

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

[paper source](#)

Techniques Summary

The researchers implemented a pipeline consisting of several stages of machine learning and searching:

1. Supervised learning of policy networks

A fast rollout policy p_π and supervised learning (SL) policy network p_σ are trained to predict human expert moves in a data set of positions.

2. Reinforcement learning of policy networks

A reinforcement learning (RL) policy network p_p is initialized to the SL policy network, and is then improved by policy gradient learning to maximize the outcome (that is, winning more games) against previous versions of the policy network.

3. Reinforcement learning of value networks

The value networks focus on position evaluation, estimating a value function $v^p(s)$ that predicts the outcome from position s of games played by using policy p for both players

4. 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.

Results

The researches evaluated the playing strength of AlphaGo with several tournament with other AI Go agent. And the results of the tournament suggest that single-machine AlphaGo is many *dan* ranks stronger than any previous Go program, winning 494 out of 495 games (99.8%) against other Go programs.

The Alpha Go, without handicap, wined the later competition with Fan Hui, a professional 2 *dan*, as well as, after this paper is published, well-known completions with Lee Sedol 9-*dan*, Ke Jie, the top ranked Go's player.

Thoughts

I am surprised that the fundamental idea of the sophisticated game agent AlphaGo is still finding legal move, evaluation and search as the isolation game. However, the legal move is defined and constrained by the policy network from supervised and reinforcement learning. The evaluation function is based on the value networks trained from self-play data set consisting of 30 million distinct positions. The Monte Carlo tree search is designed specifically for searching within policy and value networks.