

AlphaGo

Introduction

AlphaGo is a Neural Networks System. It was trained by a novel combination of supervised learning from human expert games, and reinforcement learning from games of self-play. This article also introduce a new search algorithm that combines Monte Carlo simulation with value and policy networks.

The article talks about how to reduce the search space of Go by two general principles:

1. The depth of the search may be reduced by position evaluation.
2. The breadth of the search may be reduced by sampling actions from a policy $p(a|s)$ that is a probability distribution over possible moves “a” in position “s”(state).

AlphaGo include 3 stages of the training pipeline:

1. Supervised learning of policy networks.
2. Reinforcement learning of policy networks. (Unsupervised)
3. Reinforcement learning of value networks. (Unsupervised)

To evaluate AlphaGo, Deep Mind team ran an internal tournament among variants of AlphaGo and several other Go programs.

Result

This is the first time that a computer program has defeated a human professional player in the full-sized game of Go, a feat previously thought to be at least a decade away. And using this search algorithm, AlphaGo achieved a 99.8% winning rate against other Go programs, and defeated the human European Go champion by 5 games to 0.

By combining tree search with policy and value networks, AlphaGo has finally reached a professional level in Go, and providing hope that human-level performance can now be achieved in other seemingly intractable artificial intelligence domain (eg. Star Craft).