



A Quick Look at the “Reinforcement Learning” course

A. LAZARIC (*SequeL Team @INRIA-Lille*)
ENS Cachan - Master 2 MVA

SequeL – INRIA Lille

Why

Why: Important Problems

Why: Important Problems

- ▶ Autonomous robotics



- ▶ *Elder care*
- ▶ *Exploration of unknown/dangerous environments*
- ▶ *Robotics for entertainment*

Why: Important Problems

- ▶ Autonomous robotics
- ▶ Financial applications



- ▶ *Trading execution algorithms*
- ▶ *Portfolio management*
- ▶ *Option pricing*

Why: Important Problems

- ▶ Autonomous robotics
- ▶ Financial applications
- ▶ Energy management



- ▶ *Energy grid integration*
- ▶ *Maintenance scheduling*
- ▶ *Energy market regulation*
- ▶ *Energy production management*

Why: Important Problems

- ▶ Autonomous robotics
- ▶ Financial applications
- ▶ Energy management
- ▶ Recommender systems

The screenshot shows a Google search for "bird houses". The search bar contains "bird houses" and the search button is visible. Below the search bar, the results are displayed. The first two results are "Sponsored Links":

- Bird Houses** - Scotts attracts colorful birds to your backyard! (www.Scotts.com)
- Specialty Bird Houses** - www.birds-out-back.com Roosting Boxes, Purple Martins, Bat Chalets.

Below the sponsored links, there are several organic search results:

- Bird Houses at BestNest** - www.bestnest.com Over 225 different houses in stock. Free shipping!
- Learn More About Bird Houses** - JustBirdHouses.net learn more about blue bird houses and birdhouses, along with our blue bird house and purple martin bird houses. www.justbirdhouses.net/ - Cached - Similar - ⌵
- Bird Houses** - Bird Houses and Bird Feeders for north american bird species. www.birdhouses101.com/ - Cached - Similar - ⌵
- Bird Feeders, Bird Houses - The Backyard Bird Company** - Bird Feeders - The Backyard Bird Company has a variety of bird feeders will accent your landscape and attract wildlife. Bird Houses - Birdfeeders - Decorative Bird Houses www.backyardbird.com/ - Cached - Similar - ⌵

On the right side of the search results, there are more sponsored links:

- Bird Houses** - Find All Types Of Bird Feeders And Houses At Lowe's® New Lower Price! www.Lowe.com
- Bird Houses Sale** - Authorized Dealer - New Designs. Low Price Guarantee- Free Shipping www.OutdoorLivingShowroom.com
- High Quality Bird Houses** - Nesting boxes & decorative houses. 5-Star Service. Free Shipping \$75+ www.backyardbird.com
- Decorative Bird Houses** - Beautify Your Garden With Our Wooden Bird Houses at a Discount! Birdhouse Station.com

Two red arrows point to the "Sponsored Links" labels in the top right of the search results area.

- ▶ *Web advertising*
- ▶ *Product recommendation*
- ▶ *Date matching*

Why: Important Problems

- ▶ Autonomous robotics
- ▶ Financial applications
- ▶ Energy management
- ▶ Recommender systems
- ▶ Social applications



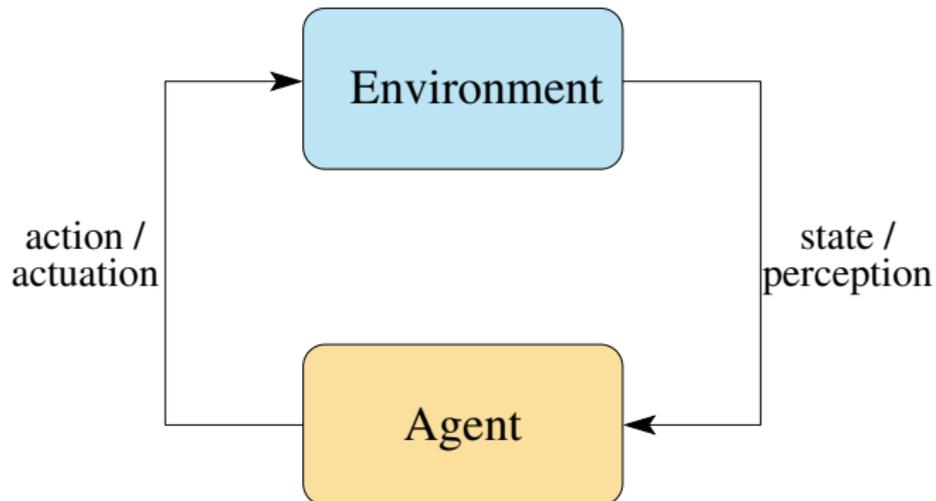
- ▶ *Bike sharing optimization*
- ▶ *Election campaign*
- ▶ *ER service optimization*
- ▶ *Resource distribution optimization*

Why: Important Problems

- ▶ Autonomous robotics
- ▶ Financial applications
- ▶ Energy management
- ▶ Recommender systems
- ▶ Social applications
- ▶ And many more...

What

What: Decision-Making under Uncertainty



How: Reinforcement Learning

Reinforcement learning is learning what to do – how to map situations to actions – so as to **maximize** a numerical reward signal. The learner is not told which actions to take, as in most forms of machine learning, but instead must discover which actions yield the most reward by **trying them (trial-and-error)**. In the most interesting and challenging cases, actions may affect not only the immediate reward but also the **next situation** and, through that, all **subsequent rewards (delayed reward)**.

*“An introduction to reinforcement learning”,
Sutton and Barto (1998).*

How: the Course



Google bird houses

Web: [Show options...](#) Results 1 - 16 of about 22,108,000 for **bird houses** (0.15)

Bird Houses
[www.Scotts.com](#) Scotts attracts colorful birds to your backyard!

Specialty Bird Houses
[www.birdandback.com](#) Spawning Boxes, Purple Martins, Blue Chaires

Bird Houses at Backyard
[www.backyard.com](#) Over 225 different houses in stock. Free shipping!

Learn More About Bird Houses

Bird House [Jardins@houses.net](#) learn more about blue bird houses and birdhouses, along with our blue bird house and purple martin bird houses.

[www.jardins@houses.net](#) - Cached - Similar - ⌵

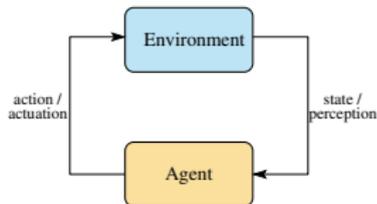
Bird Houses [www.birdhouses101.com](#) - Cached - Similar - ⌵

Bird Feeders, Bird Houses - The Backyard Bird Company
Bird Feeders - The Backyard Bird Company has a variety of bird feeders will accent your landscape and attract wildlife.

Bird Houses - Birdhouses - Decorative Bird Houses
[www.backyardbird.com](#) - Cached - Similar - ⌵

Bird Houses [www.birdhouses101.com](#) - Cached - Similar - ⌵

Decorative Bird Houses
Quality Free Garden With Our Wooden Bird Houses at a Discount [Bird-Houses-Station.com](#) - Cached - Similar - ⌵



Formal and *rigorous* approach to the RL's way to decision-making under uncertainty

What: the Highlights of the Course

How do we formalize the agent-environment interaction?

Markov Decision Process and Policy

A Markov decision process (MDP) is represented by the tuple $M = \langle X, A, r, p \rangle$ where X is the state space, A is the action space, $r : X \times A \rightarrow [0, B]$ is the reward function, p is the dynamics.

At time $t \in \mathbb{N}$ a *decision rule* $\pi_t : X \rightarrow A$ is a mapping from states to actions and a *policy* (strategy, plan) is a sequence of decision rules $\pi = (\pi_0, \pi_1, \pi_2, \dots)$.

The Bellman equations

$$V^\pi(x) = r(x, \pi(x)) + \gamma \sum_y p(y|x, \pi(x)) V^\pi(y),$$

$$V^*(x) = \max_{a \in A} \left[r(x, a) + \gamma \sum_y p(y|x, a) V^*(y) \right].$$

What: the Highlights of the Course

How do we solve an MDP?

Dynamic Programming

Value Iteration

$$V_{k+1} = \mathcal{T}V_k$$

Policy Iteration

- ▶ *Evaluate*: given π_k compute V^{π_k} .
- ▶ *Improve*: given V^{π_k} compute $\pi_{k+1} = \text{greedy}(V^{\pi_k})$

What: the Highlights of the Course

How do we solve an MDP “online”?

Q-learning

Given a observed transition x, a, x', r update

$$Q_{k+1}(x, a) = (1 - \alpha)Q_k(x, a) + \alpha(r + \max_{a'} Q_k(x', a')).$$

What: the Highlights of the Course

How do we effectively trade-off exploration and exploitation?

Multi-arm Bandit

Given K arms we define the regret over n rounds of a bandit strategy as

$$R_n = \sum_{t=1}^n X_{i^*,t} - \sum_{t=1}^n X_{I_t,t}.$$

For the UCB strategy we can prove

$$R_n \leq \sum_{i \neq i^*} \frac{b^2}{\Delta_i} \log(n).$$

What: the Highlights of the Course

How do we solve a “huge” MDP?

Approximate Dynamic Programming

Approximate Value Iteration

$$\hat{V}_{k+1} = \hat{\mathcal{T}}\hat{V}_k$$

Approximate Policy Iteration

- ▶ *Evaluate*: given π_k compute \hat{V}^{π_k} .
- ▶ *Improve*: given \hat{V}^{π_k} compute $\hat{\pi}_{k+1} \approx \text{greedy}(\hat{V}^{\pi_k})$

What: the Highlights of the Course

How “sample-efficient” are these algorithms?

Sample Complexity of LSP1

$$\|V^{\pi_K} - V^*\|_{2,\rho} \leq \inf_{f \in \mathcal{F}} \|V^* - f\|_{2,\rho} + \frac{C_\rho}{1-\gamma} \sqrt{\frac{\log(1/\delta)}{n}}.$$

See you on Tue at 11h in C103!



Who

Lectures

Alessandro LAZARIC

Sequel Team
INRIA-Lille Nord Europe

alessandro.lazaric@inria.fr
researchers.lille.inria.fr/~lazaric/

Practical Sessions

Emilie KAUFMANN

Telecom ParisTech

emilie.kaufmann@telecom-paristech.fr
perso.telecom-paristech.fr/~kaufmann/

When/What/Where

<i>Date</i>	<i>Topic</i>	<i>Classroom</i>
01/10	Intro/MDP	C103
08/10	Dynamic Programming	C103
15/10	RL Algorithms	C103
22/10	<i>TP</i> on DP and RL	C109
29/10	Multi-arm Bandit (1)	C103
05/11	<i>TP</i> on Bandit	C109
12/11	Multi-arm Bandit (2)	C103
19/11	<i>TP</i> on Bandit	C109
26/11	Approximate DP	C103
03/12	Sample Complexity of ADP	C103
10/12	<i>TP</i> on ADP	C109
17/12	Guest lectures + Internships	C103 (TBC)
14/01	Evaluation	C103 (TBC)

Lectures are from 11am to 1pm, TP should be from 11am to 1:15pm.

Evaluation

- ▶ Papers review + oral presentation
- ▶ Projects
- ▶ Stages
- ▶ PhD

Reinforcement Learning

The Inria logo is displayed in a white rounded square with a teal border. The word "Inria" is written in a red, cursive script font.

Alessandro Lazaric

`alessandro.lazaric@inria.fr`

`sequel.lille.inria.fr`