Research Review: AlphaGo by DeepMind Team

In this research review I review the game paper AlphaGo by the DeepMind Team. AlphaGo is the Artificial Intelligence (AI) program, developed by the DeepMind team (part of Alphabet), that defeated the European Go Champion by 5 games to 0 in the highly complex Go game. I use the words complex here to describe the enormous search space and the difficulty of evaluating board positions and moves. The Go game is a 19 by 19 (361 positions) sized board game with a search space of approximately 10^{170}. The branching factor (b) of Go is approximately 250 and the game depth (d) is approximately 150. Its clear that exhaustive search by recursively computing the optimal function $v^{*}(s)$ is infeasible. This is the first challenge the AlphaGo program had to overcome. To conquer Go, AphaGo introduces a number of new techniques and goals and combine these with old techniques (Monte Carlo Tree Search) to produce superior performance.

The team introduced the technique of using “value networks” to evaluate board positions and “policy networks” to select moves. They implement these value and policy networks using deep convolutional neural networks. These networks are combined with Monte Carlo Tree Search (MCTS) – that selects actions by lookahead search. Efficiently combining deep neural nets with MCTS is a new search technique on its own.

These neural networks perform an important role of reducing the search space (i.e. depth and breadth) by evaluating positions using a value network and sampling actions using a policy network. The procedure in simple terms involve two pipeline stages. In the first stage a fast rollout policy network and supervised learning policy network are trained to predict expert human moves. A reinforcement learning policy network is then training using the supervised learning network as the input. A new dataset is then produced by playing games of self-play with the reinforcement policy network. A value network [which focuses on position evaluation] is then trained by regression to predict the expected outcome.

The results of the above pipeline, when played head to head, the reinforcement learning network won more than 80% of the games against the supervised learning policy network. The reinforcement learning policy network also won 85% of games against Pachi, a top open source Go program.

The sophistication of evaluating policy and value networks requires significant computational power than traditional search heuristics so multi-threaded search is used. The final version of AlphaGo used 40 search threads, 48 CPUs and 8 GPUs. This final version won 99.8% of games against other Go programs and also won 77% of games with handicap against the powerful and commercial Crazy Stone Go program. A distributed version of AlphaGo is the one that beat the European Champion (Fan Hui) 5 games to 0.

Effectively combining deep neural networks and Monte Carlo Tree Search enable AlphaGo to play at the level of the most superhuman human players. AlphaGo evaluated thousands of times fewer positions than Deep Blue, compensating by selecting those positions more intelligently, using the policy network and evaluating them more precisely using the value network.

With AlphaGo we have reached a new benchmark in Artificial Intelligence. We are getting closer and closer to mimicking superhuman performance and from here we can only expect to see machines thinking closer and closer to humans.