Intrinsic motivation in reinforcement learning

Arthur Aubret

Université Claude-Bernard Lyon 1

15/02/2019

Université Claude Bernard Lyon 1

Learning approaches

	Feedback	No feedback
w interactions	Reinforcement learning	Intrinsic motivation
w/o interaction	Supervised learning	Unsupervised learning

Markov decision process

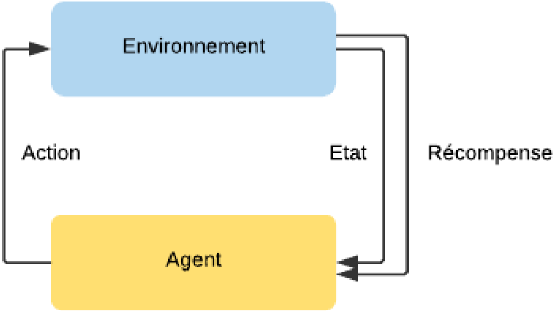


Figure: Markov decision process.

Goal : Maximize $\mathrm{E}_{a \sim \pi, s \sim \rho_\pi}[\sum_{t=0}^{T} R(s_t, a_t)]$

Reinforcement learning [20]

- Learn to accomplish a task.
- Reinforce couples (State,Action) : Which action should I do in my state ?
- No knowledge about environment's transitions $s' = T(s, a)$.
- Depending on S and A : sensori-motor interaction.

https://pythonmachinelearning.pro/
an-overview-of-reinforcement-learning-teaching-machines-to

Simple example

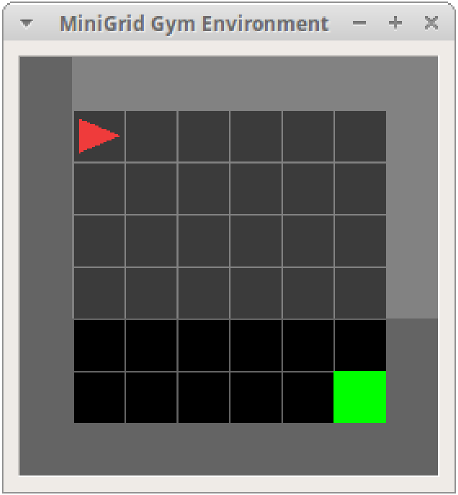


Figure: The agent have to reach the green square.

With deep architectures

Learn from high-dimensionnal input pixels :
https://www.youtube.com/watch?v=oo0TraGu6QY

Exploration

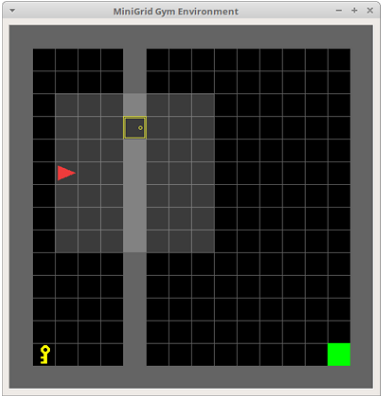


Figure: Simple gridworld with sparse reward, the only reward received is when the agent reaches the green target.

Sample efficiency

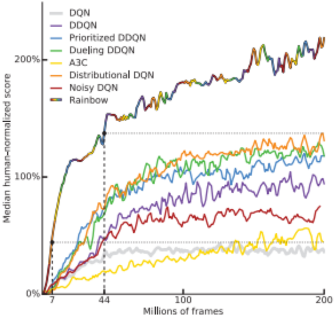


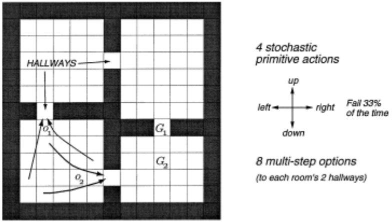
Figure: Learning to play atari [11] [13]. 200k frames = 1h humaine.

In reality : No simulator.

Multi-scale learning

- We are able to take high-level decisions.
- Easier to learn throught 10 high level action than 1000 low-level actions [21].
- How can we generate the content of those actions ? [2] [23]

https://blog.openai.com/learning-a-hierarchy/ [8]



How to solve some of these problems ?

Incorporate intrinsic motivation in the RL framework : Maximize intrinsic reward.

Intrinsic motivation

- Spontaneous exploration.
- Self-organize a curriculum of exploration and learning.
- Enough from evolution ?

Some input properties

From Berlyne [4], use information theory :

- Novelty.
- Complexity.
- Surprise.
- Incongruity.
- Ambiguity.
- Indistinctiveness.

Intrinsic reward [16]

Location of reward

- External reward : The reward comes from outside the organism.
- Internal reward : The reward comes from inside the organism.

Type of reward

- Intrinsic reward : Reward is from the relation/structure of action/observations.
- Extrinsic reward : Reward is from the meaning of action/observations.

Let's try !

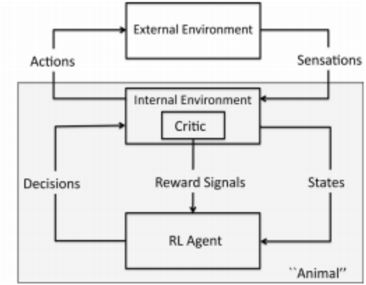
1. I want to have fun with my toys !
2. I need to stop playing because it is childish.
3. I need to stop playing because others are making fun of me.
4. I'm excited to push this unknown magical button.
5. I work to get a good grade at school.
6. I want to get stronger.
7. I'm hungry and I will look for food.
8. I like discovering and understanding math.

Let's try !

1. I want to have fun with my toys ! **Internal and intrinsic.**
2. I need to stop playing because it is childish. **Internal and extrinsic.**
3. I need to stop playing because others are making fun of me. **External and extrinsic.**
4. I'm excited to push this unknown magical button. **Internal and intrinsic**
5. I work to get a good grade at school. **External and extrinsic.**
6. I want to get stronger. **It depends why**
7. I'm hungry and I will look for food. **Internal and extrinsic.**
8. I like discovering and understanding math. **Internal and intrinsic.**

How to combine intrinsic motivation and reinforcement learning ?

The environment, in the RL meaning is not the same as the external environment : **the sources of all of an animal's reward signals are internal to the animal.**[19].



Curiosity

- Go where we can't predict the next observation [17].
  `https://pathak22.github.io/noreward-rl/`
- Maximize error prediction on the prediction error.[10].
- Attracted by states far away from states in memory[18]
  `https://ai.googleblog.com/2018/10/curiosity-and-procrastination-in.html`.
- Prediction error from a random ANN [5].
- Attracted where agent never goes [3][15].

Empowerment

Empowerment quantify the control and influence of an agent in a state[12].

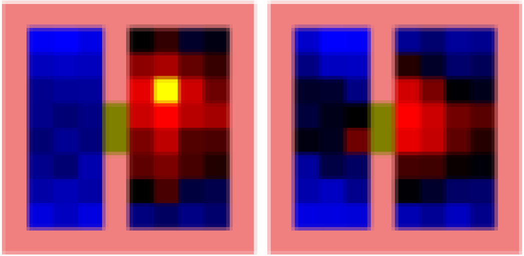


Figure: Empowerment in an environment key-door [14]. In red, states with important empowerment. Empowerment is moving when agent get the key.

Empowerment : example 2

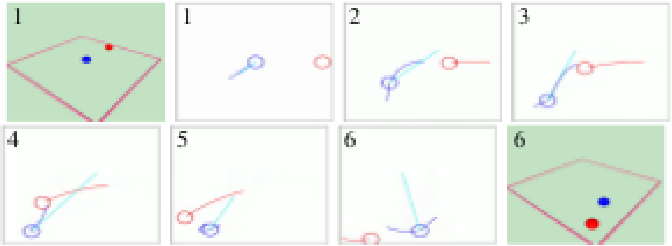


Figure: Empowerment in an environment hunter-prey[14]. In red, the hunter, in blue the prey.

Empowerment : example 3

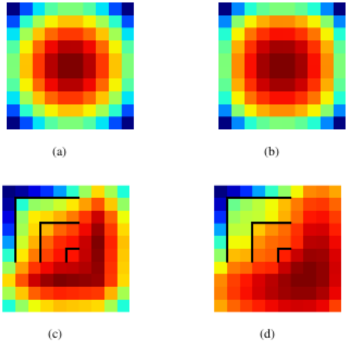


Figure: Empowerment in an environment with barriers [22]. In red, an important empowerment.

Goal generation

Being able to distinct goal accomplishment's subpolicies [9]:

Using theoretic information :
`https://sites.google.com/view/diayn/` [6]
`https://varoptdisc.github.io/`[1].
`https://www.youtube.com/playlist?list=PLEbdzN4PXRGVB8NsPffxsBSOCcWFBMQx3`[7]

My work

Références I

- Achiam, J., Edwards, H., Amodei, D., Abbeel, P.: Variational option discovery algorithms. arXiv preprint arXiv:1807.10299 (2018)
- Bacon, P.L., Harb, J., Precup, D.: The option-critic architecture. In: AAAI. pp. 1726–1734 (2017)
- Bellemare, M., Srinivasan, S., Ostrovski, G., Schaul, T., Saxton, D., Munos, R.: Unifying count-based exploration and intrinsic motivation. In: Advances in Neural Information Processing Systems. pp. 1471–1479 (2016)
- Berlyne, D.E.: Structure and direction in thinking. (1965)

Références II

- Burda, Y., Edwards, H., Storkey, A., Klimov, O.: Exploration by random network distillation. arXiv preprint arXiv:1810.12894 (2018)
- Eysenbach, B., Gupta, A., Ibarz, J., Levine, S.: Diversity is all you need: Learning skills without a reward function. arXiv preprint arXiv:1802.06070 (2018)
- Florensa, C., Duan, Y., Abbeel, P.: Stochastic neural networks for hierarchical reinforcement learning. arXiv preprint arXiv:1704.03012 (2017)
- Frans, K., Ho, J., Chen, X., Abbeel, P., Schulman, J.: Meta learning shared hierarchies. arXiv preprint arXiv:1710.09767 (2017)

Références III

- Gregor, K., Rezende, D.J., Wierstra, D.: Variational intrinsic control. arXiv preprint arXiv:1611.07507 (2016)
- Haber, N., Mrowca, D., Fei-Fei, L., Yamins, D.L.: Learning to play with intrinsically-motivated self-aware agents. arXiv preprint arXiv:1802.07442 (2018)
- Hessel, M., Modayil, J., Van Hasselt, H., Schaul, T., Ostrovski, G., Dabney, W., Horgan, D., Piot, B., Azar, M., Silver, D.: Rainbow: Combining improvements in deep reinforcement learning. In: Thirty-Second AAAI Conference on Artificial Intelligence (2018)

Références IV

- Klyubin, A.S., Polani, D., Nehaniv, C.L.: Empowerment: A universal agent-centric measure of control. In: Evolutionary Computation, 2005. The 2005 IEEE Congress on. vol. 1, pp. 128–135. IEEE (2005)
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., et al.: Human-level control through deep reinforcement learning. Nature 518(7540), 529 (2015)
- Mohamed, S., Rezende, D.J.: Variational information maximisation for intrinsically motivated reinforcement learning. In: Advances in neural information processing systems. pp. 2125–2133 (2015)

Références V

- Ostrovski, G., Bellemare, M.G., Oord, A.v.d., Munos, R.: Count-based exploration with neural density models. arXiv preprint arXiv:1703.01310 (2017)
- Oudeyer, P.Y., Kaplan, F.: How can we define intrinsic motivation? In: Proceedings of the 8th International Conference on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems, Lund University Cognitive Studies, Lund: LUCS, Brighton. Lund University Cognitive Studies, Lund: LUCS, Brighton (2008)
- Pathak, D., Agrawal, P., Efros, A.A., Darrell, T.: Curiosity-driven exploration by self-supervised prediction. In: International Conference on Machine Learning (ICML). vol. 2017 (2017)

Références VI

- Savinov, N., Raichuk, A., Marinier, R., Vincent, D., Pollefeys, M., Lillicrap, T., Gelly, S.: Episodic curiosity through reachability. arXiv preprint arXiv:1810.02274 (2018)
- Singh, S., Lewis, R.L., Barto, A.G., Sorg, J.: Intrinsically motivated reinforcement learning: An evolutionary perspective. IEEE Transactions on Autonomous Mental Development 2(2), 70–82 (2010)
- Sutton, R.S., Barto, A.G.: Reinforcement learning: An introduction, vol. 1. MIT press Cambridge (1998)

Références VII

- Sutton, R.S., Precup, D., Singh, S.: Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning. Artificial intelligence 112(1-2), 181–211 (1999)
- Tiomkin, S., Tishby, N.: A unified bellman equation for causal information and value in markov decision processes. arXiv preprint arXiv:1703.01585 (2017)
- Vezhnevets, A.S., Osindero, S., Schaul, T., Heess, N., Jaderberg, M., Silver, D., Kavukcuoglu, K.: Feudal networks for hierarchical reinforcement learning. arXiv preprint arXiv:1703.01161 (2017)