Research Review

DeepMind's world champion AlphaGo described in Nature is a combination of several several state of the art methods in parallel computing, artificial intelligence and machine learning. Each turn, AlphaGo searches for the the best action based on two-part score: the value of the action plus a bonus based on a policy network evaluated at that state. The policy network that computes the bonus score is a convolutional neural network trained on expert moves from recorded gameplay and AlphaGo's own simulated play. The search implementation uses Monte Carlo Tree Search (MCTS) to expand the tree and a hybrid evaluation function to value leaf nodes composed of a slower value network and a very fast policy iteration rollout in equal proportion. The fast policy iteration is used to look ahead very far and very fast, while the slower value network is another convolutional neural network trained through reinforcement learning. The fast policy rollout and the value network can be evaluated in parallel, as can the tree expansion, such that the single machine version of AlphaGo uses 40 cores and 8GPUs simultaneously in its search.

The effect of these combinations is unparallelled mastery of a game that dwarfs Chess in complexity. The hybrid scoring method used by search lets AlphaGo select actions based on analysis of their merit and its memory of similar actions encoded in the policy network. It is still infeasible to expand the entire Go game tree from a single state, but using MCTS expansion over the prior probability computed by the policy network, it can still explore the most promising actions in each state to significant depth. At the leaf nodes, the deep learning value network takes into account the outcomes learned from previous gameplay at similar states, while the fast policy rollout helps find winning states and avoid losing ones when the heuristic is short sighted. In essence, it thinks a lot like a human player, expanding a limited number of promising actions, estimating the value of their outcome in a limited number of moves based on board position, while seeking victory and avoiding losing positions.

AlphaGo soundly beat both open source and commercial Go agents by several orders of magnitude, and won reliably even with a four-stone handycap. At the time the article was published, AlphaGo had defeated the European Go champion. Two months later, AlphaGo defeated 18-time world champion Lee Sedol, winning four out of five rounds. A year later, a version of AlphaGo running on a specialized architecture defeated the current world champion Ke Jie in a three round match, winning all three rounds. AlphaGo has since retired, as there are no opponents at present who could challenge its mastery of Go.