Problem and Goal

The game Go has been the greatest challenge in classic games for AI research due to its enormous search space and difficulty of evaluating board position and moves. The strongest prior research is based on Monte Carlo tree search (MCTS), enhanced by policies trained to predict human expert moves. However, the prior work has been limited to shallow policies or value function based on linear combination of input features. It achieved weak amateur level play in Go. The goal of the paper is to introduce AlphaGo, a system which takes advantage of deep convolutional neural networks that enhances the winning rate against other Go programs and human professional players.

Techniques

The authors trained neural networks using a pipeline of several stages of machine learning. It reduces the effective depth and breadth of a search tree. The resulting system AlphaGo combined the neural networks with MCTS. There are four key techniques in AlphaGo.

1. Supervised learning of policy networks
The authors trained a supervised learning (SL) network from expert human moves. It is used to provide fast, efficient learning updates with immediate feedback and high quality gradient. The policy network can predict expert move with an accuracy of 57% using all input features. A fast but less accurate policy was also trained, which achieves accuracy of 24.2%.

2. Reinforced learning of policy networks
The authors trained a reinforcement learning (RL) policy network that optimise final outcome of the game by letting the current policy network play games with randomly selected previous iteration of the policy network. The RL network won more than 80% of games against SL network.

3. Reinforced learning of value networks
The authors trained a neural network that predicts the outcome from any state of the game. The neural network was trained using a new self-play data set consisting of 30 million distinct positions with final game results. The resulting single evaluation function approached MCTS rollouts but using 15,000 times less computation.

4. Combining policy and value networks
AlphaGo combines the policy and value networks in an MCTS algorithm that selects actions by lookahead search.

Result

The evaluation of AlphaGo involves tournaments among variants of AlphaGo and several other Go programs including CrazyStone, Zen, Pachi, Fuego and GnuGo. The result showed that single-machine AlphaGo ranked many dan ranks stronger than any previous Go program. It won 494 out of 495 games (99.8%) against other Go programs. Distributed version of AlphaGo was evaluated against Fan Hui, a professional 2 dan in October 2015. Alpha Go won the match 5 games to 0. This is the first time that a computer Go program has defeated a human professional player.