Udacity Artificial Intelligence Degree: Build a game agent - Publication summary

Introduction

This document is part of the Udacity Artificial Intelligence degree program, in the "Build a game agent" project. The goal is to understand and summarize a publication related to game playing.

The publication chosen is "**Mastering the game of Go without human knowledge**" (The authors explain a Deep Learning program that has beaten the best humans and fully artificial systems in the Go game with a totally unsupervised learning.

The work was developed by DeepMind employees and was published in Nature International Journal of Science (http://dx.doi.org/10.1038/nature24270).

Methods

The work is based on a previous system that the authors named AlphaGo, also developed by DeepMind. It consists of a deep neural network based on a deep convolutional residual network architecture. Deep neural networks had enormous success in the last years, due to the improvement in computational hardware, specifically in Graphic Processing Units (GPUs). Deep architectures based mainly on convolutional and fully connected layers have dramatically improved the state of the art in many problems, including some of them that were unresolved until now.

The network needs to learn from databases with thousands of human experts moves. However, in order to make the best possible prediction in each move, it uses Monte Carlo simulations with a score based on supervised human knowledge. Monte Carlo is a tree search technique based on similar concepts as the ones studied during the Udacity lesson where this work is included. The program was able not only to improve the current state of the art programs, but it also achieved superhuman knowledge defeating consistently to the current world champion at the time, Lee Sedol, in March 2016.

AlphaGo Zero, the system described in this publication, takes the wole idea a big step forward. This system makes use of a completely unsupervised reinforcement algorithm, that allows it to learn without requiring any human knowledge. Basically, the network plays against itself, optimizing its parameters based on the final winner of the game.

Results

The results of the work look absolutely stunning to me. AlphaGo Zero is able to learn many of the best human strategies without the need of any human guide in a pretty short time and using less computational power than its predecessors. Quoting the authors: "After 72 hours, we evaluated AlphaGo Zero against the exact version of AlphaGo Lee that defeated Lee Sedol, under the same 2 h time controls and match conditions that were used in the man– machine match in Seoul (see Methods). AlphaGo Zero used a single machine with 4 tensor processing units (TPUs), whereas AlphaGo Lee was distributed over many machines and used 48 TPUs. AlphaGo Zero defeated AlphaGo Lee by 100 times to 0."

But not only that. When comparing both systems, the authors describe that even though the supervised programs learn slightly faster the best human strategies, in the overall AlphaGo Zero defeats all of them. This shows that AlphaGo Zero not only learns the best human strategies with almost the same performance as a supervised system, but it also develops new ones that achieve a better global performance in the game.

In my opinion, this is what makes this work extremely exciting. Even most of the current AI systems that outperform humans in a given task usually need to be trained using expert human knowledge to solve the problem. AlphaGo proves that an AI system may achieve superhuman performance in very complex problems without requiring any human knowledge. We are living extremely exciting and unsettling times…

Jorge Onieva.

January 27, 2018