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

Summary

This paper is a work by the authors to build an AI agent that plays Go. Named AlphaGo, it is an agent with capacity to play the game without look-ahead search. This is because exhaustive search is not a practical feasibility in the game of Go. Hence, there is a lot of interest in this problem. The possible sequence of moves is in the order of $\sim 250^{150}$.

Technique(s)

AlphaGo primarily uses Monte Carlo tree search (MCTS), which in turn uses the Monte Carlo rollouts to estimate the value of each state in the search tree. They feed the human games as 19x19 image per board state to the algorithms convolution layers to construct a representation. This allows AlphaGo to train on large historic dataset of Go games, in the supervised learning (SL) step. In the next step, AlphaGo uses another layer of network called Reinforcement learning (RL), which is generated by doing self-play.

The policy network has 13 layers. SL policy was trained from 30 million positions and yielded the accuracy of 57% using all input features and about 55.7% using only raw board positions and move history as inputs. This is huge compared to previous 44.4% of other research groups.

RL policy is identical in structure to SL and was initialized with the same weight values. This is where self-play comes in use. Games with the current policy were played against a randomly selected previous iteration of the policy network. The weights were updated in the direction of maximizing expected out, given stochastic gradient ascent. The RL policy network, working this way, won 80% of the games against SL policy network. Also, tested against one of the strongest open-source Go progrman – Pachi – the network won 85% of the games.

Using the two networks above, as mainly evaluating and sampling, reduced the effective depth and breadth of the search tree

Result(s)

AlphaGo won 494 out of 495 games against other Go agents.

In the handicap stones games (free moves for the opponent), AlphaGo won an average of 87% of the games.

The distributed version of AlphaGo won 100% of the games against other programs.

On October 5 to October 9, AlphaGo won the match 5-0 against Fan Hui, the Europian Go champion of 2013, 2014 and 2015.

AlphaGo proved that combining tree search with policy and value networks, AI agents can achieve human-level performance.