Research Summary:

Mastering Alpha Go paper discusses on using supervised learning and reinforcement learning to create Policy Networks and Value Networks to evaluate the board and selecting an optimum move. It also introduces a new search algorithm based on the policy and value networks along with Monte Carlo Tree Search (MCTS) to achieve an accuracy of 99.8% winning rate.

Policy Networks are trained by expert moves data by supervised learning which is fed to next stage of training pipeline for self play as basis for reinforcement learning. Value Networks predicts the desired outcome. The paper discusses details about a 13-layer policy network which is fed with 30 million move positions and about trade-offs on accuracy and speed of the network.

Results

By combining the neural networks with MCTS, AlphaGo has achieved a greater winning rate by reducing its evaluation search space. AlphaGo outperformed DeepBlue in a more complex game by selecting those moves by policy network and evaluating using value networks.