RL seminar #1: Intro to RL & admin stuff

Maksim Kretov

MIPT, Deep learning lab & iPavlov.ai

14 Oct 2017

Outline

Class information
- Logistics
- Course outline
- Assignments
- References

RL introduction

Tools

Assignments

Table of Contents

Class information
- Logistics
- Course outline
- Assignments
- References

RL introduction

Tools

Assignments

Logistics

Course goals (seminars)

To **help** studying RL: provide structure and answer questions.

Organizer

MIPT, NN and DL Lab: `http://edu.ipavlov.ai/`

Staff & Contacts

Maksim Kretov (kretovmk@gmail.com)
Anton Karazeev (anton.karazeev@gmail.com)
Telegram: https://t.me/DeepRLfall2017
Github: https://github.com/deepmipt/deep-rl-seminars
Real world: BFK, classroom 518.

Default place and time

MIPT, BFK, every Saturday at 12pm starting 14 Oct 2017. In total 7-8 seminars.

Course outline

Put topics for discussion there: `https://goo.gl/qN6jmJ`
Write questions on the 2nd w/sheet of the same doc.

Loose plan

1. Intro to RL $\leftarrow$ TODAY.
2. MDP, Bellman equations.
3. Value-based methods.
4. Policy gradient I-II.
5. Stochastic computational graphs I-II.

Typical seminar

I tell something relevant to current lectures (NOT reproduction of other courses' slides); we discuss questions and assignments.

Assignments

Scoring system for "zachet"

50%: 3 home assignments (coding, 10p) and 5 quizzes (google doc form, 4p). Works with high similarity score are discarded.
50%: exam at the end OR top-10 in RL competition (50p)

Quizzes (min 1 week)

One quiz after seminar.

Coding assignments (min 2 weeks)

HW1: PyTorch and imitation learning.
HW2-HW3: value methods and policy gradient.

References

Basic courses

- RL boot camp: https://sites.google.com/view/deep-rl-bootcamp/
- CS294: http://rll.berkeley.edu/deeprlcourse/

Introductory courses

- CS188: http://ai.berkeley.edu/home.html
- David Silver's course: http://www0.cs.ucl.ac.uk/staff/d.silver/web/Home.html

Books

- Sutton's book: https://goo.gl/spWWVL
- RL algorithms: https://goo.gl/V4ppwA

Table of Contents

Class information
- Logistics
- Course outline
- Assignments
- References

RL introduction

Tools

Assignments

RL introduction

Task definition

Agent receive from environment state S_{t+1}, reward $r_t(s_t, a_t)$ and perform action a_{t+1}, which is determined by parametrized policy $\pi_\theta(a_t|s_t)$. Goal is to maximize cumulative reward $J(\theta) = \sum_{t=0}^{T-1} r_t$.

Cool stories about RL

Backgammon, Chess, Go, Atari, Robotics

RL introduction

Why RL is powerful?

- Universal – most of the tasks can be framed in RL setting
- Reward is sparse and non-differentiable

Cooler stories about RL

RL2[1], Neural architectures search[2], Reasoning with neural networks[3]

[1] https://arxiv.org/abs/1611.02779

[2] https://arxiv.org/pdf/1707.04873.pdf

[3] http://bair.berkeley.edu/blog/2017/06/20/learning-to-reason-with-neural-module-networks/

RL introduction

MDP and PODMP

Directed and undirected graphical models.
Observations and states: MDP and POMDP.

Imitation learning

Questions?

Table of Contents

Class information
- Logistics
- Course outline
- Assignments
- References

RL introduction

Tools

Assignments

Tools

PyTorch

- Dynamic computational graphs[4]
- Dynamic batching[5]
- Fast and clear

Implementation of REINFORCE or A2C takes about 100 lines of code, including imports and args parsing (see examples in repo[6]). Many more complicated algorithms already implemented[7].

[4] http://pytorch.org/tutorials/

[5] https://medium.com/@ilblackdragon/pytorch-dynamic-batching-f4df3dbe09ef

[6] https://github.com/pytorch/examples/tree/master/reinforcement˙learning

[7] https://github.com/ikostrikov

Table of Contents

Class information
- Logistics
- Course outline
- Assignments
- References

RL introduction

Tools

Assignments

Assignments

Reading

Lectures 2-4 of CS294.
Please, let us know about your questions: google doc, e-mail, telegram.

Doing

Quiz and home assignment are coming.