Review Of AlphaGo' Paper

The AlphaGo is one of the most famous AI application that has receive a lot of attentions last year. The Go's team employs convolutional neural network to improve the performance besides traditional methods. In general,   AlphaGo contains four components: Policy Network, Rollouts, Value Network and Monte Carlo Tree Search.

To build the model, the team trains the network by using SL(supervised learning) which has 13 layers and trained with real expert human moves.It is being called Policy Network, using stochastic gradients to maximize likelihood.Finally, the network predict the expert human's move with about 57%. Besides, there is another networked trained called Rollout to improve the calculation speed.

The next step to improve the Policy Network is gradient reinforcement learning(RL). It's initialization is the same as Policy Network. The weight updates at each time step by stochastic gradient ascent to maximize expected outcomes with a reword function at the perspective of current player at time step t. It performs better than SL and other previous state-of-the-art.

The final stage of the training is focused on Value Network which predicts the outcome of specified position. It has a similar architecture as RL but outputs a single result instead or probability distribution. The group trains the network on state-outcome pairs with 30millions distinct position from separate games. Finally It has MSEs of 0.226 and 0.234 on the training and test set.

AlphaGo combines the policy and value network in a MCTS algorithm. MCTS is a method to get a max value when searching in the tree which is popular recently. The leaf node is evaluated by the combination of value network and the outcome of a random rollout played out until terminal step. The output probability is stored as prior probability. Once the search is complete, the algorithm choose the most visited move from the root position. In this way, It requires huge CPU and GPU resource and threads.

The AlphaGo performs roughly as one amateur dan rank advantage on KGS. In addition, Using all of Rollouts, Value Network and Policy Network perform best in all permutation and combination of these function. Obviously AlphaGo defeat the strongest Go programs. It indicate that the position-evaluation which are value network and rollouts are complementary.

Different from traditional Go program, AlphaGo use a combination of neural network and Monte Carlo Search Tree to build it reaching a professional level in Go. AlphaGo resolve great difficulties of decision-making, intractable search space and complex optimal solution. What helps a lot is policy network which gives AlphaGo a vision of overall situation. Rollouts and value network help AlphaGo to get optimal solution quickly. Reinforcement Learning perform not as important as I think in AlphoGo, , and there is a lot to to let it act as a general AI.