

Summary of research paper

”Mastering the game of Go with deep neural networks and tree search”

Dovydas Čeikutka

November 24, 2017

In one of the seminal artificial intelligence papers (Silver et al. 2016) introduces AlphaGo - an intelligent agent designed to play Go using a novel approach.

1 Summary of the techniques introduced

The paper (Silver et al. 2016) introduces a novel method of training the intelligent agent using a three step process:

1. supervised learning of policy networks
2. reinforcement learning of policy networks
3. reinforcement learning of value networks

1.1 Supervised learning of policy networks

The supervised learning of policy networks consists of training a 13-layer neural network consisting of alternating convolutional layers and rectifier nonlinearities with a softmax layer at the end. The neural network is trained using 30 million positions from the KGS Go Server.

1.2 Reinforcement learning of policy networks

In the next stage the policy network is improved using policy gradient reinforcement learning. The reinforcement learning policy network has identical structure to the supervised learning policy network. The weights are also initialized using the values from the previous step. Then games between random policy from previous iteration and current policy are played to prevent overfitting to the current policy and weights are updated.

1.3 Reinforcement learning of value networks

In the final stage the value function is estimated for position evaluation. Optimal value function under perfect play is not obtainable, therefore the value function is estimated for the strongest policy using reinforcement learning.

2 Summary of the results

In order to evaluate the results of the AlphaGo agent several tournaments were ran against other Go programs and also against Fan Hui - professional Go player and winner of multiple European Go championships. AlphaGo's performance was surprising as it was believed that human-level artificial intelligence in the game of Go was decades away. This result inspires hope that artificial intelligence will reach human-level performance in other domains.

2.1 Results vs. other Go programs

AlphaGo played these Go programs:

1. Crazy Stone
2. Zen
3. Pachi
4. Fuego
5. GnuGo

All contenders were allowed 5s of computation time per move. The new techniques allowed single-machine AlphaGo to win 494 out of 495 games. Distributed AlphaGo proved to be even stonger and won all the games agains other programs and 77% of the games agains single-machine AlphaGo.

2.2 Results vs. Fan Hui

AlphaGo played a formal five-game match, which AlphaGo won 5-0. It is worth noting that during the match AlphaGo evaluated thousands of times fewer positions that Deep Blue did in tis chess match agains Kasparov - a feat which was achieved by selecting the positions more intelligently.

References

Silver, David et al. (2016). “Mastering the game of Go with deep neural networks and tree search”. In: *Nature* 529.7587, pp. 484–489. ISSN: 0028-0836. DOI: [10.1038/nature16961](https://doi.org/10.1038/nature16961). URL: <http://dx.doi.org/10.1038/nature16961>.