’s “grand challenges”. They have developed, for the first time, effective move selection and position evaluation functions for Go, based on deep neural networks that are trained by a novel combination of supervised and reinforcement learning. They have introduced a new search algorithm that successfully combines neural network evaluations with Monte Carlo rollouts. Their program AlphaGo integrates these components together, at scale, in a high-performance tree search engine

During the match against Fan Hui, AlphaGo evaluated thousands of times fewer positions than Deep Blue did in its chess match against Kasparov; compensating by selecting those positions more intelligently, using the policy network, and evaluating them more precisely, using the value network—an approach that is perhaps closer to how humans play. Furthermore, while Deep Blue relied on a handcrafted evaluation function, the neural networks of AlphaGo are trained directly from gameplay purely through general-purpose supervised and reinforcement learning methods.