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

Introduction

Board games can be split into 2 groups. The first, perfect information games, and the second which is the opposite, partial information games. In perfect information games, we can technically measure the outcome of the game for all possible configurations of the game board. In some games, measuring the value of all possible board states is infeasible due to the large number of possible moves and hence restricted by the high computational capacity required.

AlphaGo Goals & Techniques

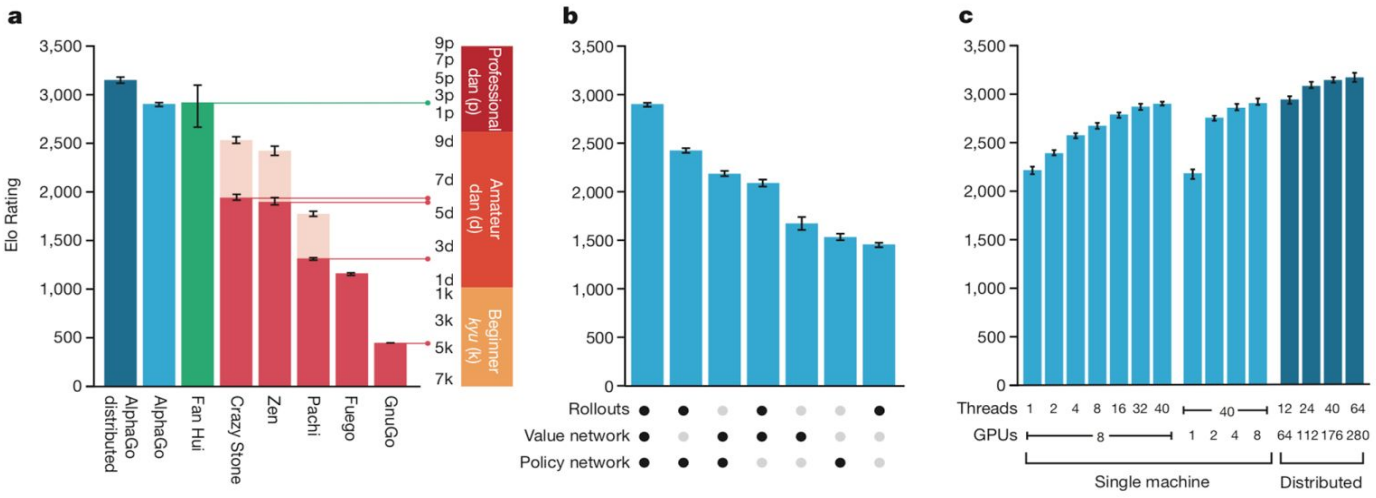
Go is a game which has a very high branching factor. This means for every possible move, there are many possible responses from the opposing player. We can measure the branching factor in board games by using the formula b^d where *b* = *breadth* and *d* = *depth* of the game tree. In chess, b = 35 and d = 80, while in Go, b = 250 and d = 150, which means an exhaustive search is not feasible.

To achieve a strong game playing agent, the AlphaGo paper introduces multiple policy networks and a value network which help tackle the high branching factor in Go. The networks are used as follows:

1. A supervised learning *policy network* is trained using expert human moves. This provides the AI agent with immediate feedback and fast high quality decisions.
2. A fast *rollout policy* which samples actions during rollouts. This network is also trained to predict human expert moves.
3. Next, a *Reinforcement Learning (RL) policy network* is introduced along with a mechanism to allow self-play. In this step the agent plays the entire game against itself to update the weights in the RL policy network.
4. Finally, a value network is trained to predict the winner based on actions taken by the RL policy network in Step 3.

AlphaGo Results

The Go playing agent was able to reach professional level of playing Go and even beat the european champion (Fan Hui - 2 dan). Later iterations of the AlphaGo agent was able to reach higher levels and beat the world Go champion (Lee Sedol - 9 dan).



Interesting Notes:

- AlphaGo used CNNs to visually understand the game and hence attempt to get a more intuitive understanding of the game.
- Improvements to AlphaGo could include Recurrent Neural Nets (RNNs) which are also improved using the RL policy network mentioned above.

References:

- [AlphaGo Nature Paper]
- [AlphaGo Website]