Mastering the Game of Go with Deep neural networks and Tree Search

Author: Erik Taheri

AlphaGo is an AI developed by Google Deepmind to play the board game Go. The game of Go presents a unique challenge to artificial intelligence due to its size and complexity. Similar to board games like chess, a computer is unable to reasonably search the game board in its entirety due to time and processing limitations.

Prior to Alpha Go, the best AI players used a method called Monte Carlo tree search. Monte Carlo alone, would not provide the means necessary for the Deepmind team to develop AlphaGo. Due to the recent advancement and power of deep neural network, the Deepmind team a had a valuable tool to significantly improve prior solutions to playing Go. Unfortunately, neural networks alone would not show any significant improvement over the current Monte Carlo solutions. Despite the independent limitations of Monte Carlo and deep neural networks, the Deepmind team was able to create a new algorithm to solve this problem. By combining the strengths of the Monte Carlo simulation with the advancements in deep neural network AlphaGo was able to win 99.8% of games against other Go AIs and even was able to defeat professional human players.

The Deepmind team trained their neural network using several stages of machine learning. The team began by training a supervised learning policy network. This network was trained directly from expert moves from 30 million positions from the KGS Go Server. Supervised learning allows for immediate feedback and to create the most optimal output. Once the supervised network was trained, they began training a reinforcement learning policy network to improve the supervised network. The supervised policy maximizes the the AIs predictive accuracy whereas the reinforcement policy maximizes the the goal of winning the game. The RL policy was able to defeat the SL policy in 80% of the games. Finally, the Deepmind team trained the reinforcement learning value network. Unlike the RL policy network which outputs a probability distribution, the value network outputs a single prediction. Using the trained policy and value networks AlphaGo combines the networks in a Monte Carlo Tree search to determine its actions using a lookahead search.

AlphaGo was evaluated by playing against variants of itself, other go programs, and finally a human professional player. AlphaGo was able to win 99.8% of games against other programs and was able to win more than 77% of games with a handicap. AlphaGo was also able to beat a professional Go player, Fan Hui, 50 games to 0. AlphaGo was ultimately able to perform at the level of the strongest humans. This achievement marks a major milestone in artificial intelligence. The ability for an AI to perform at the level of the most skilled humans is considered the pinnacle of AI innovation.