Research Review

Jaco Fourie

July 18, 2017

1 Summary of research article *Mastering the game of Go with Deep Neural Networks and tree search*

In this article the researchers describe their design of a game playing agent, which they called AlphaGo, for the game Go that was able to defeat the European Go champion. The game of Go has simple rules but has an enormous search space of possible moves for each player and a full game can last hundreds of moves to reach the end. This makes it very challenging to evaluate a game state using the traditional approach of recursively calculating the optimal move from any given position.

Due to the size of the search space an exhaustive search is considered unfeasible and the best Go game agents use statistical methods to sample from the full space of possible moves to evaluate a board state. An example of this is the Monte-Carlo tree search method (MCTS) that uses random sampling of a trained distribution of moves and outcomes to estimate the value of a move. While this approach has been somewhat successful it is limited by the shallow depth of its search tree and is restricted to only a linear combination of inputs features to describe a state.

The authors of this article chose instead to leverage the recent success of deep neural networks to efficiently evaluate a Go board and estimate the optimal move. Their solution consisted of a pipeline of several layers of machine learning starting with a convolutional neural network (CNN) to capture the state of the game and represent it as a condensed feature vector. Non-linear regression is then used to predict human expert plays from a given game state as represented by the feature vector. This part of the agent is called the policy network and uses supervised learning to predict the best move based on human expert knowledge.

The second part of the agent uses reinforcement learning to improve on the policy network by playing version of itself against each other and using the results to improve itself. Reinforcement learning is also used to build a network that evaluates a player's relative standing in the game and tries to predict the game's outcome. This is called the value network and is used to score a particular move to estimate whether it is an improvement on alternative move.

This approach proved to be remarkably effective and performed better than all previously published results. In a tournament that aimed to demonstrate AlphaGo's ability it

was able to defeat the European champion Fan Hui. The tournament consisted of 5 official games which were all won by the AlphaGo agent.