Mastering the game of Go with deep neural networks and tree search (Research Review)

Goals

- Introduce a program AlphaGo that combines efficiently the deep neural network with Monte Carlo tree search.
- Evaluating the board positions and moves given the enormous search space and the complexity of Go Game.

Techniques

- Introduce deep convolutional neural networks to reduce the depth (game length) and breadth (number of legal moves per position) of the search tree by evaluating positions using a value network, and sampling actions using a policy network. It covers the following steps for training and prediction:
    a) Passing in the board position as a 19×19 image and use convolutional layers to construct a representation of the position.
    b) Training a supervised learning policy network directly from expert human moves, providing fast, efficient learning updates with immediate feedback and high-quality gradients.
    c) Training a reinforcement learning policy network that improves the supervised learning policy network by optimizing the final outcome of games of self-play. The reinforcement learning policy network is initialized to the same weights of the supervised learning policy network.
    d) Training a value network that predicts the winner of games played by the reinforcement learning policy network against itself. It has the similar architecture to the policy network, but outputs a single prediction instead of a probability distribution.
    e) Combines the policy and value networks with Monte Carlo tree search.
- AlphaGo program exploited multiple machines, 40 search threads, 1,202 CPUs and 176 GPUs.

Results

- AlphaGo ran against several other Go programs in an internal tournament, including the strongest commercial programs Crazy Stone and Zen, and the strongest open source programs Pachi and Fuego. All of these programs are based on high-performance Monte Carlo tree search algorithms.
- For Single machine:
    a) AlphaGo performed better than any previous Go program, winning 494 out of 495 games (99.8%) against other Go programs.
    b) To provide a greater challenge to AlphaGo, games are played with free moves for the opponent, AlphaGo won 77%, 86%, and 99% 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.
- The mixed evaluation of the value network and rollouts performed best, winning ≥95% of games against other variants.