Research review of the AlphaGo paper

Zhenyuan Liu

December 2017

1

AlphaGo combines Monte Carlo simulation with deep reinforcement learning and achieved amazing performance.
It first trains a supervised learning(SL) policy network $p_\sigma(a|s)$ using human expert moves from the KGS Go Server. This process is very fast and efficient, because the gradients are high-quality. The policy network trained in this way achieves a 57.0% accuracy in the held-out test set in predicting expert moves.
In the next step, it improves the SL policy network $p_\sigma(a|s)$ using policy gradient reinforcement learning(RL). The RL policy network $p_\rho(a|s)$ has the same structure as the SL policy network, initially $\rho = \sigma$. They prevent over-fitting by playing games between the current policy network p_ρ and a randomly chosen policy network from previous iterations. The RL policy network won more than 80% of the games head-to-head against the SL policy network.
The final step focuses on a value function $v^p(s)$ that predicts the outcome from position s of games played by using the policy p for both players:

$$v^p(s) = E(z_t|s_t = s, a_{t,\ldots,T} \sim p)$$

In their paper, they approximate this value function using a value net work $v_\theta(s)$with weights θ. This neural network is similar to the policy network, but it outputs a single prediction instead of a probability distribution. The weights of this value network are calculated using regression on the state-outcome pairs (s, z). Again, over-fitting is mitigated by playing games between the RL policy network and itself. AlphaGo combines the policy and value networks in an MCTS algorithm.
The results of the tournament shows that AlphgGo outperforms other Go programs and even human professional by a large margin. It wins 494 out 495 games against other Go Programs. It also defeats a human European champion by 5-0. After two years since the initial AlphaGo paper, the evolved AlphaZero is much stronger than its predecessor, moreover, it requires no human moves for training anymore, requiring self-playing only. The techniques of deep reinforecement learning is an exciting area to explore.