Capstone Project Proposal

Deep Reinforcement Learning

November 11, 2017

1 Domain Background

1.1 General Setting

I propose to combine the two ML Nanodegree sections reinforcement learning and deep learning and undertake a project in deep reinforcement learning (DRL). The combination of the two disciplines has has generated a lot of interest lately, most notably when DeepMind's AlphaGo defeated the Go champion. It will very likely be a key ingredient in general artificial intelligence, which is one of the trends currently emerging in the tech sector. Given that the two most important companies in general AI research – DeepMind and OpenAI – are both using DRL and that both RL and Deep Learning have been covered in the Nanodegree, it is safe to say that this is a good subject for a capstone project.

1.2 Brief Historical Overview

The recent survey [Arulkumaran et al.(2017)] gives a great overview of the field of DRL. Reinforcement learning is a framework for experience–driven autonomous learning. Prior to the advent of deep learning, RL approaches lacked scalability and were limited to rather low–dimensional problems. Deep learning has helped RL to scale to decision–making problems that were previously intractable. The most notable successes of DRL have probably been

1. the development of an algorithm that could learn to play a range of ATARI 2600 video games at a superhuman level, just by looking at the pixels [Mnih et al.(2015)]

2. the development of a hybrid DRL system that defeated a human world champion in Go [Silver et al.(2016)]

Following theses successes, DRL algorithms have been applied to a wide range of problems, such as robotics, indoor navigation, computer animation and natural language processing.

I think it is safe to say that DRL will be a major ingredient in future AI systems and at least for me, it is the overall most interesting application of machine learning because of the almost endless possibilities that the combination of RL and DL bring.

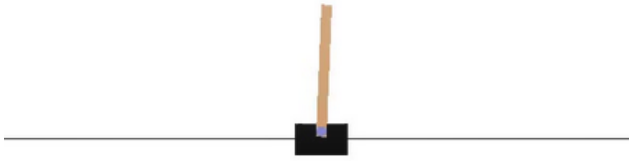

Figure 1: The cartpole-v0 environment from OpenAI.

2 Problem Statement

Thanks to OpenAI, there exists a great environment to test custom RL and DRL algorithms: the OpenAI gym. I propose the following capstone project: I will solve OpenAI's pole-balancing environment with the two different main approaches to solving RL problems, i.e., a method based on value functions and a method based on policy gradients. The problem of pole-balancing is a classical one in RL: according to [Sutton, Barto (1998)] this was first attempted in 1968: *" [T]hey applied it to the task of learning to balance a pole hinged to a movable cart on the basis of a failure signal occurring only when the pole fell or the cart reached the end of the track..."*.

OpenAI's description of the pole–balancing problem is the following: *"A pole is attached by an un-actuated joint to a cart, which moves along a frictionless track. The system is controlled by applying a force of +1 or -1 to the cart. The pendulum starts upright, and the goal is to prevent it from falling over. A reward of +1 is provided for every timestep that the pole remains upright. The episode ends when the pole is more than 15 degrees from vertical, or the cart moves more than 2.4 units from the center." [`https://gym.openai.com/envs/CartPole-v0`]*

3 Datasets and Inputs

The data (which in a reinforcement learning problem corresponds to the environment) comes from OpenAI. In April 2016, OpenAI has released Gym (`https://gym.openai.com/`), a toolkit for developing and comparing reinforcement learning algorithms. With Gym, everyone can write their own RL algorithms and test it in a number of different environments without the need to build the environment oneself. This is an ideal platform for this capstone project. I will use the the pole-balancing environment "CartPole-v0", available at `https://gym.openai.com/envs/CartPole-v0`.

4 Solution Statement

The problem for this capstone project is to write a deep reinforcement learning algorithm that solves the cartpole environment. Solving in this context is defined by OpenAI as: *"CartPole-v0 defines "solving" as getting average reward of 195.0 over 100 consecutive trials."*

5 Benchmark Model

There are many benchmark models available on OpenAI. Each user of gym has the option to write down their solution, so a benchmark model might be the current (2017/11/11) top entry on `https://gym.openai.com/envs/CartPole-v0/`, which is n1try's algorithm, available on `https://gym.openai.com/evaluations/eval_EIcM1ZBnQW2LBaFN6FY65g/`.

6 Evaluation Metrics

The metrics for this project are rather straightforward since they are given by OpenAI. I will choose the same metrics to measure success as they are: the average reward over 100 consecutive trials must exceed 195.

7 Project Design

I plan to first write a non–deep reinforcement learning algorithm, probably Q–learning – the same algorithm that was used for the smartcab problem in the RL Nanodegree section – and to observe how well/bad it does on this problem. Next, I will try out 2 DRL algorithms, one from the family of value function methods and one from the family of policy search methods and compare their performance with the simple Q–learning approach. Which exact algorithm I will use I do not know yet, I will first have to read up on which DRL algorithms are currently used.

If all goes well and one of the DRL methods is able to solve the cartpole problem, I might apply the same algorithm to a different gym environment.

References

[Arulkumaran et al.(2017)] Arulkumaran, K., Deisenroth, M. P., Brundage, M., & Bharath, A. A. 2017, arXiv:1708.05866

[Mnih et al.(2015)] Mnih, V., Kavukcuoglu, K., Silver, D., et al. 2015, Nature, 518, 529

[Silver et al.(2016)] Silver, D., Huang, A., Maddison, C. J., et al. 2016, Nature, 529, 484

[Sutton, Barto (1998)] Sutton, R.S., Barto, A.G, 1998, MIT Press