AlphaGo Research Review

by Menghe Lu

This paper introduces AlphaGo, a new approach for training an artificial intelligence to play the game of Go, a problem which has been notoriously difficult to address due to the dimension of the search space. Through a combination neural networks and tree search, the authors are able to improve the state of the art and eventually beat a professional player in a formal game.

Neural networks are the first component of this approach. Inspired by recent advances of deep learning in image processing, the authors use stacked convolutional layers to learn representations of the 19 by 19 board. The networks are used to learn a policy (i.e. find the most probable moves from a given position) or to directly learn to evaluate a configuration of the game (i.e. quantify how much it is possible to win from that position). How to train them is the key to success in this paper.

First, a first policy network is training by supervised learning using 30 million positions from the KSG Go Server and trying to predict the next move. A smaller but faster network is trained the same way although less accurate, it will be used a quick guide to tree search.

The second stage of training uses a network with the exact same shape, initialized as the network obtained by supervised learning. This policy network is trained using reinforcement learning by playing against different versions of itself.

The last stage of training consists in learning to evaluate a position. This is modeled as approximating the expected reward from a given position. Instead of relying on the KSG dataset (which resulted in overfitting at this stage), the model is learnt from self-play using the policy network obtained at the second stage. The resulting network is called value network.

Finally these networks are used to guide tree search during the play. Tree search is a classical approach to solve games of perfect information, it consists in finding the current best move by enumerating the different possibilities and outcomes. Because it is intractable to explore the complete tree, AlphaGo guides the search by using a policy network to restrain an in-depth exploration (reaching terminal state) and a value network to quantify the quality of a position without looking ahead. The final strategy is a tradeoff between these two approaches.

AlphaGo has been evaluated against other IA and professional players, beating its opponents most of the time, which was an unprecedented breakthrough in the domain of artificial intelligence.