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

SUMMARY

A team of researchers used 'value networks' to evaluate board positions and 'policy networks' to select moves in creating an artificial intelligent agent called AlphaGo that could play the complex game of Go. 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. A new search algorithm combines Monte Carlo simulation with value and policy networks, AlphaGo achieved a 99.8% winning rate against other Go players.

AlphaGo treats the game board of Go as a 19 x 19 image. First, a supervised learning (SL) policy network is trained directly from expert human moves. Then, a reinforcement learning (RL) policy network is trained to improve SL policy network to adjust policy towards the goal of winning games rather than maximizing predictive accuracy. Finally, a value network is trained such that it predicts the winner of the games played by the RL policy network against itself. These trained policy and value networks are then combined with Monte Carlo Tree Search (MCTS) to estimate the value of each state in a search tree.

RESULTS

To evaluate AlphaGo, the researchers ran an internal tournament among variants of AlphaGo and several other Go programs. It showed that a single machine AlphaGo won 99.8% against previous Go programs. Even when opponents were allowed four free moves, AlphaGo won most the games. The distributed version of AlphaGo was even stronger. Even using just the value networks, AlphaGo exceeded the performance of all other Go programs, demonstrating that value networks provide a viable alternative to MCTS only method. But, a mixed evaluation (combination of MCTS and value networks) performed the best. Lastly, the distributed version of AlphaGo played against a 3-year winner of European Go championships winning a series of 5 matches 5-0.