

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

Goals

The goal of this paper is to introduce techniques that can be used to reduce the search space and reduce the difficulty of evaluating board positions for an AI to play the classic game of GO.

Techniques Introduced

Below approach have been used to achieve the goals:

1. *Value networks to evaluate the board.*
2. *Policy networks to select the moves.*

The training pipeline:

1. Supervised learning of policy networks directly from human experts: The network consists of alternate layers of convolution layer and rectifiers nonlinear. The output softmax function gives a probability distribution over all the legal moves.
2. Improve the policy network by policy gradient reinforcement learning: this layer is similar to previous. Here train the network with a random opponent to prevent overfitting to the current policy.
3. Train the value network using reinforcement learning for position evaluation. This neural network is similar to policy network except that here we get a single prediction instead of a probability distribution. Here the network is made to play against itself to mitigate the issue of overfitting.

Searching:

The value network and policy network are combined using Monte Carlo tree search and next move is selected using lookahead search. Each search is optimized for a single best move.

Results

Using the above techniques the AlphaGO was able to defeat a human player for the first time in the fullsize game of GO. AlphaGO achieved a 99.8% winning rate against other AI programs. AlphaGo combines multiple neural networks with monte carlo algorithms in a scalable way.