AlphaGo by the DeepMind Team: A Researh Review

Summary of the paper

The Game of Go is a board game for two players in which the aim is to win more territory than the opponent. Each player places a black or white stone in the board. For a long time the Game of Go was thought to be a very hard problem to solve due to the vast search space(19×19 lines). This paper introduces a new approach with value and policy neural networks: Value networks evaluate board states and policy networks select moves. AlphaGo also takes advantage of human expert games and experience playing against itself. It achieves a very high win rate (99.8%) against other programs and also won against a professional Go Player. AlphaGo results is likely a revolution in the AI field because it's the first time that a game of such complexity is conquered with deep learning and new AI techniques.

New techniques introduced

Until now, in Games like Chess,Othelo etc. ,we try to find the optimal value function often with using a policy function (A probability distribution over legal actions). We can find the optimal value function (determines how good is each move) using minimax or DFS with alphabeta prunning. However, in Go these haven't been as effective. Until this paper, Monte Carlo tree search (MCTS) and reinforcement learning has been used and AlphaGo improves on those techniques. Monte Carlo tree search determines the value function by approximations. In AlphaGo slower but more powerful representations of the value and policy functions are used. Reinforcement learning improves the value function by playing games against itself. AlphaGo uses Reinforcement learning with Neural Networks creating a combined supervised learning and a reinforcement learning policy networks, the first one improving the last one. In short, AlphaGo uses a neural network training pipeline in which each stage improves the policy and value functions.

More specifically, the process begins by training a supervised policy network (slow) from human expert moves. Then a reinforcement learning policy (faster) network is trained that improves the previous network by self-play. In the final step of the pipeline there's a value network that predicts the winner of the reinforcement network. AlphaGo combines the policy and value networks in an MCTS algorithm that selects actions by lookahead search.

Key results

The research demonstrated that value networks provide a viable alternative to Monte Carlo evaluation in Go although the mixed evaluation achieves higher win rates. This shows that the two position evaluation networks are complimentary. Also, AlphaGo won 5-0 games against the European Champion player, **Fan Hui** and achieved a high win Rate against other programs. Until now this was a feat thought at least a decade away and it gives hope to find solutions to other similar hard problems. The game of Go is thought as a game that requires creativity. Until now, many people believed that creativity could no be demonstrated by computers.