

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

<https://storage.googleapis.com/deepmind-media/alphago/AlphaGoNaturePaper.pdf>

The paper describes the design that the Google DeepMind team took to develop AlphaGo: the first computer program to defeat a human expert in the game of Go; a game that has been seen as one of the most challenging classic games for artificial intelligence due to the large search space and difficulty on evaluating board positions and moves.

The AlphaGo agent uses Monte Carlo Tree Search (MCTS) to reduce the depth of the search tree by estimating the value of each state in the search tree. Monte Carlo rollouts are used to estimate the value of each state in the search tree.

Prior to AlphaGo the best Go programs were based on MCTS using policies that were trained by expert moves achieve an amateur level of play. On AlphaGo the DeepMind team uses deep convolutional neuronal networks to reduce the depth and breadth of the search tree. Two different types of neuronal networks are used: a "value network" used to evaluate positions and a "policy networks" used to sample actions.

Supervised Learning (SL) is used in the first stage to train the policy neuronal network. In this step the neuronal networks is trained by moves provided by human experts. This is a fast and efficient way for the neuronal network to learn. During rollouts, a fast policy networks is also train. The next step uses Reinforcement Learning (RL) to train the policy network optimizing the final outcomes of the game. The purpose is to adjust the policy toward the goal of winning games instead of focusing on maximizing predictive accuracy.

The last stage is to train the Value Network used to estimate the value function for the strongest policy derived from RL policy network. The architecture of this network is similar of the policy network but the output is a single prediction. They train the weights of the value network by regression on state-outcome pairs, using stochastic gradient descent to minimize the MSE between the predicted value, and the corresponding outcome.

With AlphaGo the DeepMind team developed an effective move selection and position evaluation functions for the game of Go using deep neuronal networks, supervised learning, reinforcement learning and a new search algorithm that combines neuronal networks with Monte Carlo rollouts. All these components are integrated together at scale, providing a high-performance tree search engine. Using these techniques AlphaGo achieved a 99.8% winning rate against other Go programs and became the first computer program to beat a human professional in the Game of Go in a complete match winning 5-0 against the European Champion, Fan Hui.

Similar how the introduction of MCST to Go computer programs led to advances in other domains the achievements the DeepMind team on AlphaGo are applicable beyond the game of Go. These techniques could be used in any domain where there challenging decision making tasks, massive search spaces and a solution that appears infeasible to approximate using a policy or value function.