Research Review of AlphaGo

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

Summary of the paper's goals or techniques introduced

The paper aims to build an AI agent win the game Go. Go is a chess game with breadth 250 and depth 150, made traditional exhaustive search, such as Min-max, infeasible.

Traditional search use two method to reduce the search space: **Position Evaluation** and **Policy Sampling Distribution**. **Position Evaluation** truncating the search tree at state s and replacing the subtree below s by an approximate value that predicts the outcome from s. Position Evaluation reduce the depth of the search, however, Go is too complex to use that. **Policy Sampling Distribution**, such as Monte Carlo tree search, try to sample possible moves a in position s. It reduced the breadth of search, but the agent is only weak amateur level play in Go.

Based on traditional tech, the paper tries to introduce new tech, neural network, to improve the performance. They use these 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. Another tech is **Rollouts**, using traditional linear regression instead of the neural network.

They start with a supervised learning **policy network**, trained to best predicted expert human moves. Then train a reinforcement learning **policy network**, optimize the outcome from predict expert human moves to winning games. After then, train a **value network** predict the winner of games in policy network, and try to optimize the definition of winning games.

The **Policy network** uses the current situation as input, forecasting / sampling the next step. It not only selected the best option, but also assign scores to all the possible next step. To increase the speed, they use 192 filters other than 384, the width of the board. It set the timeout period equal to 0.1 second. The choice of each step is based not the on the highest score, but the most appropriate confidence interval.

The **Value network** uses only one situation in the whole round as input to avoid overfit. It gave a long run position evaluation method. Value network did not dramatically improve the

performance of policy network, however, it works independently from the rollout, which help the AlphaGo performance increased from Professional to World Champion level.

The **Rollouts** used traditional logical regression instead of the neural network, devote the performance to the response speed. It made the evaluation not trapped by timeout.

Summary of the paper's results

They introduce new search algorithm that successfully combined neural network evaluations with Monte Carlo rollouts. Compared to Deep Blue, they evaluated thousands time fewer position, use general purpose supervised and reinforcement learning methods, instead of handcrafted evaluation function. Most important, they provide human-level expert performance using AI agents.

As we can see, there are three techs used in AlphaGo: Policy Network, Value Network and Rollouts. Traditional agent can only work as beginner, around 1000 EIO rating. The policy network improved 700 EIO rating, make the agent as efficient as an Amateur. The value network improved 480 EIO rating, make the agent as efficient as a Professional. The Rollouts Network improved 600 EIO rating, make the agent as efficient as human-level expert.