Karol Misiarz

11.01.2018

Go playing agent- research review

This document reviews the article “Mastering the game of Go with deep neural networks and tree search”.

The article is describing the method proposed by Google to create agent playing the game of Go – complex and abstract game with great amount of possible moves and complicated position evaluation.

Because of the size of the board and amount of possible moves, it is very important to reduce the size of search space. The breadth of search can be reduced by sampling actions using Monte Carlo algorithm instead of traversing the tree. The depth of the search can be reduced by position evaluation. Authors started with the help of convolutional neural networks when they feed the network with already played positions and master moves as a target for supervised learning. Then the agent by playing with itself, used reinforcement learning algorithms to improve that policy.

The ‘policy network’ is 13 layer deep neural network that was trained on 30 million positions from KGS Go Server. The second policy that was used - ‘rollout policy’ was about 1500 times faster but much less precise. It was used for quick estimation of the leaf nodes move quality (Q value).

The ‘value network’ was a reinforcement learning network that was used to select moves and has been trained by agent playing games between current policy network and randomly selected previous one. The reward for winning was +1 and for loosing was -1. There was no other penalty during the game.

Agent combined ‘policy network’ and ‘value network’ in Monte Carlo Tree Search algorithm. Each node of the tree stores the action value (Q), visit count, and policy. Each action is selected by sum of Q value and value ‘u’ that increases with prior policy but decays with visit count to encourage exploration. At the end, the algorithm chooses the most visited move from the root position.

Evaluation of the agent shown that ‘value network’ provides good alternative to Monte Carlo evaluation, however mixing both evaluations produced best results.