UNIVERSITÉ DE TECHNOLOGIE DE BELFORT-MONTBÉLIARD

Robot Learning

RO51 - Introduction to Mobile Robotics

Zhi Yan

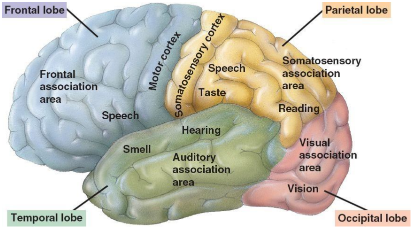
April 24, 2024

https://yzrobot.github.io/

www.utbm.fr

What is learning?

- For humans, **learning** is a process of transforming the cerebral cortex.

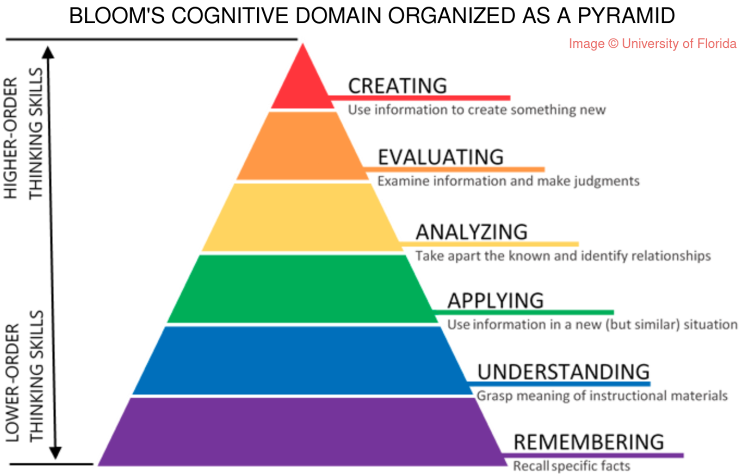


- Learning is a **process** (of transformation), not a collection (of knowledge).
- Types of learning:
  - Learning knowledge
  - Learning how to learn (meta-learning)

Human learning

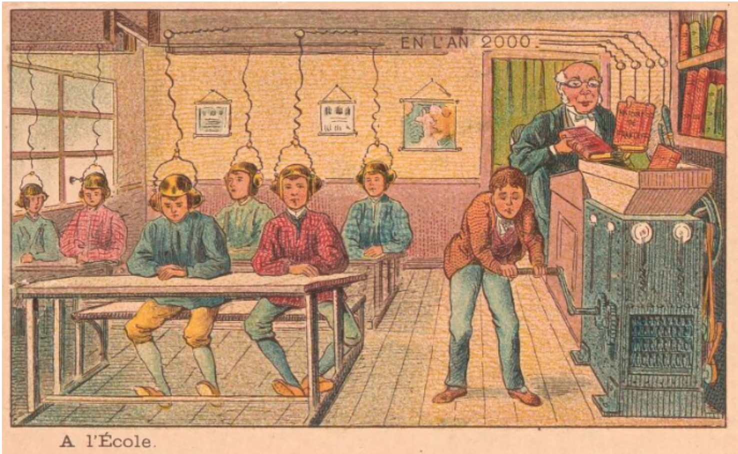
- **Why?** Learning is fundamental to our existence, and a constant process that fuels our growth and shapes who we are.
- **When?** Humans are capable of learning throughout their entire lives!
- **How?**

Human learning



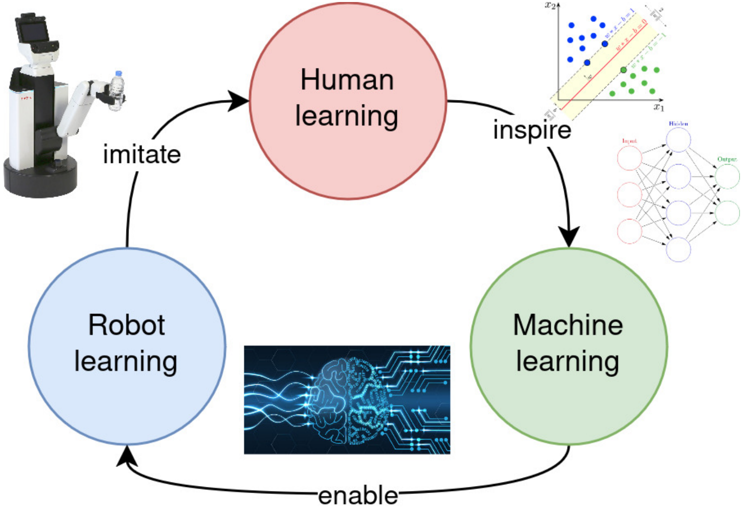
Bloom's taxonomy, Benjamin Bloom, 1956

Human learning



En l'an 2000, Jean Marc Cote, 1901

From human to robot



Machine learning

Recommended reading:

Robot learning

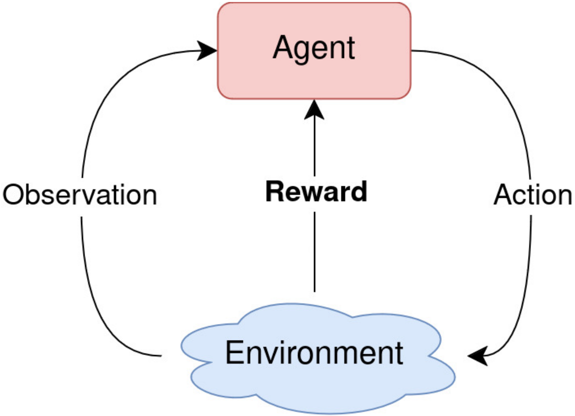
This lecture focuses on **Robot Learning**, especially those relevant methods:

- Reinforcement learning
- Imitation learning
- Lifelong learning

derived from one of the unique characteristics that distinguishes it from **Machine Learning**:

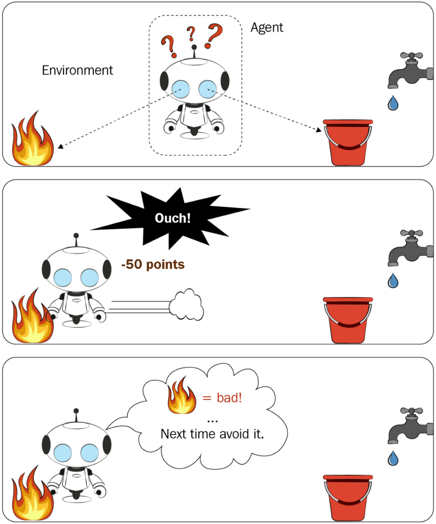
The robot can actively interact with the environment physically.

Reinforcement learning



- The goal of reinforcement learning is to produce good **policies**.
- Typical policy structure: **IF-THEN-ELSE**

Reinforcement learning



1. Observe

2. Select action using policy

3. Action

4. Get reward or penalty

5. Update policy (learning step)

6. Iterate until an optimal policy is found

Reinforcement learning

A specific example: *Towards Learning Robot Table Tennis*

Reinforcement learning

Advantages

- Reinforcement learning allows robots to learn on their own with little to no human intervention during the learning process.
- Reinforcement learning can cope with complex robot control problems, such as robot navigation, object manipulation, etc.

Challenges

- While reinforcement learning is well suited to robotics, deploying it is not easy: **sometimes the cost of trial-and-error can be very high!**
- Understanding how reinforcement learning agents make decisions can be difficult, especially with complex models: **debugging becomes more difficult, leading to reliability and safety issues in method deployment.**

Reinforcement learning

Mathematical model:

- At each time point, the agent is in a state $s \in S$ of the environment.
- After the agent takes an action $a \in A$, it will receive a reward $r(s, a)$, and according to an environment state transition function $p(s' \mid s, a)$ transitions to the next state s'.
- The goal of the agent is to learn a set of policies $\pi(a \mid s)$ (i.e., a set of correspondences from the current state to the actions to be taken, and essentially $a = \pi(s)$) that **maximize** the total reward it receives.

$\Longrightarrow$ **Markov decision process (MDP)**

Reinforcement learning

Algorithm Q-learning

Initialize $Q(s, a)$ for all states s and actions a (e.g., to 0)
Set learning rate α $(0 < \alpha <= 1)$
Set discount factor γ $(0 < \gamma <= 1)$
repeat(for each episode):
    Initialize state s
    repeat(for each step in episode):
        Choose action a from s using an ε-greedy policy
        Take action a, observe reward r and next state s'
        Update $Q(s, a) \leftarrow Q(s, a) + \alpha[r + \gamma \max_{a'} Q(s', a') - Q(s, a)]$
        $s \leftarrow s'$
    until s is terminal
until termination condition is met

Reinforcement learning

Q-learning:

- One of the most representative reinforcement learning algorithms.
- Q stands for quality, corresponding to obtaining an estimate of $r(s, a)$.
- According to line-1 of the algorithm, we can imagine that a table is generated and will be maintained, namely **Q-table**.

Q-table	**a1**	**a2**
s1	Q(s1,a1)	Q(s1,a2)
s2	Q(s2,a1)	Q(s2,a2)

Table: An example of Q-table

Reinforcement learning

Q-learning:

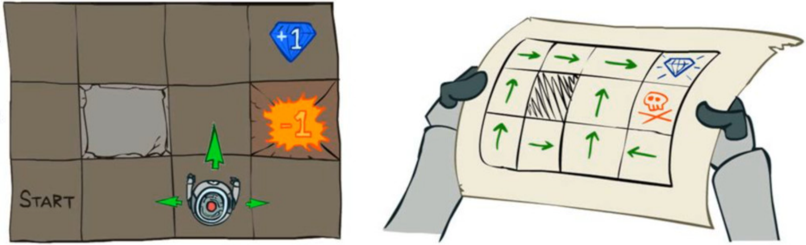
- α: The larger the value, the more the agent learns.
- γ: The larger the value, the more far-sighted the agent is.
- **episode**: How many rounds are you going to try.
- **state**: The environment the agent is in at a certain time.
- **step**: Completing an action is considered a step.
- **action**: Taken by the agent that affect the environment.
- ε-**greedy policy**:

$$\pi(a \mid s) = \begin{cases} \arg\max_a Q(s, a) & \text{probability } 1 - \varepsilon \\ \text{a random action from A(s)} & \text{probability } \varepsilon \end{cases}$$

- r: reward.

Reinforcement learning

Q-learning:

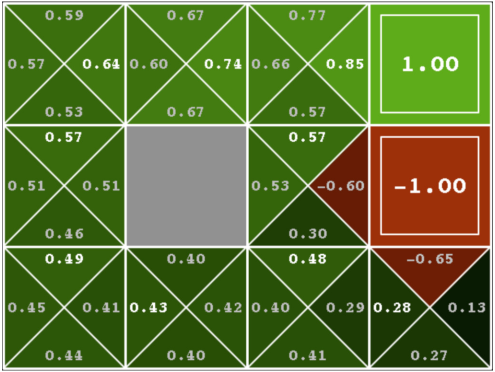


Source: P. Abbeel and D. Klein

An example

Reinforcement learning

Q-learning:



Q-values after 100 iterations

Reinforcement learning

Q-learning:

- Trace the source of line-9 in the algorithm:

$$Q_\pi(s_t, a_t) = E[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \cdots \mid s_t, a_t]$$

$\Longrightarrow$ **Bellman equation**

- Objective:

$$\max_\pi E[\sum_{t=1}^{T} \gamma^t R(S_t, A_t, S_{t+1}) \mid \pi]$$

- Recall line-9:

$$Q(s,a) \leftarrow Q(s,a) + \alpha\underbrace{[\underbrace{r + \gamma \max_{a'} Q(s',a')}_{\text{estimated future rewards}} - \underbrace{Q(s,a)}_{\text{current Q-value}}]}_{\text{error}}$$

Reinforcement learning

Other well-known methods:

- State–action–reward–state–action (SARSA)
- Policy gradients
- Deep Q-Network (DQN)
- ...

Imitation learning

- **What?** Robots learn how to perform tasks by observing or interacting with human experts (i.e. robots are like apprentices and learning from experienced people).
- **Why?** Some tasks may be difficult or impossible to program in traditional ways.
- **How?**
  - Behavioral cloning
  - Inverse reinforcement learning
  - Learning from demonstration
  - ...

Imitation learning

Imitation learning can be seen as a combination of reinforcement learning and **supervised learning**, which accelerates reinforcement learning through the supervision of "experts".

	Imitation learning	**Reinforcement learning**
Optimal strategy[1]	Provided by experts, imitated by robots	Robots need to explore on their own
Reward	Dispensable	Necessary

[1] i.e. $\{s_t, a_t\}$

Imitation learning

Behavioral cloning:

- Experts provide the following data:

$$B = \{(s_1, a_1), \ldots, (s_t, a_t)\}$$

  From the perspective of supervised learning, s is the sample and a is the label.

- Objective function of learning:

$$\theta^* = \arg\min_{\theta} E_{(s,a)B}[L(\pi_\theta(s), a)]$$

  L is the **loss function**. If a is discrete, L can be the maximum likelihood estimation (MLE). If a is continuous, L can be the mean square error (MSE).

Imitation learning

Algorithm Behavioral Cloning

Input:

- Dataset of expert demonstrations $B = \{(s_t, a_t)\}_{t=1}^{T}$
- Policy representation π_θ
- Loss function L

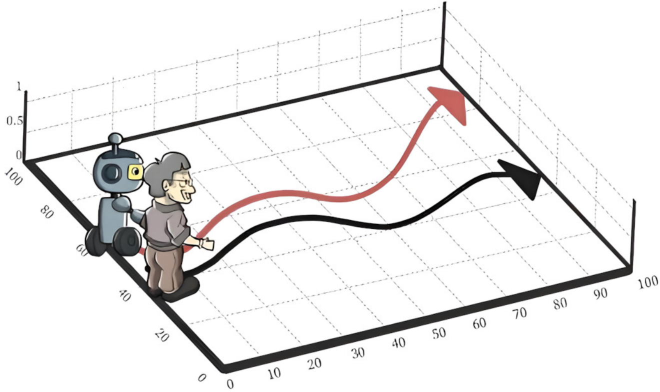
Output:

- Policy parameters θ^*

Initialize policy parameters θ
repeat
    Sample a mini-batch of demonstrations $B^* \subset B$
    for $(s_t, a_t) \in B^*$ **do**
        Predict action $a'_t = \pi_\theta(s_t)$
        Calculate loss $l_t = L(a'_t, a_t)$
    end for
    Update θ using gradient descent: $\theta \leftarrow \theta - \eta \nabla_\theta \sum_{t \in B^*} l_t$
until Convergence criterion met

Imitation learning

Behavioral cloning:

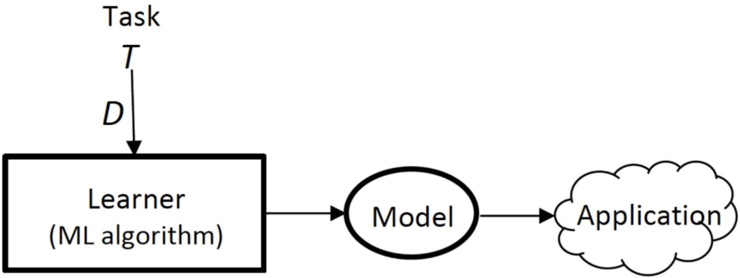


Compounding error problem

Lifelong learning

- **What?** Robots continuously learn from and adapt to the environment throughout their operational lifespan.
- **Why?** By analogy with humans, it's easy to understand that lifelong learning is one of the hallmarks of true robot intelligence.
- **How?**
  - Transfer learning
  - Multi-task learning
  - Online learning
  - Reinforcement learning
  - ...

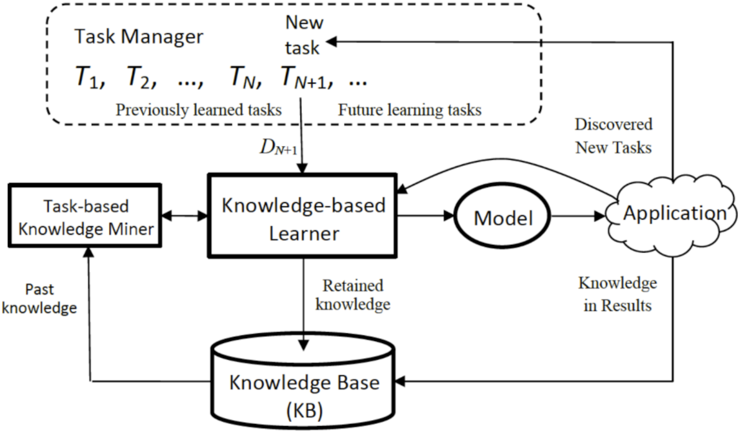
Lifelong learning



The classic machine learning paradigm[2]

[2] Z. Chen and B. Liu. Lifelong machine learning. *Morgan & Claypool Publishers*, 2018.

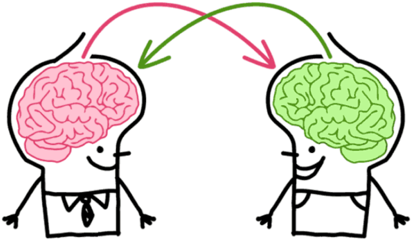
Lifelong learning



The lifelong machine learning system architecture[3]

[3] Z. Chen and B. Liu. Lifelong machine learning. *Morgan & Claypool Publishers*, 2018.

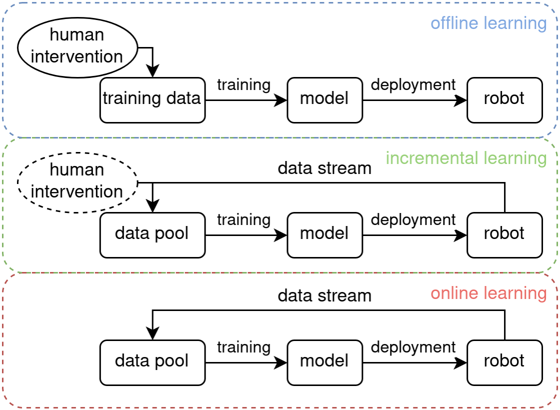
Lifelong learning



Transfer learning: Given a source domain $\mathcal{D}_{\mathcal{S}}$ and learning task $\mathcal{T}_{\mathcal{S}}$, a target domain $\mathcal{D}_{\mathcal{T}}$ and learning task $\mathcal{T}_{\mathcal{T}}$, transfer learning aims to help improve the learning of the target predictive function $f_{\mathcal{T}}(\cdot)$ in $\mathcal{D}_{\mathcal{T}}$ using the knowledge in $\mathcal{D}_{\mathcal{S}}$ and $\mathcal{T}_{\mathcal{S}}$, where $\mathcal{D}_{\mathcal{S}} \neq \mathcal{D}_{\mathcal{T}}$, or $\mathcal{T}_{\mathcal{S}} \neq \mathcal{T}_{\mathcal{T}}$.[4]

[4] J. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 2009.

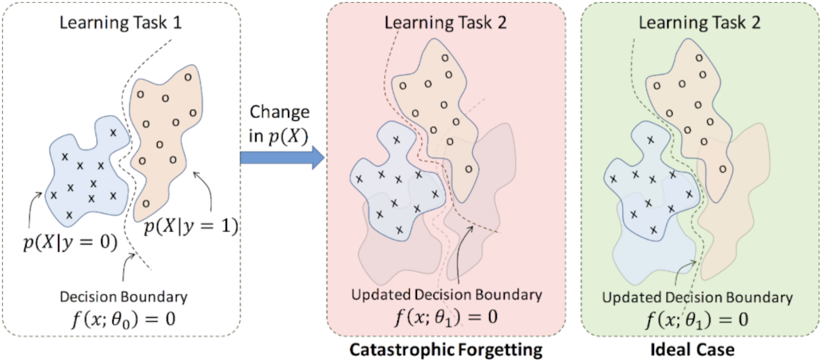
Lifelong learning



Online learning[5]

[5] Z. Yan, L. Sun, T. Krajnik, T. Duckett, N. Bellotto. Towards long-term autonomy: A perspective from robot learning. *AAAI-23 Bridge Program on AI & Robotics*, 2023.

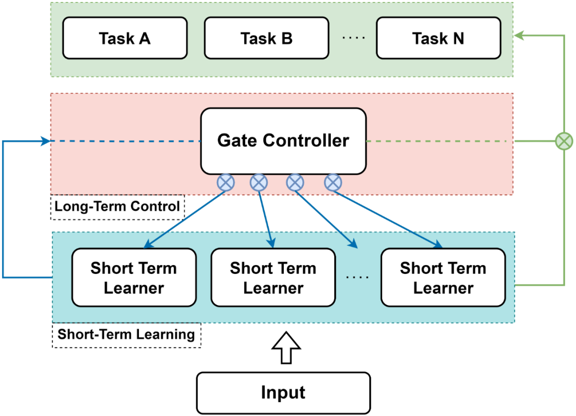
Lifelong learning



Catastrophic forgetting in machine learning[6]

[6] S. Kolouri, N. Ketz, X. Zou, J. Krichmar, and P. Pilly. Attention-based selective plasticity. *arXiv*, 2019.

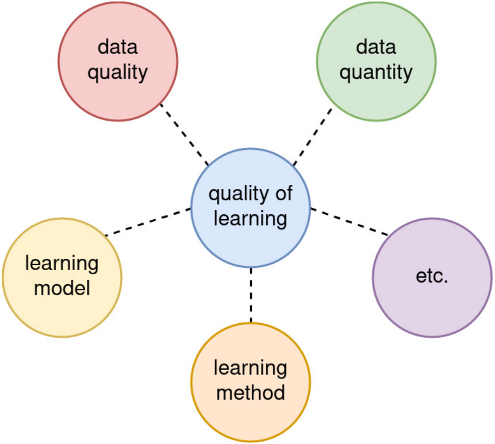
Lifelong learning



Long-Short-Term Online Learning (LSTOL)[7]

[7] R. Yang. Online continual learning for 3D detection of road participants in autonomous driving. *Ph.D. Thesis, UBFC*, 2023.

Data for learning



Robot learning, or machine learning more broadly, requires data.

Data for learning

The robot acquires data from sensors (c.f. Lecture 4).

Summary

- Human, machine, and robot learning
- Reinforcement, imitation, lifelong learning
- Data for learning

The end

Thank you for your attention!

Any questions?