Review: A Brief Survey of Deep Reinforcement Learning

Norio Kosaka

December 2018

1 Profile of the paper

- Title: A Brief Survey of Deep Reinforcement Learning
- Authors: Kai Arulkumaran, Marc Peter Deisenroth, Miles Brundage, Anil Anthony Bharath
- Published Year: Sep 2017

2 Sections in the paper

1. Introduction
2. Reward-driven Behaviour
    - Markov Decision Processes
    - Challenges in RL
3. Reinforcement Learning algorithms
    - Value Functions
    - Sampling
    - Policy Search
    - Planning and Learning
    - The Rise of DRL
4. Value Functions
    - Function Approximation and the DQN
    - Q-Function Modifications
5. Policy Search
    - Backpropagation through Stochastic Functions

    - Compounding Errors
    - Actor-Critic Methods
6. Current Research and Challenges
    - Model-based RL
    - Exploration vs. Exploitation
    - Hierarchical RL
    - Imitation Learning and Inverse RL
    - Memory and Attention
    - Transfer Learning
    - Benchmarks
7. Conclusions: Beyond Pattern Recognition

3 Abstract

This survey covers central algorithms in Deep Reinforcement Learning(DRL).

4 Introduction

The two principle properties of *Deep Learning*, which are **function approximation** and **representation learning**, help RL to address to the bigger scale problems. Among recent works, they have chosen two breakthroughs;

- Kickstarting the revolution in DRL: Development of DQN that could learn to play a range of Atari 2600 video games at a superhuman level.
- Development of a hybrid DRL: AlphaGo defeated a human world champion in *Go*.

And they have showcased some of real-world applications as follows:

- classic Atari 2600 video games
- TORCS car racing simulator
- Robotics arm control
- wheeled mobile robot control
- Image caption trained by utilising reinforcement learning approach

5 Reward-driven Behaviour

Before jumping into the contributions of deep neural networks to RL, they have introduced the field of RL in general.

- MDP
- Challenges in RL

6 Reinforcement Learning Algorithms

In this section, they have briefly reviewed some aspects of RL following their categorisation of RL as follows:

- **Value Functions**: state-value function, state-action-value function, dynamic programming, SARSA, TD, policy iteration(policy evaluation + policy improvement)
- **Sampling**: importance sampling, Advantage function
- **Policy Search**: gradient-free methods, gradient-based methods, policy gradient, actor-critic methods
- **Planning and Learning**: model-based RL, model-free RL
- **The Rise of DRL**: backpropagation, gradient vanishing problem in long-term

7 Value Functions

One of the earliest success in RL agents is **TD-Gammon**, which combined TD and neural network.

- **Function Approximation and the DQN**: it is based on NFQ(neural fitted Q iteration) and involved two techniques(experience replay and target network)
- **Q-Function Modifications**: Double-Q learning

8 Policy Search

- Backpropagation through Stochastic Functions: *REINFORCE*, hard attention, stochastic value gradients(SVGs)
- Compounding Errors: guided policy search(GPS), trust region using Kullback-Leibler(KL) divergence, TRPO(trust region policy optimisation), GAE(generalised advantage estimation), PPO(proximal policy optimisation)
- Actor-Critic Methods: DPG(deterministic policy gradients), DDPG(deep DPG), A3C, A2C, Gorilla for parallel computation.

9 Current Research And Challenges

- Model-based RL: successor representation(SR)
- Exploration vs. Exploitation: Bootstrapped DQN, UCB(Upper confidence bound)
- Hierarchical RL: top-level policy, high-level options, primitive actions
- Imitation Learning and Inverse RL: behavioural cloning, IRL(is to estimate an unknown reward function from observed trajectories that characterise a desired solution), generative adversarial imitation learning(GAIL)
- MARL
- Memory and Attention: it converts DQN into an RNN, which allows the network to better deal with POMDPs by integrating information over long time periods, deep attention reccurent Q-network(DARQN), Memory Q-network(MQN)
- Transfer Learning
- Benchmarks: standard benchmarks are Cartpole and Mountain Car, Atari2600, VizDoom, Facebook's TorchCraft, Deepmind's StarCraft II and Quake III Arena first-person shooter engine, Microsoft's Project Malmo, OpenAI Gym.