Research Review: AlphaGo

Markus Meier

October 2, 2017

Abstract

Go seems to be a quite difficult game (b 250, d 150). There are some heuristics to truncate the search tree. However, most of these heuristics lead to amateur level game play. With AlphaGo, the power of deep learning was applied for the first time in Go. They use deep learning quite exhaustively for several purposes.

- A supervised learning (SL) policy network. A convolutional neural network that is trained on expert human moves and predicts the likelihood of different moves given the state of the board.
- A reinforcement learning (RL) policy network. It has the same architecture as the SL policy network. It is initialized with the training from the SL policy network. The training is done with games of the RL policy network against (randomly chosen) previous iterations of the network. It optimizes with respect to the game result (winning or losing).
- A value network tries to predict the outcome of games given the state of the board. Trained on games of the RL policy network against itself.

All three networks are used in a Monte Carlo Tree Search, that samples search paths in the decision tree given the predicted probabilities from the trained policy and value networks.

The required CPU/GPU power is still very high, but AlphaGo is able to compete with expert players.