## Review of the DeepMind's paper about AlphaGo

**Uirá Caiado**

In 2016, the DeepMind's AlphaGo defeated Lee Sedol, one of the world's best players of Go. This ancient game is a board game so complex that computers had not been expected to master it for another decade at least, according to [1]. In their paper "Mastering the Game of Go With Neural Network and Tree Search", [2] described the technical details of the system.

As explained in the Deepmind's blog post[1] about the computer program, due to the enormous search space of Go, traditional AI methods, which construct a search tree over all possible positions, are not suited to this board game. In general, we can reduce the search space in games using two principles: position evaluation and sampling actions from a policy. However, even though the application of both principles has provided superhuman performance in games like backgammon and Scrabble, as explained by [2], they only produced a weak amateur level play in Go.

Similarly to the approach described above, AlphaGo uses two deep convolutional neural networks in the place of the general principles. One neural network, the "value network", predicts how likely a move can lead to a win after a sequence of optimal moves, helping the program reducing the depth of the search. The other neural network, the "policy network", helps reduce the breadth of the search selecting the next move to play. Both neural networks are then combined with an advanced tree search called Monte Carlo Tree Search (MCTS) that effectively selects actions by look-ahead search.

This paper has introduced many different approaches to train the model. First, the policy network was trained directly from expert human moves, using supervised learning. This first stage helped AlphaGo develops its understanding of what reasonable human play looks like. Next, the policy network already trained by supervised learning was improved using reinforcement learning. The system played against different versions of itself, learning from its mistakes. Finally, the value network, which evaluates the value of a position of the program on the board, also was created using reinforcement learning using self-play data set.

After combining the policy and value network with MCTS, the paper reports that the program achieved 99.8% winning rate against other Go programs, and defeated the human European Go champion by five games to 0. AlphaGo is based on deep neural networks that are trained by a combination of supervised and reinforcement learning, avoiding the construction of handcrafted evaluation function, as was did for DeepBlue, for example. The search algorithm developed to the program could be applied to other domains, as general game-playing, scheduling, and constraint satisfaction.

# References

[1] The Economist. The future of computing, 2017.

[2] Silver, D, Huang, A, Maddison, C J, Guez, A, and Sifre, L. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016.

[1] Source: `https://deepmind.com/research/alphago/`