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

Original Article: https://storage.googleapis.com/deepmind-media/alphago/AlphaGoNaturePaper.pdf

AlphaGo Techniques

The AlphaGo algorithm combines Supervised Learning (SL) from human Go champions, Reinforcement Learning (RL) from self played games, the Monte Carlo Search Tree (MCST) with Policy Network (PN) and Value Network (VN). The combination achieves the highest win rate and performance to date.

SL PN has 13 layers, trained from 30 million positions and provides fast instant feedback & accurate moves, by alternating between convoluted layers with weights and rectifier non linears.

PN alone receives simple board state representation. It reduces breadth of search by narrowing it down to sample actions of high probability during fast rollouts. A final softmax layer outputs probability for all legal moves.

RL PN optimizes final outcome of self played games. Policy adjusts toward the correct goal of winning the game - instead of the incorrect goal of maximizing predictive accuracy of the current move. It predicts outcome for both players and outputs a single prediction.

RL VN predicts winner against itself.

AlphaGo Results

AlphaGo achieves a 99.8% win rate, minimizes computational time and improves performance (speed & accuracy) compared to traditional algorithms and the world's top Go game champions and older algorithm solutions.

Prepared May 27th by monigarr@monigarr.com