

Mastering the game of Go without human knowledge

Previously, the best algorithm for playing Go was AlphaGo Lee. AlphaGo Lee defeated Lee Sedol, a world champion Go player, 4:1. AlphaGo Lee used a supervised learning based algorithm, it was trained on human data. The algorithm that is the subject of this paper, AlphaGo Zero, used a reinforcement learning based algorithm. It played against itself millions of times over the course of 40 days, and achieved a superhuman level of skill that was able to defeat AlphaGo Lee 100:0.

AlphaGo Zero uses a Monte Carlo tree search (MCTS) to select a sequence of moves to evaluate. Unlike a typical MCTS, AlphaGo Zero does not play out games until the end in each search. Instead the game state is evaluated by a neural network with one convolutional layer and 40 residual layers. The convolutional layer reduces the sparsity of the high dimensional data, making it easier to identify patterns. AlphaGo Zero was run on four tensor processing units, custom hardware created by Google for machine learning algorithms.

Machine learning algorithms have achieved superhuman performance in the game of Go.

A reinforcement learning algorithm performed better at Go than an equivalent supervised learning algorithm.

AlphaGo Zero was able to discover conventional strategies commonly used by experts and entirely new strategies without any information other than the rules of the game.

Questions

What are possible industry applications of this algorithm?

What improvements could be made to this algorithm?

What will reinforcement learning never be able to do?