N. Veiga (nohe.veiga@gmail.com)
Artificial Intelligence Nanodegree
Aug 2017

# Research Review

AlphaGo by the DeepMind Team,  484 NATURE, Vol. 529. 28 January 2016

---

# Summary

This paper exposes one of the biggest breakthrough in AI in the recent years. Describes a new approach to the computer game Go based on Deep Learning plus Monte Carlo Tree Search (MCTS) to generate a values network to evaluate board positions and a policy network to select the moves.

In large games as Go, an exhaustive search tree is not feasible due to the nature of the board, so the search space must be reduced. In order to do that, is common to follow one of these approaches: (a) by position evaluation truncating the search tree state and replacing the subtree below s by an appropriate value function that predicts the outcome from state s. However this approach it was believed to be intractable to Go due to the complexity of the game, or (b) reducing the search by sampling action from a policy, that is a probability of distribution over the possible moves (a) in a position (s). This paper follows the second approach but instead of using sophisticated Monte Carlo algorithms for tree search as most of the commercial solutions, they use a combination of Deep Convolutional Neural Networks with the Monte Carlo rollouts. The training for the CNN's consists in a pipeline of several stages of machine learning:

1.  They start by training by training a supervised learning (SL) policy network, directly from expert human moves. The inputs s to the policy network is a simple representation of the board state. It is trained randomly sampled state action pairs (s, a) using SGD (stochastic gradient descent). The convolutions is 13 layer deep.
2.  Train a reinforcement learning (RL) policy network that improves the SL (step 1) by optimizing the final outcome of games of self-play.
3.  Finally they train a value network that predicts the winner of the games played by the RL policy network (3) against itself.

In order to evaluate AlphaGo they ran a tournament among AlphaGo variations and other Go programs including the strongest commercial software, most of them, based on MCTS algorithms. AlphaGo defeated them with a 99.8% rate. Besides, to make it more challenging, they also played games with four free moves to the opponent and still the winning rate were over the 75%. A distributed version of AlphaGo played against the last years European Go championship Fan Hui in a formal five game match, winning all the rounds.

# Conclusion

The new search algorithm proposed by DeepMind Team is a new approach to computer Go, based on a successful combination of deep neural networks (convolutions) evaluations and tree search, is able to play at the level of the strongest human players, and it has been revealed as one of the great achievements in AI in the recent years. During the match against Fan Hui, AlphaGo evaluated thousands of times fewer positions than Deep Blue did, using the policy network and evaluating them more precisely using the value network, and approach that is perhaps more close to how a human plays.