

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

In this paper you are going to read a brief summary of the paper "Mastering The Game of Go With Deep Neural Networks and Tree Search".

Goal or Technique Introduced

- Normally, when we are creating a game agent, one of the most efficient way to make the best move is to explore all possible outcomes of player-opponents moves.
- However, it is not possible in the game of Go because, given that there are $361!$ combinations, it is nearly impossible to calculate the best move just by exploring all possibilities.
- To overcome this issue, we use another way to solve this problem by using the neural network approach.
- There are mainly two parts in the AlphaGo design: Policy network and value network.
- Policy Network: Given a particular state of a game (like the arrangement of stones), the probability of placing stone in coordinate X. In other words, it is $P(\text{the probability of playing X} | \text{Game state})$.
- Value network: An evaluation method that evaluate the overall game state after playing stone in coordinate X.
- In other words, if you have a perfect policy network, you will know the biggest probabilities of playing coordinate X in a given state. If you have a perfect value network, you will know which moves gives you the best overall game state.
- To achieve that, there are several steps to take.
- First, the research team uses 160,000 games played by KGS 6 to 9 dan human players, which contains 29.4 million positions, to training the policy network for classification.
- In order to do the training, there are two key technologies to use: Deep Convolutional Neural Network (DCNN) and Monte Carlo Tree Search (MCTS).
- Then the researchers use a reinforcement learning (self-playing games) to optimize its policy network, based on the result AlphaGo is getting using Policy Gradient Method and MCTS.
- Finally, based on the optimized policy network, the researchers use supervised learning - regression to calculate the value network, which is a 15 layers CNN.
- Using Policy network and value network, AlphaGo can obtain the best move given the current game state.

Result:

With the above neural network architecture, AlphaGo/Distributed AlphaGo can achieve Elo of 2890 and 3140, which beats all mainstream Go programs like CrazyStone (1929) and GnuGo(431). Recently it also beat Professional Go player, **Lee Se-dol 9 Dan(Korean) and World #1 Go Player Ke Jie.**