> source: https://storage.googleapis.com/deepmind-media/alphago/AlphaGoNaturePaper.pdf

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

Introduction

The game of Go is maybe the most complex classic game: It's a game of perfect information and as such may be solved by recursively computing the optimal value function in a search tree containing approximately b power d possible sequences of moves, where b is the game’s breadth (number of legal moves per position) and d is its depth (game length). In large games, such as chess (b≈35, d≈80) and especially Go (b≈250, d≈150), exhaustive search is infeasible.

DeepMind's team has achieved a breakthrough result in this field: For the very first time, an AI has won against a professional player !

In the following chapter, I'll summarize the methods and technologies they used to achieve this amazing performance.

The concept

The main problem to solve is how to reduce breadth and depth of search tree.

We can reduce the breadth of search tree by using a policy (a probability distribution over possible action in a position) reducing the search to a probabilities. Depth of search tree is reduced by position evaluation which replaces subtree into an approximate value. Truncating the search tree at state s and replacing the subtree below s by an approximate value function v(s)≈v*(s) that predicts the outcome from state s. DeepMind's team has chosen the Monte Carlo tree search (MCTS). It is based on Monte Carlo rollouts to estimate the value of each state in a search tree. Monte Carlo rollouts maximize depth without branching.

DeepMind's team has created a suite composed by a combination of deep neural networks and tree search to achieve these objectives

1st stage: The Policy NN

Three policy networks are used:

- A supervised learning policy network
- A fast policy network
- A reinforcement learning policy network

The supervised learning (SL) policy network has been designed with expert human moves. The Reinforcement learning (RL) policy network is in the second stage. This layer is structurally identical to the SL network, its weights being initialized with the sames weights in the SL policy network. Many games are then played between the current network and randomly chosen previous iteration of the RL network (to prevent overfitting). They trained also the fast policy network, much faster but less accurate that can rapidly sample actions during rollouts.

2nd stage: Value networks NN

AlphaGo uses value networks to improve position evaluation. It is used to estimate the value function for the strongest policy derived from RL policy network. The architecture is the same as the policy network. But contrary to the policy network, it will output a single prediction. To train the weights of this neural network, they used stochastic gradient descent (minimizing the MSE between prediction and the awaited value. To prevent overfitting, they generated 30M game moves with distinct positions and played these games between RL policy network and itself.

Searching with policy and value networks

AlphaGo combines the policy and value networks in an Monte Carlo Tree Search (MCTS) algorithm that selects actions by lookahead search. Each edge is evaluated using the value network and by running a rollout to the end of the game with the fast rollout policy. At the end of simulation, the action values and visit counts of all traversed edges are updated. Once the search is complete, the algorithm chooses the most visited move from the root position.

Outcome

AlphaGO has then been tested against the best existing AI, beating them all. This mix of MCTS with supervised and reinforcement learning Convolutional Neural Networks has achieved a 99.8% winning rate against other Go programs, and defeated the human European Go champion by 5 games to 0

AlphaGO is for the first time an effective search algorithm that successfully combines neural network evaluations with Monte Carlo rollouts.

The previous major breakthrough in computer Go was MCTS which led to corresponding advances in many other domains. By combining tree search with policy and value networks, AlphaGo is a new milestone promising seminal advances in other seemingly intractable artificial intelligence domains.