

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

This research article introduces a new approach to winning Go by using Value networks to evaluate board positions and Policy Network to select moves. Go is a challenging AI game because it involves enormous search space and high complexity to evaluate board positions and moves. The article discusses about the authors' program AlphaGo that achieved 99.8% winning rate against other Go programs and defeated the human European champion by 5 games to 0. This is the first time a computer program defeated a professional human player in the full-sized game of Go.

Earlier strong Go programs were based on Monte Carlo tree search which uses Monte Carlo rollouts to estimate the value of each state in a search tree. The earlier programs were enhanced by policies that are trained to predict human expert moves. But most of these policies were shallow and limited to linear combinations of inputs.

Alpha Go takes advantage of the recent developments in Deep Learning Visual image processing. It passes the board position as 19x19 image and uses Convolutional layers to represent the position. Neural networks reduce the effective depth and breadth of the search tree. First, a policy network is trained using Supervised learning directly from expert human moves for quicker feedback and fast learning. Then they train a fast policy that can rapidly sample actions during rollouts. Then a reinforcement learning policy network is trained that improves the supervised policy network by optimizing the outcome of games of self-play. Finally, they train a value network to predict the winner of games played by the reinforcement policy network against itself.

AlphaGo neatly combines the policy and value networks with the Monte Carlo Tree search. The article also describes the evaluation of the playing strength of AlphaGo. An internal tournament was held among AlphaGo and variants of AlphaGo and other Go programs including Crazy Stone, Zen, Pachi, Fuego, GnuGo. AlphaGo was clearly the winner. It won 494 out of 495 games. Even the variations in Alpha Go without rollouts exceeded the performance of all other Go programs which shows that the value networks provide a better alternative to Monte Carlo evaluation in Go. Finally, a distributed version of AlphaGo was evaluated against a human champion. AlphaGo won the match 5 games to 0.

This is the first time a Go program has defeated a human Go Champion. The deep neural networks trained by a combination of supervised and reinforcement learning played a main part in its victory. This is the first time a neural network has been successfully combined with Monte carlo rollouts. Tree search combined with policy and value networks should hopefully lead to solutions for some unsolved problems.