Research Review of Paper: AlphaGo

Jinxiang Ruan

Goals

Due to the complexity and the large set of possible moves when playing the game of Go, exploring the entire game tree is infeasible. This paper introduce a new approach to computer Go that uses 'value networks' to evaluate board position and 'policy networks' to select moves.

Techniques introduced

The authors of the paper suggest an innovative approach to use 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. The neural networks are trained using a pipeline consisting of 3 stages of machine learning:

1. A supervised learning (SL) policy network trained with human expert moves. At this stage, a 13-layer network is trained using image representations of the board, with 30 million positions taken from the KGS GO server.
2. A reinforcement learning (RL) policy network that evaluates self-play outcomes of the current state of the game. By making it play against itself million times and beat earlier incarnations of itself, keeping the network weights of the winner, it became much stronger.
3. A RL value network that focuses on position evaluation, it estimate a value that predicts the outcome of the game played by using the RL policy network for both players.

AlphaGo program combines neural network evaluations with Monte Carlo rollouts together, at scale, in a high-performance tree search engine. AlphaGo uses an asynchronous multi-threaded search that executes simulations on CPUs, and computes policy and value networks in parallel on GPUs. Single-machine version used 40 search threads, 48 CPUs, and 8 GPUs. Distributed version of AlphaGo exploited multiple machines, 40 search threads, 1202 CPUs and 176 GPUs.

Results

The paper shows that AlphGo significantly outperforms previously existing Go-playing AIs and obtained a superhuman performance in the game while beating human experts.

With the breakthrough of combining tree search with policy and value networks, AlphGo 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.