Artificial Intelligence Nanodegree

Research Paper Review – AlphaGo

Ollie Graham
25th September 2017

What is AlphaGo?

AlphaGo is the first computer program to defeat a professional human Go player, the first program to defeat a Go world champion, and arguably the strongest Go player in history.

AlphaGo's first formal match was against the reigning 3-times European Champion, Mr Fan Hui, in October 2015. Its 5-0 win was the first ever against a Go professional, and the results were published in full technical detail in the international journal, **Nature**. AlphaGo then went on to compete against legendary player Mr Lee Sedol, winner of 18 world titles and widely considered to be the greatest player of the past decade.

AlphaGo's 4-1 victory in Seoul, South Korea, in March 2016 (which took place after this paper was written) was watched by over 200 million people worldwide. It was a landmark achievement that experts agreed was a decade ahead of its time, and earned AlphaGo a 9 dan professional ranking (the highest certification) - the first time a computer Go player had ever received the accolade.

Reference: https://deepmind.com/research/alphago/

What is Go?

The game of **Go** originated in China 3,000 years ago. The rules of the game are simple: players take turns to place black or white stones on a board, trying to capture the opponent's stones or surround empty space to make points of territory. As simple as the rules are, Go is a game of profound complexity. There are an astonishing 10 to the power of 170 possible board configurations - more than the number of atoms in the known universe - making Go a **googol** times more complex than Chess.

What were the new methods?

All games of perfect information have an optimal value function, v* (s), which determines the outcome of the game, from every board position or state *s*, under perfect play by all players. These games may be solved by recursively computing the optimal value function in a search tree containing approximately b^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).

The team behind AlphaGo used deep convolutional neural networks combined with Monte Carlo Tree Search (MCTS) for the first time to achieve their unprecedented game playing results.

They began by training a supervised learning (SL) policy network directly on expert human moves from prior games. In addition to this they also trained a Fast Policy network which could rapidly sample moves during Monte Carlo rollouts.

Next the team trained a reinforcement learning (RL) network that improves the SL policy network by adjusting the policy towards the correct goal of winning games, rather than predictive accuracy.

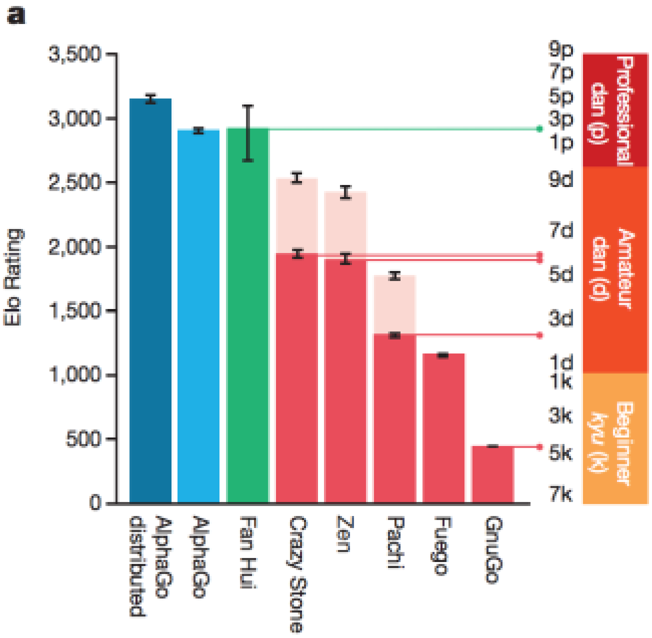
Finally, the team trained a value network that predicts the winner of games played by the RL policy network against itself.

The efficient combination of these policy and value networks using MCTS for the first time was the reason for the increased performance of AlphaGo over its predecessors.

What was achieved?

To evaluate its effectiveness AlphaGo was placed in a tournament against other leading commercial Go software.

AlphaGo won 494 out of 495 games played (99.8%) in the tournament. The following chart shows AlphaGo's ranking compared to the other software players and Fan Hui the European champion:



AlphaGo ranked many dans above the other software, according to the game's rating system for players.

The team behind AlphaGo for the first-time integrated neural network evaluations and Monte Carlo rollouts, at scale, in a high-performance tree search engine. The engine evaluated far fewer moves than previous game playing AI such as Deep Blue, compensating by selecting those positions more intelligently, using the policy network, and evaluating them more precisely, using the value network.

Go is exemplary in many ways of the difficulties faced by artificial intelligence: a challenging decision-making task, an intractable search space, and an optimal solution so complex it appears infeasible to directly approximate using a policy or value function. The previous major breakthrough in computer Go, the introduction of MCTS, led to corresponding advances in many other domains; for example, general game-playing, classical planning, partially observed planning, scheduling, and constraint satisfaction. By combining tree search with policy and value networks, AlphaGo has reached a world beating professional level in Go, providing hope that human-level performance and beyond can now be achieved in other seemingly intractable artificial intelligence domains.

Reference: AlphaGo Nature Paper