Research Paper Review

Mastering the game of Go with deep neural networks and tree search, David Silver et al (2016)

Method summary

AlphaGo uses Monte Carlo Tree Search (MCTS) to estimate the value of each state in it' search tree. MCTS is used in the best Go programs currently available. AlphaGo uses deep convolutional neural network structure, which has been successful in image classification and recognition tasks, to reduce the breadth and depth of the search tree by evaluating positions using a value network and sampling actions using a policy network.
Several levels of machine learning are used to train the AlphaGo neural network. 1(a) Supervised Learning policy that is trained using human expert moves. This is a 13-layer network trained using 30 million positions from KGS Go server. It could predict 57% of expert moves using all the input features (other research group reported only 44.4% accuracy) 1(b) A fast rollout policy that can sample actions and thereby reduce the search space is also trained. It took only 2us to sample an action with 24.2% accuracy, 2) The supervised learning policy is then strengthened using reinforcement learning (RL). The RL policy network has a similar structure as the SL network. RL network is initialised with the same weights as the SL network and then plays against randomly selected (to prevent overfitting) opponents where the opponents are previous iterations of the RL network. and 4) a value network is trained so as to predict the winner of the game played by the reinforcement policy.

Result summary

AlphaGo program was tested against opponents which uses high performance Monte Carlo Tree Search (MCTS) such as different variants of AlphaGo, commercial Go programs Crazy Stone and Zen, open-source Go programs Pachi and Fuego and open-source Go program called GnuGo which uses non-MCTS algorithm. The time limit used for computation time was 5 seconds per move. AlphaGo performance was tested for both single machine version as well distributed version.
Single machine AlphaGo won 494 out 495 games i.e., 99.8% win rate. AlphaGo performed better even when opponents were given 4 free moves. With the free moves for the opponent, it had a win rate of 77% against Crazy Stone, 86% against Zen and 99% against Pachi. The distributed version of AlphaGo had a 100% win rate against all the opponents and 77% win rate against the single machine AlphaGo.
Different variants of AlphaGo- with position evaluation with only rollout, only value network, combination of both and only policy network(no search) was also assessed. The best performance was observed where all three components were used (>= 95% win rate against all other variants).
The distributed version of AlphaGo also beat a 2 dan professional player 5-0, which is the first time a computer beat a professional Go human player. AlphaGo evaluated thousands of times fewer moves during the match than Deep Blue did during its match against Gary Kasparov. This breakthrough in Go game can potentially lead to solutions in areas considered impossible for AI to solve.