

# Paper Reading of AlphaGo

Jiyao Li

February 20, 2017

The paper I read is “Mastering the game of Go with deep neural networks and tree search”. Someone says that 2016 is the starting year of Artificial Intelligence. Although AI technology started since 1950s and even earlier, AlphaGo wining Lee Sedol, once the best professional Go player in the world, by 4 to 1, created shocking news to the public and again attracted all the public attention (and capital money) to the booming AI industry. Later, an improved “AlphaGo” player, hiding itself with a nickname “Master”, wins over 60 the Korean, Chinese and Japanese best professional Go players straight in a row, declaring another unchallenged dominance area of AI. So I am very happy to read the paper published on Nature by the Google DeepMind team.

The Nature paper is already very short. And I am afraid that I could not describe the techniques as clearly and accurately as the original authors. So I am just going to summarize the Key techniques and Result of their work.

The major breakthrough of this work is the combination of Deep Neural Network and Monte Carlo Tree Search (MCTS). As we learned from the Udacity class, the technique to make decision in game playing is to do extensive tree search over all possible states and select the next move with the highest evaluation score. The caveat is that searching in game-playing is an exponential problem, especially for Go. And due to the complexity and large states space, there was no satisfying way to build an advanced Go player. The key idea to hack the exponential game-playing search is to reduce the search depth and branching factor. However, traditional way of approximating value function  $v(s) \approx v^*(s)$  to limit search depth failed in Go. And using MCTS to avoid branching only create amateur level Go player. Previously, Convolutional Neurual Network with supervised learning was tested, but can not even beat the MCTS player.

In this paper, the revolutionary work they developed is to build several neural network to give general guidance, and combine this knowledge with the power of MCTS to give better estimate of the next best move. In general, they developed a policy network, giving a probability distribution of legal moves, making the MCTS simulation (rollout) fast. In this process, the policy network was first trained using supervised learning directly from expert human moves. Later, human influence was ignore and policy network is trained in self play using Reinforced learning to maximize the goal of winning. Then another neural net, Value network, is also created from the stronger Reinforced Learning policy network. Value network predict the expected outcome of the game given a state of the board. I feel that this should be a sophisticated Heuristic evaluation function to reduced the search depth. Although without value network, MCTS using the policy network can achieve similar accuracy on expert games but much slower. After the policy network and value network are built, AlphaGo will combine the two network in an MCTS algorithm that selects actions by lookahead search.

The result of this work is creation of a powerful and intelligent computer Go player. They ran tournament evaluation of AlphaGo with all other best computer Go player, Crazy Stone, Zen, Pachi, Fuego and GnuGo, and the best professional Go player in Europe, Fan Hui. The result is that AlphaGo defeat them all. More importantly, before AlphaGo, the previous state-of-the-art computer Go players can only reach Human Amateur level. AlphaGo is the first computer Go player to reach Human professional level and later, as I described in the beginning, defeated all best Human professional Go players... Just one last thing to say, one thing confuse me is several placed mentioned that the AlphaGo distributed

plays better than AlphaGo. Shouldn't that be obvious? Using several GPUs and even computer clusters better gives improvements over AlphaGo with less computing power. If the time limit for one move is still 5s, AlphaGo distributed should perform better than AlphaGo-fight-alone.