Summary of AlphaGo paper by DeepMind team

Before this breakthrough paper, building an agent with human level performance in the game of Go was considered far from reality. This task was challenging because of its enormous search space of approximately b^d to large average branching factor ($b \approx 250$) and large depth ($d \approx 150$). In this paper search space is reduced by both reducing both b and d in the following ways :

1. Depth of the search tree is reduced by truncating the tree at node s by its approximate value function v(s) that closely mimics the ground truth value function v*(s).
2. The breadth of the search tree is reduced by sampling the moves (actions) using a policy distribution P(a | s) which gives the probability of action to take given the current state.

Both value function and policy are built using neural networks. First a supervised learning policy network p_σ is learned directly from expert human moves, in addition a fast policy p_π network is also trained that can rapidly sample actions. Then a reinforcement learning policy network p_ρ, is trained which improves on p_σ by setting the goal of winning games (in RL setting) instead of just maximizing the supervised learning objective of maximizing the accuracy. In the final step a value network v_Θ is trained by playing Go using the p_ρ policy network against itself. During evaluation AlphaGo effectively combines the policy and value networks with Monte Carlo tree search to make all the moves.

Results summary

AlphaGo was evaluated by playing it against some Go programs - Crazy Stone, Zen (commercial) and Pachi and Fuego (open source) (all based on high performance MCTS algorithms). All programs were allowed 5 seconds of computation time per move and AlphaGo won 494 out of 495 games played (99.8%). AlphaGo also competed against Fan Hui (European Go championship winner from 2013-2015) and AlphaGo won a formal five game match 5-0.