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

Thilina Shihan Weerathunga

June 2017

1 Abstract

In the field of Artificial intelligence the game of Go is considered as the most challenging classical game due to its extremely large search space and complicated nature of evaluating board positions and moves. In this paper, authors introduce a new approach to computer Go player which uses 'value networks' to evaluate board position and 'policy networks' to select moves. These value networks and policy networks are deep neural networks (DNNs) which are trained by combination of supervised learning from human expert games and reinforcement learning from games of self-play.The authors introduce a new algorithm which combines Monte Carlo simulation with value and policy neural networks that achieved a 99.8% winning rate against other existing Go programs.

2 Methodology

Games with perfect information have optimal value function $v^*(s)$ which is a function of board position or state s. The game may be solved by recursively calculating $v^*(s)$ in a search tree. The dimensions or the possible sequence of moves of the search tree is given by b^d, where b is the game's breadth and d is the depth of the tree. Here breadth represents the number of legal moves for per position and the depth represents the length of the game. For the game Go b^d creates extremely large search space ($b \approx 250$ and $d \approx 150$) which makes exhaustive search unrealistic.

However, the depth of the search tree can be reduced by position evaluation which will truncate the search tree at state s and replacing the sub tree below s by approximating the optimal value function. Furthermore, the breadth of the search tree can be reduced by introducing a new sampling strategy which comes from a well defined policy distribution $p(a|s)$. Here, a represents possible moves in state s.In the AlphaGo algorithm, authors uses two deep neural networks named value network and policy network. The position evaluation is done using the value network and sampling strategy is decided by using the policy network.These two networks are trained using the following strategies.

- Supervised learning of policy networks.
- Reinforcement learning of policy networks.
- Reinforcement learning of value networks.

Finally these two network outputs are integrated with Monte Carlo tree search (MCTS) to play the game Go.

3 Conclusion

A new algorithm (AlphaGo) has been developed to play the game Go using DNNs and Monte Carlo search trees. The DNNs are trained by using combination of supervised and reinforcement learning.Outputs from the neural networks are combined with Monte Carlo rollouts to produce the final output value.

AlphaGo has achieved a 99.8% winning rate against other existing Go programs and won against the human European Go champion by 5 games to 0. This is the first time that computer player wins against a human professional in the full-sized game of Go. Also during the match, AlphaGo evaluated significantly fewer position than Deep Blue did when it was playing chess against the world champion Kasparov. This performance is achieved by selecting positions more wisely using the policy network and evaluating them precisely using the value network.

Previously, Go game achieved satisfactory results by including Monte Carlo tree search strategies. After integrating deep neural network strategies with Monte Carlo tree search, AlphaGo gives exceptional results which exceed human level performances in complicated artificial intelligence domains.

References

[1] doi:10.1038/nature16961