# AlphaGo paper review

By Daniele Storni

AlphaGo is an AI built by Google DeepMind to play board game Go. Before AlphaGo was released researchers in the field of artificial intelligence thought that we were about one decade away from having system with AlphaGo skills.

Go is a game of perfect information which means that each player is perfectly informed of all the events that have previously occurred. This games have an optimal value function through which it is possible to determine outcome from every state of the game assuming that each player makes the best possible move. To solve game of perfect information, agent has to evaluate each possible move by means of simulations which consists of search tree containing approximately b^d possible sequences of moves, where b is a number of possible legal moves and d is a game length. Approximate numbers are: $b \approx 35$, $d \approx 80$ for chess ($10^{123}$ different games) and $b \approx 250$, $d \approx 150$ for Go ($10^{360}$ different games). To evaluate game states the Monte Carlo tree search is used. Roughly speaking MCTS consists of on many playouts, where the game is played out to the very end by selecting moves at random. Results of each playout is then used to weight the nodes in the game tree so that better nodes are more likely to be chosen in future playouts. AlphaGo uses both policy and value networks in Monte Carlo search tree. In particular it contains three different CNN of two different types: two networks are policy networks and one is a value network. Both types of networks take as input the current game state, represented as an image. The policy networks provide guidance regarding which action to choose, given the current state of the game. To train policy network, there are a couple stages in machine learning: fisrt, a policy network was trained on 30 million positions from games played by human experts, which reached an accuracy of 57.0%. Than a smaller policy network is trained as well. This network has a lower accuracy around 24%, but it is much faster. The second part of the training stage is Reinforcement Learning: while Supervised Learning policy network is good in predicting next most likely moves, Reinforcement Learning helps with prediction of the best possible winning moves. In this phase to improve generalization capabilities and avoid overfitting AlphaGo played more than one milion games with itself.

## Results

- More computional resources deliver an higher performances: the distributed version of AlphaGo beat the single machine version 77% of the times.

- AlphaGo consistently beats the other AI designed to play GO as well as the european champion and even without using all of its networks, AlphaGo reaches performances comparable to the other AI systems..