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

A Research Review - Summary by Pasteur Tran

Goals and Techniques Introduced

Exhaustive breadth and depth search is infeasible in a game such as go with a breath of 250, and a depth of 150. Depth can be reduced by position evaluation (searching the tree at the current state). THe breath can be reduced by actions by *policies.* Which are based on the probability distribution over possible moves in the current position. This however only succeed with strong amateur play. Instead, by passing the board as an image and using convolutional layers to construct positions they are able to reduce the effective breadth and depth.

There are a few stages to creating this:

1. Supervised learning is used to on *previous* expert moves in the game. At this point, it was able to predict 57% expert moves - but small improvements resulted in larger playing strength. Although larger networks could be used - but they were a lot slower.
2. Reinforcement learning - the second stage involved improving the policy - this is done by using the *current policy* vs. an *older* policy network (randomly selected). Randomizing is to stop *significant overfitting.* Using a reward system, the weights are updated.
3. Reinforcement on position - using networks which predict the outcome from incomplete games is best (as it would otherwise overfit). Successive positions are correlated.
4. By combining these networks, it selected by looking ahead and each tree is traversed by a simulation. The leaf nodes are evaluated by the value network and the outcome itself.
5. It ended up using 45 CPUs and 8 GPUs!

Evaluation of AlphaGo was done by internal tournament amongst variants. Winning 494 out of 495 games! Even providing handicap stones, it was able to win between 77-99% of games.

Summary

In Summary, the Go program developed is based on a combination of deep neural networks (images) and tree search. IT achieved one of artificial intelligence's 'grand challenges' and was trained with a combination of supervised and reinforced learning. Although AlphaGo evaluated thousands of times fewer than Deep Blue - it made more intelligent positions using policy networks and value network - which is perhaps 'closer' to how humans played. The network was also trained directly from gameplay purely through general-purpose supervised and reinforcement learning methods.