Takahiro Ishikawa
AIND: Build a Game-Playing Agent project

Research Review

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

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

Goal and Techniques Introduced

Goals are playing Go that has long been viewed as the most challenging of classic games for artificial intelligence owing to its enormous search space and the difficulty of evaluating board positions and moves, and achieving superhuman performance. Techniques introduced on this paper are below.

1. Using 'Value Network' to evaluate board positions and 'Policy Network' to select moves. These networks are deep convolutional neural networks.
2. Training a supervised learning policy network directly from expert human moves.
3. Training a reinforcement learning policy networks that improves the a supervised learning policy network by optimizing the final outcome of games of self-play.
4. Training a value network that predicts the winner of games played by the reinforcement learning policy network.
5. Efficiently combining the policy and value networks with Monte Carlo Tree Search (MCTS).

Architecture

AlphaGo uses an asynchronous multi-threaded search that executes simulations on CPUs, and computes policy and value networks in parallel on GPUs. A single-machine version of AlphaGo used 40 search threads, 48 CPUs, and 8 GPUs. A distributed version of AlphaGo used multiple machines, 40 search threads, 1,202 CPUs and 176 GPUs.

Results

AlphaGo is stronger than any previous Go program, including the strongest commercial programs Crazy Stone and Zen, and the strongest open source programs Pachi and Fuego, that are based on high-performance MCTS algorithms. Winning rate is 99.8% (winning 494 out of 495 games).

And AlphaGo played games with four handicap stones (free moves for the opponent). AlphaGo won 77%, 86%, and 99% of handicap games against Crazy Stone, Zen and Pachi, respectively. The distributed version of AlphaGo was significantly stronger, winning 77% of games against single-machine AlphaGo and 100% of its games against other programs.