Artem Arakcheev
October 6, 2017

# Research Review

based on the paper
“Mastering the game of Go with deep neural networks and tree search“

## Paper’s goals

The game of Go is one of the most challenging games for Artificial Intelligence. Go has enormous search space and the difficulty of evaluating board positions and moves. The paper introduces a new approach for algorithm playing Go that uses ‘value networks’ to evaluate board positions and ‘policy networks’ to select moves. The Go algorithm uses deep neural networks are trained by a novel combination of supervised learning from human expert games, and reinforcement learning from games of self-play. Paper introduces a new search algorithm that combines Monte Carlo simulation with value and policy networks.

## Techincs introduced

Authors used deep convolutional neural networks for the game of Go. They passed in the board position as a 19 × 19 image and used convolutional layers to construct a representation of the position. They used neural networks to reduce the effective depth and breadth of the search tree: evaluating positions using a value network, and sampling actions using a policy network.

AlphaGo program efficiently combines the policy and value networks with Monte Carlo tree search (MCTS). MCTS uses Monte Carlo rollouts to estimate the value of each state in a search tree. As more simulations are executed, the search tree grows larger and the relevant values become more accurate. The policy used to select actions during search is also improved over time, by selecting children with higher values.

Authors trained the neural networks using a pipeline consisting of several stages of machine learning:

1. They begin by training a supervised learning (SL) policy network directly from expert human moves. This provides fast, efficient learning updates with immediate feedback and high-quality gradients.

2. They trained a fast policy that can rapidly sample actions during rollouts.

3. They trained a reinforcement learning (RL) policy network that improves the SL policy network by optimizing the final outcome of games of self-play. This adjusts the

policy towards the correct goal of winning games, rather than maximizing predictive accuracy.

4. Finally, they train a value network that predicts the winner of games played by the RL policy network against itself.

## Papers’s results

Authors of paper have developed a Go program, based on a combination of deep neural networks and tree search, that plays at the level of the strongest human players, thereby achieving one of artificial intelligence’s “grand challenges”. By combining tree search with policy and value networks, AlphaGo has finally reached a professional level in Go, providing hope that human-level performance can now be achieved in other seemingly intractable artificial intelligence domains.

The AlphaGo program achieved a 99.8% winning rate against other Go programs, and defeated the human European Go champion by 5 games to 0. This is the first time that a computer program has defeated a human professional player in the full-sized game of Go, a feat previously thought to be at least a decade away.

## References

1. “Mastering the game of Go with deep neural networks and tree search“ - https://storage.googleapis.com/deepmind-media/alphago/AlphaGoNaturePaper.pdf