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

By Michel Angeles Ortiz

In 2016, DeepMind released AlphaGo, the first Go program capable of beating a professional human Go player.

1 Goals

The goal of AlphaGo was to develop a program capable of reaching a professional level on Go, which due to the enormous search space and the difficulty of evaluating board positions, had been a hard problem in Artificial Intelligence.

2 Techniques

AlphaGo reduces the effective search space of moves using two main principles:

- For evaluating a board position, the search tree is truncated at state s and replaced the subtree by an approximate value function $v(s) \approx v^*(s)$.
- Reducing the breadth of the search space by sampling actions from a policy $p(a \mid s)$, which is a distribution over all possible moves a in position s.

A combination of neural networks and tree search was used for achieving the mentioned tasks. The neural networks reduce the depth and breadth of the search tree. For evaluating position a value network is used, whereas for sampling action a policy network is used.

To summarize the whole pipeline used in AlphaGo, is the following:

1. A policy network $p_\sigma(a \mid s)$, made of convolutional layers, nonlinearities (relus), and weights σ is trained via supervised learning (SL), to predict expert moves. The input of this network is a representation of the board state s. The dataset used was taken from the KGS Go Server, 30 million of positions were used. This network was able to predict actions in 3ms.

   Similarly, another network p_π was trained to make faster predictions $2\mu s$, however this network was less accurate when predicting actions.
2. Another policy network p_ρ with identical structure as p_σ is initialized with the same weights $\rho = \sigma$. This network is trained via Reinforcement Learning (RL), this means that the network is able optimize its parameters by playing with opponents using a random pool of old policies of the same network.
3. A final neural network v_θ is trained to predict the winner of games played by the RL network against itself, using the strongest policy RL.
4. Monte Carlo tree search is used as a search algorithm in combination with the RL and SL networks.

3 Results

When playing against another Go programs, Alpha go was able to win in 494 out of 495 games (99.8 %). AlphaGo was also tested against the European Champion Fan Hui, in which AlphaGo took the victory with a score of 5-0, in October 2015.
In March 2016, AlphaGo played against Lee Sedol, one of the best Go player at the time, with a result of 4-1, respectively.
In October 2017, DeepMind published a new article, introducing AlphaGo Zero, a version that is able to learn without help of human data, by playing only against itself. AlphaGo Zero is able to beat any past versions of AlphaGo by a big margin.

4 Resources

Silver, David, et al. ”Mastering the game of Go with deep neural networks and tree search.” Nature 529.7587 (2016): 484-489.
Silver, David, et al. ”Mastering the game of go without human knowledge.” Nature 550.7676 (2017): 354-359.