Research Review: A dangerous combo of Deep Learning and Tree Search to beat Go expert

Eduardo Carrasco

In October 2015 Google DeepMind created a game agent called AlphaGo. The Alphago agent played against Lee Sedol in a five-game match. The advantage that AlphaGo had over other Go game agents is that it was not handicapped, it was able to play and decide on moves that were set up in a 19x19 grid. The paper introduces the usage of deep neural networks with a combination of search trees to allow optimal search space.

The authors of the paper suggest an innovative use to approach that leverages deep neural networks, more specifically three types of networks:

1. A *Supervised Learning of Policy Network* trained with human expert moves.
2. A *Reinforcement Learning policy network* that evaluates self-play outcomes of the current state of the game.
3. Another reinforcement learning layer that predicts the winner of the games played with the first reinforcement learning layer.

The AlphaGo agent uses a combination of policy and value Monte Carlo Simulation Tree effectively.

At the supervised learning stage a 13-layer network was trained using image representations of the board with moves taken from the KGS Go server. This is a similar strategy used in convolutional neural network like Google's ImageNet. This allowed the network to achieve a fairly high prediction rate per move made at about 57%.

The first neural network performs a classification which produces a policy gradient. The second layer produces a probability distribution of moves and passes this into the third layer

which then performs a regression to establish the winner of the moves evaluated from the given move set.

This after the network were displayed the Monte-Carlo search tree which utilize random sampling of the tree with evaluation of each game tree branch. At each visit evaluations are stored so as to quantify a bonus for each branch that effetely determines the most promising moves in order to win the game. The value of each explored node is the mean of an evaluation function over the number of visits on that node.

The paper shows that this method has not only obtained a superhuman performance in the game, but also – thanks to the combination of policy and value networks with tree search that the game agent has evaluated far fewer moves than DeepBlue did in its chess match against Kasparov, while beating human experts at a game that is far more complex than chess.

The other huge advantage that AlphaGo has over DeepBlue is that function where not hardcoded in its possible decision space but rather used machine learning methods to maximize like hood of winning.