RESEARCH REVIEW

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

Antonio Rincon

Contents

1. Goals and techniques 3

2. Results 3

1.Goals and techniques

The game of Go was considered as the most challenging of classic games for artificial intelligence due to its enormous search space 250^{150} , calculated by its breath (number of legal moves per position) and depth (game length), and the difficulty of evaluating board positions and moves. Current state of the art algorithms based on Monte Carlo tree search (MCTS) only achieved amateur performance, it was considered a problem to be solved in 10 years to achieve superhuman performance.

The new approach used by AlphaGo presented in this article uses deep neural networks. Neural networks have been used in visual domains with unprecedented performances where, neural networks using many layers of neurons to construct abstract representations of an image. A similar architecture for the game of Go has been used, the board was represented as a 19x19 image and convolutional layers has been used to construct a representation of the position. Neural networks reduce the effective depth and breadth of the search tree evaluating positions using a value network, and sampling actions using a policy network.

These deep neural networks have been trained by a combination of supervised learning from human expert games, and reinforcement learning from games of self-play.

For the first stage of the training pipeline, policy networks where trained using supervised learning to predicting expert moves in the game of Go. The SL policy network alternates between 13 convolutional layers with weights and rectifier nonlinearities, the final softmax layer outputs a probability distribution over all legal moves, the network was trained from 30 million positions from the KGS Go Server, more complex networks achieve better accuracy but were slower to evaluate during search. While evaluating this network a 55.7% accuracy was achieved, compared to the previous 44% accuracy obtained by previous state of the art research groups.

The second stage of the training pipeline aims at improving the policy network by reinforcement learning (RL). The RL policy network is identical in structure to the SL policy network. The RL policy network won 85% of games against other Go algorithms, compared to the previous state-of-the-art based only on supervised learning that only won 12%.

The final stage of the training pipeline focuses on training value networks using reinforce learning. This neural network has a similar architecture to the policy network, but outputs a single prediction instead of a probability distribution.

2.Results

Using the search algorithm, the program AlphaGo achieved a 99.8% winning rate against other Go programs, including variants of AlphaGo, commercial programs Crazy Stone and Zen and open source programs Pachi, Fuego and GnuGo, that used MCTS and state-of-the-art search methods that preceded MCTS.

AlphaGo also defeated the human European Go champion Fan Hui by 5 games to 0. This was the first time that a computer program defeated a human professional player in the full-sized game of Go, a goal previously thought to be at least a decade away.