RESEARCH REVIEW

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

JULY 1, 2018

SUBMITTED BY: MOHAMED AYMAN NAGUIB

Contents

Introduction

After completing my game-playing agent which plays the game isolation on different board sizes, using various techniques which are limited depth Mini-max algorithm, alpha-beta pruning and combining that with iterative deepening, I had the chance to review a really interesting paper which is “Mastering the game of Go with deep neural networks and tree search”, which explains how did the researchers create the first computer program that has defeated a human professional champion in the full-sized game of Go.

Techniques Adapted

To start off, the main goal of many game-agents is to determine the outcome of the game depending on a specific state, by recursively computing the optimal value function in a search tree. But, for large games with a high game’s breadth (b) and game length (d) values, there is a need for reducing the search space. Such technique is Monte Carlo rollouts which in prior work it showed great results, nevertheless it only showed amateur level play in Go as policies or value function used with it were based on linear combinations of input features. However, the researchers used supervised learning (SL) to train their policy network p_σ directly from expert human moves In addition to fast training policy p_π that can act optimally during rollouts Furthermore, training a reinforcement learning (RL) policy network p_p that improves the SL policy network. Finally, training a value network v_Θ that tries to predict the winner of the games played by the RL policy network adapted earlier against itself.

Supervised learning of policy networks

They trained a 13 layer policy network, which as described earlier is the SL policy network, it basically trains from 30 million Go game positions. The policy network alternates between convolutional layers in order to predict human plays.

Reinforcement learning of policy networks

It is almost the same as the previous network, but by playing itself 1.2 million times and defeat earlier forms. At the end it keeps the weight of the winner, it becomes much more optimal and stronger.

Reinforcement learning of value networks

For this stage the main focus was to evaluate the position estimating a value function that predicts the outcome of the given positions of games played by the policy network for both players. They faced a setback as their network memorized the data set that was given to it rather than generalizing. They created a different set with 30 million distinct problems and amazingly it showed better results than the Monte Carlo rollouts as it was more accurate.

Results

The paper presents different results starting with their Go game-agent outperforms Go-playing AIs from at that moment and previous AIs also, in addition to competing against the European champion and defeating him by a huge difference. Secondly, it was shown if not all of its neural networks are being used it still performs as good as other AIs. They introduced a new search algorithm that combines Monte Carlo rollouts with deep neural networks evaluations at high performance. In the match against the European champion, Alpha go evaluated fewer positions then Deep Blue by selecting positions more intelligently.