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

paper's goals

1 - They introduce a new search algorithm that combines Monte Carlo simulation with value and policy networks.

2 - The strongest current Go programs are based on MCTS(Monte Carlo tree search), enhanced by policies that are trained to predict human expert moves in order to narrow the search to a beam of high-probability actions, and to sample actions during rollouts.

paper's results

By Using this search algorithm, their program AlphaGo achieved a 99.8% winning rate against other Go programs, and defeated the human European Go champion by 5 games to 0 and this is the first time that a computer program has defeated a human professional player in the full-sized game of Go, a feat previously thought to be at least a decade away.

Steps:

1 - Supervised learning of policy networks: -

A - They build on prior work on predicting expert moves in the game of Go using supervised learning.

2 - Reinforcement learning of policy networks:-

A - The second stage of the training pipeline aims at improving the policy network by policy gradient reinforcement learning.

3 - Reinforcement learning of value networks:-

A - The final stage of the training pipeline focuses on position evaluation, estimating a value function vp(s) that predicts the outcome from position s of games played by using policy p for both players.