Research Review

By Michael Shum

23 September 2017

Paper: AlphaGo by the DeepMind Team

Goals and Techniques Introduced:

AlphaGo is a game engine that plays the game of Go, which is so complicated with 250^150 potential moves that computers were not expected to beat humans for another decade at least. Existing systems used Monte Carlo Tree Search (MCTS) with policies that tried to emulate expert human players.

AlphaGo uses a different method by applying deep convolutional neural networks to the problem. Board positions are passed into the system as 19x19 images. Two main neural networks are created: a supervised learning (SL) policy network, and a reinforcement learning (RL) policy network.

The SL policy network is trained using a dataset of expert human games to predict how the best players would play. It is a 13-layer policy network that uses stochastic gradient ascent to calculate the probabilities of every available move. Several versions of this method are created, one optimized for accuracy and one optimized for speed.

The RL policy network is trained against itself and augments the SL network but the goal is to win games rather than predict correct moves. The network plays against a random past iteration of itself to prevent overfitting. A third network, the value network, is trained to predict the winner of games as the RL policy network trains by playing itself.

Finally, AlphaGo combines the policy networks with MCTS to create a winning combination that is more effective than both alone. It does this by using MCTS to traverse game trees, but each node is processed by the policy networks, and each node is evaluated using the value network and through a rollout to the end of the game.

Results:

The SL policy network could predict expert human moves with an accuracy of 57% at its peak, although the faster network predicted moves with a 24.2% accuracy.

The RL policy network could win greater than 80% of its games vs the SL policy network. Without using MCTS, the RL policy network was also to win 85% against one of the top competing Go-playing AI systems.

On an Elo scale to measure its proficiency, AlphaGo using just its policy network had an Elo of approximately 1500, using just the value network it had an Elo of about 1700, but by combining rollouts with the policy and value networks, its Elo jumped to about 2800. This is an Elo comparable to Fan Hui, the European champion (Elo about 2800), and much ahead of the previous best AI, Crazy Stone, which had an Elo of about 2000. In its final form, AlphaGo won 494 out of 495 games against competing Go programs. It was also able to beat Fan Hui 5 games to 0.