AlphaGo Research Review

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

This paper introduces how AlphaGo program succeeded to create a professional level play to compete in the game of Go. Although,this game is known by its enormous search space and the difficulty of evaluating board positions and moves, AlphaGo managed to handle this complexity, using Monte Carlo tree search and deep neural network (value networks and policy networks).It also combines supervised learning from human experts games and reinforcement learning from the games of self play.

Prior work based on linear combination of input features or minimax algorithm to calculate the optimal value function were not effective because of the wide search space (b to the power of d with b= 250 and d = 150 in Go).
Instead, AlphaGo algorithm aim to evaluate positions by truncating the search tree at state s and replacing the subtree bellow s by an approximate value function that predicts the outcome from states. It also reduce breadth of the search by sampling action using policy networks.
AlphaGo combines different new technologies:

- Monte Carlo tree Search (MCTS): It uses Monte Carlo rollouts that can search to maximum depth without branching by sampling long sequences of actions for both players. It estimates the value of each state in a search tree. As more simulation are executed, the search tree grows larger and the relevant values become more accurate.
- Policy Networks:Used to evaluate board positions using supervised learning trained by human experts and reinforcement learning that output a probability distribution of values over possible moves in the direction to maximize the expected outcome.
- Value networks: Used to select move using reinforcement learning on position evaluations.
  Witch led to a new search algorithm that combines Monte Carlo simulation with the value and policy networks.

Results

Combining all these technologies with each other made AlphaGo many rank stronger than any previous Go program with a win rate of 99.8%.
DeepMind team also tested many variant of AlphaGo using only value network ($\lambda = 0$) , just rollouts ($\lambda = 1$) and mixed evaluation ($\lambda = 0.5$) and the last one result of a winning rate greater than 95% because the position-evaluation mechanisms are complimentary: the value network approximates the outcome of game played by the strong but impractically slow policy. And the rollouts can precisely score and evaluate the outcome of games played by the weaker but faster rollouts policy.
To challenge AlphaGo even more , it been tested with 4 handicap stone and scored 77% to 99% win rate against other paid and open source Go programs.
The distributed version scored 77% against single machine Alpha-Go and 100% against others and beats for the first time in history the world champion Fan Hui in 5 games by a score of 5 to 0.

Dalila Boutoumilate

[1]

1 479 Words