AlphaGo

By DeepMind Team

Overview

Go is considered as most challenging AI game in the world due to its enormous search space and difficulty of evaluating board positions and moves. AlphaGo is the AI computer programme that plays the game GO. It was developed by Alphabet's Google DeepMind in October 2015.

How

AlphaGo introduced a new algorithm which was a combination of the algorithm Monte Carlo and value and policy network. Using this newly developed algorithm AlphaGo defeated human European Go championship by 5 games to 0 and have 99.8% winning rate compare to other Go programs.

Milestones

The complete milestone of state of the art technology was achieved by a number of the steps

1. Supervised learning of Policy networks

   In this stage policy networks were trained using supervised learning to achieve a better accuracy. This was also included with the training of faster but less accurate rollout policy with a linear softmax of small features with weights to achieve an accuracy of 24.2%.

2. Reinforcement learning of Policy networks

Aimed at improving the policy network by policy gradient reinforcement learning. In this stage it plays game between randomly selected policy network and previous selected policy network, by this way it improves the network by stabilizing the network and by preventing overfitting.

3. Reinforcement learning of Value networks

This stage is focused on position evolution and estimating the position outcome from the game played, Hence it predicts the value function for strongest policy using RL policy network.

Evaluation

AlphaGo team evaluated the overall performance of the Go programme by running the internal tournament and measuring the Elo rating of the programmes. All the programmes received the maximum of 5s computation time per move. The probable estimation of one programme beats other was calculated using **logistic function** and estimation of ratings was done by the Bayesian logistic regression.

With the exception of distributed AlphaGo, each computer Go program was executed on its own single machine, with identical specifications.

The match against Fan Hui was arbitrated by an impartial referee. Five formal games and five informal games were played with 7.5 komi, no handicap, and Chinese rules. AlphaGo won these games 5−0 and 3−2 respectively.

Following are the results of the different Go programmes played in internal tournament-

Short name	Computer Player	Version	Time settings	CPUs	GPUs	KGS Rank	Elo
α_{rvp}^{d}	Distributed AlphaGo	See Methods	5 seconds	1202	176	–	3140
α_{rvp}	AlphaGo	See Methods	5 seconds	48	8	–	2890
CS	CrazyStone	2015	5 seconds	32	–	6d	1929
ZN	Zen	5	5 seconds	8	–	6d	1888
PC	Pachi	10.99	400,000 sims	16	–	2d	1298
FG	Fuego	svn1989	100,000 sims	16	–	–	1148
GG	GnuGo	3.8	level 10	1	–	5k	431
CS_4	CrazyStone	4 handicap stones	5 seconds	32	–	–	2526
ZN_4	Zen	4 handicap stones	5 seconds	8	–	–	2413
PC_4	Pachi	4 handicap stones	400,000 sims	16	–	–	1756

Each program played with a maximum of 5 s thinking time per move; the games against Fan Hui were conducted using longer time controls, as described in Methods. CN_4, ZN_4 and PC_4 were given 4 handicap stones; komi was 7.5 in all games. Elo ratings were computed by BayesElo.

By combining tree search with policy and value networks, AlphaGo has finally reached a professional level in Go, providing hope that human-level performance can now be achieved in other seemingly intractable artificial intelligence domains.