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

D. Silver et al.
Summary by Christine Payne

Goals

The authors aimed to create an artificially intelligent GO player, that would be able to beat the best human player. Prior to these results, AI agents were only able to reach the level of strong amateur human players.

Techniques

The authors combined two known techniques to make an agent that was significantly better than anything seen previously. First, they train a policy network p_σ (they use a neural net, training based on moves expert humans would make). They were able to achieve 57% accuracy, when previous research groups had only reached 44.4% accuracy. This yielded a large improvement in playing strength.

They next improve this policy network by policy gradient reinforcement learning. They play games between the current policy network and a randomly selected previous version of the network. Weights within the network are updated based on whether the current policy wins or loses. At this point, they ran a tournament between the current policy and those from before this reinfocement learning step. The current ones won 80% of the time. Also, when tested against the strongest open-source Go program (Pachi), these policy networks won 85% of games.

The final stage of training focused on position evaluation. It is computationally too intensive to compute the value for each state by determining perfect play through the remainder of the game. Instead, they use the policy network (described in the previous two paragraphs) to simulate playing the game. The tree is traversed by looking at each step and choosing the next state based on the policy network's choice. This way they are able to greatly reduce the number of branches that need to be evaluated, instead of considering every single possible next state of the game.

Results

To evaluate AlphaGo, the authors ran an internal tournament against variants of AlphaGo and against the strongest commercial and open source pro-

grams. All programs were allowed 5 seconds of computation time per move. AlphaGo won 494 out of 495 games against other Go programs. Even with four handicap stones (free moves for the opponent), AlphaGo won 77-99% of the games.

Finally they evaluated AlphaGo against Fan Hui, a championship human player. AlphaGo won the match 5 games to 0.