Review : Mastering the game of Go with deep neural networks and tree search Silver et al. 2016

This paper presents an AI agent (AlphaGo) that can achieves a 99.8% winning rate against other AI solutions to Go. The agent also beat the European Go champion 5 games to 0. The paper introduces a new approach using 'value networks' to access board positions and 'policy networks' to select moves.

Each of these networks is a deep neural network (DNN). To train a policy network for selecting actions a DNN was first trained using supervised learning on human expert moves. Given the current state of play the network was trained to recreate the actions of human experts. The policy network was then improved by training a reinforcement learning (RL) policy network. To begin this network was initialised to the trained supervised DNN. The RL network then improves itself by playing many games against itself and learning which actions are best given a current state. A fast rollout DNN was trained similar to the supervised learning DNN. The simpler network architecture returns policy much faster than the final RL policy network, but with less accuracy though the speed is important when searching which moves to make.

The value network was also trained using RL with the aim of predicting the result of the game from a given state suing data from self play. The agent plays then game using Monte Carlo tree search with the value and policy networks. Actions are selected by lookahead search. Each edge of the tree stores an action value, a visit count and a prioir probability. A node is expanded and each child node is scored by the policy network and the prior probability of the is updated according to this score. A bonus score is also calculated proportional to the prioir probability, but decayed by the current visit count so as to encourage exploration of other regions of the search tree. The child is also evaluated by the value network and the fast rollout policy network which along with the visit count are used to calculate the action score.

Through the use of generalized supervised and RL machine learning AlphaGo can outperform humans at Go. Go is orders of magnitudes more complex than chess yet AlphaGo evaluated thoudsands fewer positions than Deep Blue during its match against Kasperov.