The large search space and the difficulty of evaluating board positions make the game of GO as one of the most challenging classical games for artificial intelligence. This paper introduces an AI agent named AlphaGo using deep neural network to achieve superhuman performance in the game of GO.

Implementation

This approach uses 'value networks' to evaluate board positions and 'policy networks' to select moves.

The neural network in is trained using a pipeline of several stages of machine learning.

1. Training a supervised learning policy network (SL) directly from expert human moves.
2. Training a reinforcement learning policy network (RL) that improves SL policy network that evaluates the self-play outcomes of the current game state.
3. Training a reinforcement learning value network that predicts the winner of games played by the RL policy against itself.

This program efficiently combines the value and policy networks with Monte Carlo Search tree (MCTS).

The SL policy network tries to predict human expert moves by training a 13 layer policy network from 30 million positions from the KGS Go Server. The RL policy network is initialized to the SL policy network. Then the RL policy network is improved by policy gradient obtained from the SL policy network to learn maximize the outcome against previous versions of the policy network. The RL policy network passes a probability distribution of moves to the RL value network. The value network is trained by regression to predict the expected outcome. Then the policy and value networks are combined in an MCTS algorithm that selects actions by look-ahead search.

In Monte Carlo tree search each simulation traverses the tree by selecting the edge with maximum action value plus bonus calculated using store prior probabilities. The resulting leaf nodes are expanded and the new nodes are processed once by the policy network and the output probabilities are store as prior probabilities for each action. At the end of the simulation the leaf nodes are evaluated using the value network and the outcome of a random roll out played till termination. Actions values and visit counts of all traversed edges are then updated.

Results

The AlphaGo program was able to win 99.8% of the games played against any previous Go programs in markets. AlphaGo was also able to beat the Euro Go Champion by evaluating thousands of times fewer positions than Deep Blue did in its chess match against Kasparov. Unlike Deep blue which relies on handcrafted evaluation functions AlphaGo trains directly from pure gameplay through general purpose machine learning methods.