Research Review

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

The Google-owned research lab, DeepMind, created a new AI agent called AlphaGo, to play the classic game of Go. Go shares some similarities to the game of Isolation that was used for this project. They are both two-player games that provide perfect information, however Go has a much larger state space, which makes it a much more difficult task for machine learning. Until the performance of AlphaGo was shown to the world, it was thought that Go was too complex of a game for current AI techniques to perform well. Nonetheless, AlphaGo shocked the world when it defeated the human Go European champion 5 game to 0 in October 2015.

To achieve such a feat, new techniques were produced by DeepMind. Value networks were created to evaluate board positions and policy networks to select moves. At a basic level, AlphaGo operates in a similar way to the game-playing agent in the Isolation project, i.e. searching through a game tree to find the optimal move. However, given the breadth and depth required by an agent to play Go well, AlphaGo instead uses Monte Carlo Tree Search with a deep convolutional neural network to evaluate and select the best moves. The final layer of the CNN is a softmax layer, which outputs the probability of each potential move leading to victory.

To train their agent, DeepMind created a pipeline consisting of several stages of machine learning. First, expert moves from humans are used to train a supervised learning policy network (SLPN). This provides both quick and accuracy training. Next, a reinforcement learning policy network (RLPN) is trained to improve the results of the SLPN through games of self-play. Lastly, a value network is trained to predict the winner of games played by the RLPN against itself.

One of the issues of this complex network is the amount of computation that is required for evaluation. AlphaGo uses an asynchronous multi-threaded search that executes simulations on CPUs, and computes policy and value networks in parallel on GPUs. The distributed version of AlphaGo runs on 1,202 CPUs and 176 GPUs. Thanks to this amount hardware, the distributed version of AlphaGo is frequently able to select a move in under five seconds.

Through these innovative methods, DeepMind was able to produce other impressive results. To further prove its domination of the game, AlphaGo earned a winning rate of 99.8% against other Go programs. Its predictive ability is also very strong, as it achieved state-of-the-art results by anticipating expert moves with an accuracy of 57.0%, compared to previous records of 44.4%.