PROJECT: AIND-ISOLATION SHIV SHAH

Mastering the Game of Go with Deep Neural Networks and Tree Search

Summary of the goals and techniques

This paper comes from a leading Artificial Intelligence Research team working under Google called DeepMind. The paper focuses on designing a program called AlphaGo that can achieve super human intelligence in playing a Chinese Board Game called Go. Even though there have already been several Go engines that can play this game at a human level performance, none of them exceeds human capability. Here, the problem is attributed to the sophisticated nature of the game with vast number of possibilities or branching factor, making it difficult to evaluate positions and traverse the search tree. While Chess has been mastered using evaluation function and Monte-Carlo search, this approach creates only weak amateur level play in Go. So, the goal of the paper is to introduce new techniques for solving this grand challenge.

The techniques that are introduced are based on deep neural networks that are trained by the combination of supervised learning from human expert games and reinforcement learning from games of self-play. This new approach uses value networks to evaluate board positions and policy networks to select moves. Additionally, a new search algorithm is introduced that combines this training with the Monte-Carlo Tree search. Training process takes place using a pipeline consisting of several stages of machine learning. At the first stage of training, a database of human expert moves is used to train a Supervised Learning(SL) policy network. This learning allows the policy network to predict the accurate moves. The second stage includes training a Reinforcement Learning(RL) policy network that improves the SL policy network by optimizing the final outcome of games of self-play. This helps the policy network to focus not only towards the accurate moves but also towards the correct goal of winning games. The final stage of training focuses on position evaluation estimating a value function that predicts the outcome for a given position. This is achieved by training a value network that predicts the winner of games played by the RL policy network against itself. In the end, AlphaGo efficiently combines the policy and value networks with Monte Carlo Tree Search that selects actions by lookahead search.

Result

The results of using these techniques was quite remarkable. When AlphaGo played tournaments against several other strongest Go programs based on MCTS algorithms on a time limit of 5 seconds, it proved to be stronger than all of them, winning 99.8% of the time. Even when those programs were allowed 4 free moves at the beginning, AlphaGo was able to beat them at least 77% of the time. The distributed version of AlphaGo faced Fan Hui, the winner of 2013, 2014, and 2015 European Go championships, in a five-game match, and the result was 5 to 0 in favor of AlphaGo. During the match, its gameplay was closer to humans, selecting positions intelligently, using the policy and value networks and evaluating thousands of times fewer positions than Deep Blue did in its chess match against Kasparov. This was the first time that a computer Go program had defeated a human professional player, in the full game of Go, achievement that was believed to be at least a decade away.