Research Review of Google's AlphaGo

Yazhuo Li

1 Introduction

The paper demonstrated the new algorithm introduced in AlphaGo program to play the game Go. The algorithm is based on deep neural "policy networks" and "value netowrks". It helped AlphaGo achieved a high winning rate against other Go programs and defeated human professional player for the first time on a full-size game Go board.

2 Techniques

2.1 Training Pipeline

AlphaGo uses a training pipeline consists of "value networks" and "policy networks", combined with the traditional Monte Carlo Tree Search to achieve high winning rate. It consists of the following stages:

- Build a supervised learning (SL) policy network based on expert human moves. At the same time, build another faster but less accurate roll out policy. These networks are used to predict human expert moves in a data set of positions.

- Use reinforcement learning (RL) to improve policy network through self-play. The RL policy network constructed is identical to SL policy network. The self-play happens between the current policy network and a randomly selected previous iteration of the policy network. This helps to optimize the policy network to the final outcomes of the game.

- Last stage of the pipeline is a value network that focuses on position evaluation and predicts the final outcome of the games played by the RL policy against itself.

2.2 Search Algorithm

The newly introduced search algorithm used in AlphaGo combines Monte Carlo Tree Search (MCTS) with the policy and value networks that are based on deep neural networks of many layers. These policy and value networks help the traditional MCTS to narrow search actions to the highly probable moves and to sample actions during roll-outs.

The training pipeline of valued networks and policy networks here are used in combination with MCTS to select, expand and evaluation moves.

3 Results

AlphaGo game engine has achieved outstanding result against both computer and human players.

Single-machine AlphaGo has won 494 out of 495 games (99.8%) against other computer Go programs. It also won 77%, 86% and 99% of handicap games against Crazy Stone, Zen and Patchi respectively - some of the most strong Go computer programs so far.

AlphaGo also beat European Go Champion Fan Hui, winning the match 5-0, the first time that a computer program has beaten a professional human player on a full-size board.