Review on the Research Paper "Mastering the game of Go with deep neural networks and tree search"

First I will give a short summary on the goals of the paper and afterwards I describe the results the researchers have found.

Summary of the paper's goals or techniques

The game of Go is a very difficult game because of the enormous search space and the difficulties of evaluation the board positions and moves. Therefore this game is a tremendous challenge on Artificial Intelligence.

Researchers have used a combination of deep neural networks that were trained on supervised learning from real human plays as well as on gameplays from self-play by reinforcement learning.

To overcome the problem of complexity researchers used a multilayer AI-architecture. They combined deep convolutional nets and trained thirteen layers of a supervised learning policy network. This supervised learning policy was able to predict about 57% using all input features and 55.7% using only a raw board position. This was a great improvement in comparison to the 44.4% that was state-of-art then.

On top of the supervised learning policy they stacked a reinforcement learning policy by policy gradient learning. When they played against the supervised policy they won more than 80% of the games.

On the final training stage they included position evaluation that predicts the outcome from position s of games played by using policy p for both players.

For searching researchers included a search tree. Each edge (s, a) of the search tree stores an action value Q(s, a), visit count N(s, a), and prior probability P(s, a). The tree is then traversed descending the tree - starting from the root tree. When the traversal has reached a leaf node, it may be expanded.

At the end of the simulation the action values and visit counts of all traversed edges are updated.

Because this evaluation policy and value networks consume several orders of magnitude more computation than traditional search heuristics, AlphaGo uses an asynchronous multi-threaded search that executes simulations on CPUs, and computes policy and value networks in parallel on GPUs. The final version of AlphaGo used 40 search threads, 48 CPUs, and 8 GPUs. They also evaluated a distributed version of AlphaGo that exploited multiple

Summary of the paper's results

With the described multi-layer architecture researchers achieved a winning rate of 99.8% against other Go programs (they have won 494 out of 495 games)

Lastly this AI-Go also won against Fan Hui - the winner of the European Go championships in 2013, 2014 and 2015. This AI-Go won 5 games out of a total of 5 games in a classic Go-Match.