RESEARCH REVIEW ON A PAPER ON ALPHAGO

Paper read: Mastering the game of Go with deep neural networks and tree search.[1]

PROBLEM DESCRIPTION

In theory Go is a game of perfect information and a search tree containing b^d sequences of moves, where b is the game's breadth (number of legal moves per position), and d is it depth (games) should tell us the perfect play in each position. Unfortunately, for Go b is around 250, and d is around 150. Exhaustive search is not possible. [2]

PAPER'S GOALS AND TECHNIQUES INTRODUCED

This paper aims at overcoming the problems of enormous search space and the difficulty of evaluating board positions and moves. The new approach introduced uses "value networks" to evaluate board positions and "policy networks" to select moves. The training of these deep neural networks is done by a combination of supervised learning from human expert games and reinforcement learning from games of self-play. A new search method is also used that combines Monte Carlo simulation with value and policy networks.

A summary of the procedure used in the process of making AlphaGo game playing agent:

1. Training a supervised learning policy network directly from expert human moves to provide fast, efficient leaning updates with immediate feedback and high-quality gradients.
2. Training a fast policy that can rapidly sample actions during rollouts.
3. Training a reinforcement learning policy network that improves the supervised learning policy network by optimizing the final outcome of games of self-play.
4. The advantage of the above move is that it adjusts the policy towards the correct goal of winning games rather than maximizing predictive accuracy.

Sited below are the list of steps taken to train the AlphaGo's game playing agent:

Step 1: Supervised learning of policy networks.

Step 2: Reinforcement learning of policy networks

Step 3: Reinforcement learning of value networks: game played between the RL policy network and itself until the game terminated which helped in the reduction of mean squared errors.

Step 4: Searching with policy and value networks

RESULTS

The results that the authors achieved as listed in the paper are:

1. Against the supervised policy network the reinforced learning policy network won more than 80% of games.
2. Against the strongest open-source Go program, Pachi, the reinforced learning policy network alone won 85% of the times while supervised learning won only a 11% of the times.
3. Supervised learning policy network performed better in AlphaGo than the stronger reinforcement learning policy network presumably because humans select a diverse beam of promising moves, whereas reinforced learning optimizes for the single best move.
4. Even without rollouts AlphaGo exceeded the performance of all other Go programs, demonstrating that value networks provide viable alternative to Monte Carlo evaluation in Go.
5. The mixed evaluation (value networks with rollouts) performed best by winning more than 95% of games against other variants.
6. AlphaGo achieved a whopping 99.8% winning rate against other Go programs.
7. AlphaGo also won 5 games to 0 against the European champion of Go in a full-sized game of Go.

REFERENCES

[1] Mastering the game of Go with deep neural networks and tree search David Silver, Aja Huang, Chris J. Maddison , Arthur Guez , Laurent Sifre , George van den Driessche , Julian Schrittwieser , Ioannis Antonoglou , Veda Panneershelvam , Marc Lanctot , Sander Dieleman, Dominik Grewe , John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach , Koray Kavukcuoglu , Thore Graepel & Demis Hassabis

[2] https://blog.acolyer.org/2016/09/20/mastering-the-game-of-go-with-deep-neural-networks-and-tree-search/