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

The game of Go has for long been considered a game too complex to be played professionally by a computer. The game's search space is incredible large ($b^d = 250^{150}$) making exhaustive search infeasible.

The authors of the paper demonstrates a clever solution to reducing the depth and breadth of the search, achieving world class performance in the game of Go – AlphaGo. Game Playing is all about making the right decisions given some knowledge and constraints.

AlphaGo is trained using clever 'policy networks' and 'value networks' to evaluate the best moves playing against an opponent.

The Policy Network (PN) uses a combination of Supervised Learning (SL) and Reinforcement Learning (RL).

The SL policy network learns to classify human expert positions to determine the optimal policy. Alternating between 13 convolutional layers with weights and rectifier nonlinearities, the network outputs a probability distribution over all legal moves.

Building upon the SL policy network the RL policy network improves the pipeline by policy gradient reinforcement learning to simulate games with different iterations of the policy network. Calculating the terminal reward at the end of each game for the current player and using stochastic gradient ascent to maximize the expected outcome drastically improves the policy network.

With a strong policy network the agent needs to be able to evaluate positions in the game. AlphaGo uses reinforcement learning in it's value networks to evaluate the best moves to make. By minimizing the mean squared error between the predicted value and corresponding outcome of a policy.

A Monte-Carlo Tree Search algorithm is used to traverse the nodes in the game space and find the best move. By evaluating every node based on its action value, visit count and prior probability, AlphaGo achieves world class performance in the game of Go.

The agent architecture demonstrated in this paper reveals the power of deep learning techniques combined with lots of computational power (CPUs and GPUs).