Tianwei Zhu
Dec 22, 2017

Summary of AlphaGo Zero by the DeepMind

The DeepMind Team introduced a new Go program called AlphaGo Zero, which can master the game Go from the scratch without any human knowledge. AlphaGo Zero is based on a technology called reinforcement learning, which makes the computer to be trained from their own experience. The main idea is to use a neural network to predict better move selections while the computer practices Go with itself. In other words, the neural network is trained by a self-play reinforcement learning algorithm that uses Monte Carlos Tree Search to play each move.

There are also some exciting lessons learned from AlphaGo Zero. By discovering new strategies beyond the traditional ones, the program proved that the existent human Go knowledge is still limited. At the same time, it also showed that human and machines learn to play Go in fundamental different ways. AlphaGo uses two deep neural networks in a Monte Carlo tree search (MCTS): a policy network that outputs move probabilities and a value network that outputs a position evaluation. Reinforcement learning of policy networks try to improve the prediction of best moves and was trained to recognize patterns of play by experts. The value network, on the other hand, only calculates the chances of winning. It is used for evaluating moves during gameplay using MCTS. This value network is trained from games for which the outcome is already known, by using the strongest policy they have trained.

By comparing the performance of AlphaGo Zero with all previous versions such as AlphaGo Master, Lee and Fan, Zero outperformed all of them. By training with itself, AlphaGo Zero easily beat Alpha Go Lee in 3 days, with Elo rating of 3500. Then the Elo rating of Zero is gradually approaching to the AlphaGo Master and eventually surpass the it after 30 days. AlphaGo Zero keeps improving itself gradually with Elo rating higher then 5000. Another interesting discovery is that the performances of AlphaGo Zero and Master were similar. This is because both have used the same technology, the only difference is that Master used human data and features instead of self-playing.

In conclusion, AlphaGo Zero proved that the universal application of reinforcement learning technology is possible and can be implemented into many other different fields such as self-driving car technology. For the first time, this version of AlphaGo can learn from itself without any prior human knowledge. More importantly, it would allow to rediscover the beauty of Go by providing new inspirations and insights. If we looked back the history of AlphaGo, when it won with European champion Fan Hui. It was the very first time when machine defeated a professional in the game of Go. During the game with Fan Hui, AlphaGo evaluated a thousands times fewer moves than Deep Blue did in its chess match against Kasparov what suggests that positions are selected more intelligently. And this intelligence is growing with emerging improved versions of AlphaGo. All the conclusions clearly show that thanks to AlphaGo we are another step closer to the future.