Udacity Artificial Intelligence Project2 : Build a Game-Playing Agent

Part2: Research Review – AlphaGo by DeepMind

Julianne Hong

The paper talks about how DeepMind team achieved the Go game agent AlphaGo. The general techniques of building a game agent in artificial intelligence is to have a tree that covers all the possible states of a game (b^d) and recursively compute the evaluations of each nodes. However, doing that with games like Go that the size of the node is approximately 250^{150} will take more than life time. What DeepMind adopted to solve this challenge is Monte Carlo Tree Search(MCTS) combined with value network and policy network.

MCTS is an algorithm consists of a loop with four simple stages: Selection, Expansion, Evaluation, and Backup. The agent traverse the tree and when it reaches the node that does not have children nodes, it expends limited number of random nodes (maybe selected by rollout policy). The agent evaluates the expanded nodes, then traverses back to the current state of the node and recalculate more informed value with data from the random simulation. The agent does this in a loop until the evaluation's termination conditions meet.

Supervised learning (SL) of policy network is a prior work before playing actual moves. This policy is trained by randomly sampled state-action pairs. This policy is to be updated as game goes on from the data from the simulations of MCTS. During the game, the agent does Reinforced learning (RL) policy network. Initially, RL policy is the same as RL policy. During the simulation, the agent play game against random selections from the previous iteration of the policy network. It uses the reward function to update the RL policy each time. Then, it calculates the value of the state based on the strongest policy, RL at that moment. MCTS's Backup phase allows the agent to update at all the action values and traversed node in each simulation.

By combining MCTS algorithm with Policy network and value network, AlphaGo achieved a Go playing agent that can play against world champions and defeat strong other Go program with wining rate of more than 99%.