Review of "Mastering the Game of Go with Deep Neural Networks and Tree Search"

Jean-Marc Beaujour
AIND - Project 3

May 26, 2017

Abstract

In this we article, we review a novel approach training a AI agent to play the Go game. The authors of "Mastering the Game of Go with Deep neural Netowrks and Tree Search" [*Nature* **529**, pp.484-489 (2016)] demonstrate the use of Deep Learning and Reinforcement learning to train the agent. Perfect information games such as Backgammon can be solved using Tree Search algorithm that exploits the future of every possible move in term of win or loss at a particular stage in the game. In order for a Tree search to be scalable, a number of heuristic evaluation function can be exploited. The game of Go is of particular interest to the AI community because it has a large number of possible moves, i.e a large search space.

1 Summary of the paper techniques

Alpha Go uses Three convolutional networks: two policy networks and one value network. The convolutional Network enables to learn the evaluation function. That is different than the approach used in previous Go AI for which the evaluation function was designed.

1.1 Supervised Learning of the Policy Network

The policy Neural Network predicts the best possible move at a state of the game. The Policy Network is made of 13 layers. The input features are the

current game state (an image of the board and additional input features) and the output is a softmax, where each node computes the probability for each legal move to be the actual next move. The network is trained on randomly sampled state-action pairs from 30 millions positions (a dataset of human experts moves).

1.2 Reinforcement Learning

To further improve the Policy Network, **Policy gradient Reinforcement Learning** is used. The Policy Network was opposed to instances of the policy network that are randomly drawn from previous iterations. With this approach, the agent improves its performance while also preventing overfitting.

1.3 Value Function

Another Neural Network, which has the same architecture as the Policy Network, is used to compute the value function. In this case, the output layer is made of a single node that represents the probability of a win.

2 Summary of the paper results

The Novelty of the approach is that the Neural Networks are trained by using a combination of Supervised Learning and Reinforcement Learning: i.e the AI is trained on Human expert games and against itself. With this search algorithm, Alpha Go achieves a winning rate of 99.8% when playing against other Go programs. Furthermore, it successively defeated the 3 best world players.