Author: Ilton Garcia dos Santos Silveira
Date: December, 06, 2017

AlphaGo Research Review

It paper's goals was to show how DeepMind defeated the best AlphaGo player introducing a new technique approach, minimizing the enormous search space of Go created by traditional AI methods, which create a search tree for each possible position. Using Deep Learning principles, the DeepMind implemented two deep convolution neural networks (CONVNET). One CONVNET, called "value network", predicts how likely a move can lead to win after optimal moves. The other, called "policy network", helps to reduce the breadth of the searching move to play & using Monte Carlo Tree Search (MCTS), that effectively selects actions by look-ahead search it could "prune" even more the search tree.

This paper introduced many AI training perspectives: 1. Supervised Learning using human skills to train the CONVNET; 1. Reinforcement Learning, the system played against different versions of itself, learning from its mistakes; 1. Finally the evaluation function was created using Reinforcement Learning using self-play data set.

AlphaGo trained a 13-layer policy network, from 30 million positions from the KGS Go Server. The network predicted expert moves on a held out test set with an accuracy of 57.0% using all input features, and 55.7% using only raw board position and move history as inputs, compared to the state-of-the-art from other research groups of 44.4% at date of submission24, giving 2% of error average to improve the performance. Small improvements in accuracy led to large improvements in playing strength; larger networks achieve better accuracy but are slower to evaluate during search. This was why AlphaGo trained a faster but less accurate rollout policy $p\pi(a|s)$, using a linear softmax of small pattern features with weights π; this achieved an accuracy of 24.2%, using just 2µs to select an action, rather than 3ms for the policy network, now giving 20.2% of accuracy to be 15.000 times faster.

As final result, it program defeated 99.8% of others Go programs and even defeated the human European Go champion, so it proves that sometimes we don't need the most accurate system, we can give away of some accuracy & it can be better than human & using the available time instead spend so much more time than we can wait for a "simple" prediction.