RLVS-ANITI

Event Schedule

Thu, Mar 25, 2021

9:00am

Opening remarks

9:00am - 9:10am, Mar 25

Recordings Video
Materials Slides

Speaker

Sébastien Gerchinovitz IRT Saint-Exupery and Institut de Mathématiques de Toulouse

9:10am

RLVS overview

9:10am - 9:30am, Mar 25

Recordings Video
Materials Slides

Abstract This short introduction will present the organization of RLVS 2021 and the progression of ideas across days and sessions.

Speaker

Emmanuel Rachelson ISAE-SUPAERO

9:30am

RL fundamentals

9:30am - 1:00pm, Mar 25

Recordings Video
Materials Slides

Abstract The overall goal of this session is to provide the RLVS participants with an overview of RL and an understanding of (some) key challenges. It should also introduce participants to issues that will be tackled in more detail in the following classes. We will follow the course of a notebook together and we will alternate between a "lecture" mode and small illustrative exercises (with corrections) that the participants will have time to try themselves. We will introduce the modeling fundamentals of Makov Decision Processes. Then we will move our way up from toy examples and fundamental algorithms to key challenges in RL. In particular, we will focus on three key structuring issues for the RL practitioner: the exploration/exploitation tradeoff, value function approximation, and optimality search. Along with the session, we will link each topic to the corresponding classes in RLVS.

Speaker

Emmanuel Rachelson ISAE-SUPAERO

2:00pm

Introduction to Deep Learning

2:00pm - 4:00pm, Mar 25

Recordings Video
Materials Slides

Abstract Deep Learning, a form of machine learning inspired by biological learning, has powered innovations in many fields, including Reinforcement Learning. While artificial neural networks, the core component of deep learning, have been used in artificial intelligence for decades, recent advances in the domain of deep learning such as GPU computation, convolutional layers, and data availability have led to breakthroughs in machine learning, computer vision, and many applications of deep learning such as protein folding. Deep Reinforcement Learning uses deep neural networks as a central component for many algorithms, such as Deep Q Networks and Soft Actor Critic; it is therefore important to understand deep neural networks for Reinforcement Learning. In this session, we will introduce the theoretical foundations of Deep Learning, namely backpropagation, stochastic gradient descent, and layer operations such as convolution. We will apply these concepts in exercises using the PyTorch library on supervised learning examples.

Speaker

Dennis G Wilson ISAE-Supaero and University of Toulouse

4:30pm

Reward Processing Biases in Humans and RL Agents

4:30pm - 5:30pm, Mar 25

Recordings Video
Materials Slides

Abstract Drawing inspirations from studies of human behavior, we propose a general and flexible parametric framework for sequential decision-making based on a two-stream mechanism for processing positive and negative rewards. Our framework extends standard problem settings, such as multi-armed bandits (MAB), contextual bandits (CB) and general reinforcement learning (RL), allowing to incorporate a wide range of reward-processing biases -- an important component of human decision making which can help us better understand a wide spectrum of multi-agent interactions in complex real-world socioeconomic systems, as well as various neuropsychiatric conditions associated with disruptions in normal reward processing. The reward processing biases are modeled by a combination of weights on incoming two-stream rewards, as well as on memories about the prior reward history, resulting into more flexible parametric approaches that can outperform standard algorithms for sequential decision making, such as, for example, Q-Learning and SARSA methods, as well as recently proposed Double Q-Learning, on a variety of simulated and realistic tasks.

Speaker

Irina Rish Université de Montréal

5:45pm

Introduction to Hierarchical Reinforcement Learning

5:45pm - 6:45pm, Mar 25

Recordings Video

Abstract Hierarchical reinforcement learning refers to a class of computational methods that enable artificial agents that train using reinforcement learning to act, learn and plan at different levels of temporal abstraction. In this talk, I will review the main ideas of these computational approaches and present some recent advances in this field. In addition to computational results, I will draw some connections between the algorithms' hierarchical reinforcement learning approaches and existing similar models of human and animal decision making.

Speaker

Doina Precup McGill University and DeepMind

Fri, Mar 26, 2021

10:00am

Stochastic Bandits

10:00am - 12:00pm, Mar 26

Recordings Video
Materials Slides, Code, Notebook on colab (implementing the same code), The Bandit Algorithms book

Abstract The bandit framework specializes the reinforcement learning setup by removing the (controlled) state. Bandits provide all the essential ingredients to study the exploration/exploitation dilemma, with many principles derived for bandits generalizing to the reinforcement learning setting. The simplification has the advantage that it permits a more complete understanding and practical algorithms. Furthermore, bandits are a good model for many applications.

I will introduce bandit problems and present the most well-known algorithms based on the principle of optimism in the face of uncertainty. There will be a live coding demo and discussions of the many extensions needed in practical applications.

Speaker

Tor Lattimore DeepMind

2:00pm

Monte Carlo Tree Search

2:00pm - 4:00pm, Mar 26

Recordings Video
Materials Slides (starts at slide 41), Code, Notebook on colab (implementing the same code, with solutions to the exercises), The Bandit Algorithms book

Abstract Monte Carlo Tree Search is an algorithm for game tree search most famous for its application in AlphaGo. I will give a tutorial on this algorithm, which will include a significant practical component in Python. A very basic understanding of Python and Numpy will be useful for students wishing to complete the practical component independently.

Speaker

Tor Lattimore DeepMind

4:30pm

Multi-Armed Bandits in Clinical Trials

4:30pm - 5:30pm, Mar 26

Recordings Video
Materials Slides

Abstract Bayesian bandit problems have been described in the theoretical statistics literature since 1933. I'll say how I got into the area as a graduate student and some of my contributions to the bandit theory. My motivation was clinical trials, even though I came to learn that this application was summarily rejected by all clinical trialists ... at the time. I'll give you a few-minute tour through several decades of my life where I worked within the conventional clinical trial establishment trying to change the paradigm. In the last decade, I have had more than a modicum of success. I'll describe actual national and global clinical trials that I've design and that are being conducted today. I call them Bayesian adaptive platform trials, which is a euphemism for multi-armed bandits. They are getting a surprising amount of support from regulators. For example, the Director of the Center for Drugs at the FDA says, "These have to be the future."

Speaker

Donald Berry University of Texas and Rice University

Sat, Mar 27, 2021

Sun, Mar 28, 2021

Mon, Mar 29, 2021

Tue, Mar 30, 2021

Wed, Mar 31, 2021

Thu, Apr 01, 2021

9:00am

Deep Q-Networks and its variants

9:00am - 12:30pm, Apr 1

Abstract:
We will present in a coherent paradigm the different breakthroughs that led to the seminal paper Deep Q-Networks (DQN). Starting from dynamic programming and the value iteration algorithm, we will show how DQN can be seen as a particular instance of an approximate value iteration algorithm. Then, we will present an open-source codebase released by DeepMind called dqn_zoo that implements in JAX the DQN algorithm. Finally, we will present the variants of DQN, from Double DQN (DDQN) to Implicite Quantile DQN (IQN) and more.

Prerequisites:

- Understanding of Markov Decisions Processes and Bellman equations (starting point of the presentation)
- Basic notions of Python and JAX.
- (Optional) Run the dqn_zoo codebase.

Speakers

Bilal Piot DeepMind

Corentin Tallec DeepMind

12:30pm

Lunch Break

12:30pm - 1:30pm, Apr 1

1:30pm

Coffee Break (Social)

1:30pm - 2:00pm, Apr 1

2:00pm

Deep Q-Networks and its variants

2:00pm - 3:00pm, Apr 1

Abstract:
We will present in a coherent paradigm the different breakthroughs that led to the seminal paper Deep Q-Networks (DQN). Starting from dynamic programming and the value iteration algorithm, we will show how DQN can be seen as a particular instance of an approximate value iteration algorithm. Then, we will present an open-source codebase released by DeepMind called dqn_zoo that implements in JAX the DQN algorithm. Finally, we will present the variants of DQN, from Double DQN (DDQN) to Implicite Quantile DQN (IQN) and more.

Prerequisites:

- Understanding of Markov Decisions Processes and Bellman equations (starting point of the presentation)
- Basic notions of Python and JAX.
- (Optional) Run the dqn_zoo codebase.

Speakers

Bilal Piot DeepMind

Corentin Tallec DeepMind

3:00pm

Coffee Break (social)
3:00pm - 3:15pm, Apr 1

3:15pm

Regularized MDPs
3:15pm - 4:15pm, Apr 1

Abstract:
Many recent efficient deep reinforcement learning algorithms make use of some sort of regularization. This tutorial will review these approaches through the lens of regularized approximate dynamic programming, which allows connecting the different algorithms and explains theoretically why regularization works.

Speaker

Matthieu Geist Google Brain

4:30pm

TBA
4:30pm - 5:30pm, Apr 1

Speaker

Mengdi Wang Princeton University

5:30pm

Tea Break (Social)
5:30pm - 6:00pm, Apr 1

Fri, Apr 02, 2021

9:00am

Policy Gradients and Actor Critic methods
9:00am - 12:30pm, Apr 2

Materials Slides 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14

Abstract:

Starting from the general policy search problem and direct policy search methods, I will give a didactical presentation of the Policy Gradient Theorem and explain some variants of the REINFORCE algorithm. From there, I will move step by step to presenting more advanced methods such as TRPO, PPO and Actor-Critic methods such as DDPG and SAC.

Outline:

1. The policy search problem
2. Policy search methods: direct policy search versus policy gradient
3. Policy gradient derivation
4. Understanding the policy gradient
5. From policy gradient with baseline to actor-critic
6. Bias-variance trade-off
7. On-policy vs off-policy
8. TRPO, ACKTR and PPO
9. DDPG and TD3
10. SAC
11. Wrap-up

Speaker

Olivier Sigaud Sorbonne Université

12:30pm

Lunch Break

12:30pm - 1:30pm, Apr 2

1:30pm

Coffee Break (Social)

1:30pm - 2:00pm, Apr 2

2:00pm

Pitfalls in Policy Gradient methods

2:00pm - 3:00pm, Apr 2

Materials Slides 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14

Abstract In this talk, I will present the behavior of variants of the REINFORCE algorithm using simple gym classic control benchmarks (CartPole, Pendulum, MountainCar...) and various stochastic policy representations (Bernoulli, Gaussian, squashed Gaussian). I will highlight difficulties faced by these algorithms on those simple environments and draw lessons about the necessity to better understanding how they work or why they don't before moving to more advanced methods and more complex benchmarks.

Speaker

Olivier Sigaud Sorbonne Université

3:30pm

Exploration in Deep RL

3:30pm - 5:30pm, Apr 2

Abstract:
One of the major challenges in reinforcement learning (RL) is the trade-off between exploration of the environment to gather information and exploitation of the samples observed so far to execute "good" (nearly optimal) actions. In this seminar, we review how exploration techniques are paired with function approximation in continuous state-action spaces. In particular, we will focus on the integration of exploration mechanisms with deep learning techniques. The seminar should provide enough theoretical and algorithmic background to understand existing techniques and possibly devise novel methods. Throughout the talk, we will discuss open problems and possible future research directions.

Speaker

Matteo Pirotta Facebook AI Research

5:30pm

Tea Break (Social)

5:30pm - 6:00pm, Apr 2

Sat, Apr 03, 2021

Sun, Apr 04, 2021

Mon, Apr 05, 2021

Tue, Apr 06, 2021

Wed, Apr 07, 2021

Thu, Apr 08, 2021

9:00am

Evolutionary Reinforcement Learning

9:00am - 11:00am, Apr 8

Abstract:
Reinforcement learning traditionally takes inspiration from operant conditioning, that is, trial-and-error learning during the lifetime of the agent. However, evolution is another trial-and-error process that is very successful in nature. This process inspired many algorithms that can also solve reinforcement learning problems while using a very different set of metaphors; in that case, learning happens at the phylogenetic timescale, from generation to generation. While evolutionary learning has its strengths, it also raises its own challenges. In this class, we will focus on the representation problem "How can an artificial genotype describe a neural network that could be as complex as a brain?"—and the stepping-stone problem—"what intermediate steps lead to an artifact as sophisticated as a brain?". We will also draw parallels with traditional reinforcement learning methods and attempt to understand the strengths and weaknesses of each family of methods.

Readings:
[1] https://www.nature.com/articles/s42256-018-0006-z
[2] https://www.cell.com/iscience/fulltext/S2589-0042(20)30928-7

Speakers

Dennis G Wilson ISAE-Supaero and University of Toulouse

Jean-Baptiste Mouret Inria

11:00am

Coffee Break (social)

11:00am - 11:30am, Apr 8

11:30am

Evolving Agents that Learn More Like Animals

11:30am - 12:30pm, Apr 8

Abstract:
Deep neuroevolution, a combination of deep neural networks and evolutionary algorithms, has recently shown to be a competitive alternative to other deep reinforcement learning approaches. In this talk, I will present some of our work on creating agents that evolve to learn complex tasks, such as playing games or controlling robots. I will show that these algorithms do not only allow agents to perform in simple environments but also enable them to (1) learn 3D tasks directly from pixels, (2) learn models of the world for rapid planning, and (3) adapt quickly to task changes through a biologically-inspired form of meta-learning.

Speaker

Sebastian Risi IT University of Copenhagen

12:30pm

Lunch Break

12:30pm - 1:30pm, Apr 8

1:30pm

Coffee Break (Social)

1:30pm - 2:00pm, Apr 8

2:00pm

Micro-data Policy Search

2:00pm - 4:00pm, Apr 8

Abstract:
Most policy search algorithms require thousands of training episodes to find an effective policy, which is often infeasible when experiments takes time or are expensive (for instance, with physical robot or with an aerodynamics simulator). This class focuses on the extreme other end of the spectrum: how can an algorithm adapt a policy with only a handful of trials (a dozen) and a few minutes? By analogy with the word "big-data", we refer to this challenge as "micro-data reinforcement learning". We will describe two main strategies: (1) leverage prior knowledge on the policy structure (e.g., dynamic movement primitives), on the policy parameters (e.g., demonstrations), or on the dynamics (e.g., simulators), and (2) create data-driven

surrogate models of the expected reward (e.g., Bayesian optimization) or the dynamical model (e.g., model-based policy search), so that the policy optimizer queries the model instead of the real system. Most of the examples will be about robotic systems, but the principle apply to any other expensive setup.

Readings:
[1] https://arxiv.org/abs/1807.02303

Speakers

Konstantinos Chatzilygeroudis University of Patras

Jean-Baptiste Mouret Inria

4:00pm

Coffee Break (social)
4:00pm - 4:30pm, Apr 8

4:30pm

Efficient Motor Skills Learning in Robotics
4:30pm - 5:30pm, Apr 8

Abstract:
As a fundamental cornerstone in the development of intelligent robotic systems, the research community on robot learning has addressed autonomous motor skill learning and control in complex task scenarios. Imitation learning provides an efficient way to learn new skills through human guidance, which can reduce time and cost to program the robot. Robot learning architectures can provide a comprehensive framework for learning, recognition and reproduction of whole body motions. The inference mechanism can be applied not only to learn the robot's free body motion but also to learn physical interaction tasks, including human robot interaction. In this talk, I will introduce robot learning algorithms including learning from human demonstrations, incremental learning and the extension of learning from simple movement primitives to complex tasks, such as context aware manipulation task, and human robot collaboration.

Speaker

Dongheui Lee Technical University of Munich

5:30pm

Tea Break (Social)
5:30pm - 6:00pm, Apr 8

Fri, Apr 09, 2021

9:00am

RL in Practice: Tips and Tricks and Practical Session With Stable-Baselines3
9:00am - 1:00pm, Apr 9

Abstract:
The aim of the session is to help you do reinforcement learning experiments. The first part covers general advice about RL, tips and tricks and details three examples where RL was applied on real robots. The second part will be a practical session using the Stable-Baselines3 library.

Pre-requisites:
Python programming, RL basics, (recommended: Google account for the practical session in order to use Google Colab).

Additional material:
Website: https://github.com/DLR-RM/stable-baselines3
Doc: https://stable-baselines3.readthedocs.io/en/master/

Outline:
Part I: RL Tips and Tricks / The Challenges of Applying RL to Real Robots

1. Introduction (3 minutes)
2. RL Tips and tricks (45 minutes)
    1. General Nuts and Bolts of RL experimentation (10 minutes)
    2. RL in practice on a custom task (custom environment) (30 minutes)
    3. Questions? (5 minutes)
3. The Challenges of Applying RL to Real Robots (45 minutes)
    1. Learning to control an elastic robot - DLR David Neck Example (15 minutes)
    2. Learning to drive in minutes and learning to race in hours - Virtual and real racing car (15 minutes)
    3. Learning to walk with an elastic quadruped robot - DLR bert example (10 minutes)
    4. Questions? (5 minutes+)

Part II: Practical Session with Stable-Baselines3

1. Stable-Baselines3 Overview (20 minutes)
2. Questions? (5 minutes)
3. Practical Session - Code along (1h+)

Speaker

Antonin Raffin DLR

1:00pm

Lunch Break
1:00pm - 2:00pm, Apr 9

2:00pm

Coffe Break (Social)
2:00pm - 2:30pm, Apr 9

2:30pm

Symbolic representations and reinforcement learning
2:30pm - 3:30pm, Apr 9

Abstract:
A remarkable property of deep learning algorithms is their ability to learn useful task-specific representations from data directly without the need for hand-crafted feature engineering. As they have grown in popularity over the past decade deep neural networks (NNs) have been successfully applied to a wide range of machine learning tasks, achieving state of the art results across many research areas. However, as the complexity of the research problems increase some of the limitations of NN become increasingly clear: NNs suffer from interpretability issues, poor generalisation that leads to very data-

hungry algorithms and the inability to be combined with other old, well established AI algorithms. Some of the research tackling these drawbacks takes inspiration from symbolic AI. It focusses, for example, on obtaining interpretable representations from NNs or thinking about objects and relations when building network architectures. This talk reviews symbolic approaches and properties that might be interesting to keep in the back of our heads for current representation learning and reviews current research that merges deep and symbolic methods with an emphasis on methods applied to reinforcement learning.

Speaker

Marta Garnelo DeepMind

3:30pm

Coffee Break (social)

3:30pm - 3:45pm, Apr 9

3:45pm

Leveraging model-learning for extreme generalization

3:45pm - 4:45pm, Apr 9

Early RL was almost completely focused on learning a policy, sometimes via a value function, for a fixed reward function. If the reward function is really fixed and if the agent really needs a quick reaction in all world states, that's the right strategy. But those conditions don't always hold. "Goal-conditioned" RL attempts to relax the first condition by allowing the reward function to be, in essence, part of the state. That addresses the first problem but makes the second one much worse! Now we are have to find a quick reaction to all state-goal pairs. I will extol the virtues of learning models instead, with a focus on models that generalize extremely aggressively via factoring and quantification, and argue informally that such models can be learned with relatively less computation and data, and can be acquired incrementally without interference. "Compilation" methods can still be used to make reactions fast for very common cases while retaining the ability to deliberate to solve unusual or difficult problems. Depending on time and interest, I might also talk about the substantial partial observability that arises in long-term household robot problems, discuss some of our approaches to it, and encourage participants to help think of others.

Class material:
Blog post

Speaker

Leslie Kaelbling MIT

4:45pm

Coffee Break (social)

4:45pm - 5:00pm, Apr 9

5:00pm

RLVS Wrap-Up

5:00pm - 6:00pm, Apr 9

Speaker

Emmanuel Rachelson ISAE-SUPAERO