

Research Review

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

Methodology

The AlphaGo team introduces a new approach to play Go game. New search algorithm combines Monte Carlo simulation with value and policy networks. Value networks are used to evaluate board positions and policy networks are used to select moves. To construct a representation of the position they take a state of the board as a 19×19 image and use 13-layer convolutional neural networks.

Training pipeline consisting of several stages of machine learning:

1. Training a supervised learning (SL) policy network directly from expert human moves. The SL policy network alternates between convolutional layers and rectifier nonlinearities. A final softmax layer outputs a probability distribution over all legal moves. The network predicted expert moves on a held out test set with an accuracy of 57.0% using all input features, and 55.7% using only raw board position and move history as inputs, compared to the state-of-the-art from other research groups of 44.4%.
2. Training a reinforcement learning (RL) policy network that improves the SL policy network by optimizing the final outcome of games of self-play. The RL policy network is identical in structure to the SL policy network. When played head-to-head, the RL policy network won more than 80% of games against the SL policy network.
3. Training a value network that predicts the winner of games played by the RL policy network against itself. This neural network has a similar architecture to the policy network, but outputs a single prediction instead of a probability distribution. A single evaluation approached the accuracy of Monte Carlo rollouts using the RL policy network, but using 15,000 times less computation.

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

Results

AlphaGo program achieved a 99.8% winning rate against other Go programs, and defeated the human European Go champion by 5 games to 0.