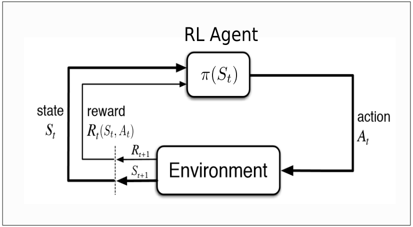
Multitask learning, GCRL and Autotelic agents

Olivier Sigaud

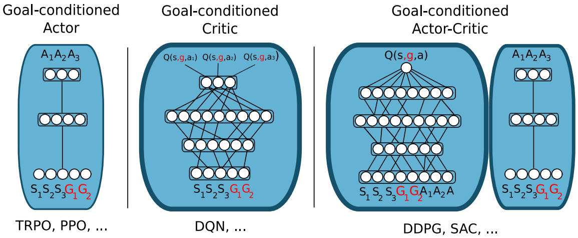
Sorbonne Université
http://people.isir.upmc.fr/sigaud

Introduction



- The standard RL framework addresses a single task which is only specified through a reward function
- RL agents are not autonomous: they depend on the design of an external reward function
- Reward engineering is a known challenge
- Not rich enough to account for many learning phenomena when we face multiple tasks/goals: transfer learning, curriculum, etc.
- Goal-conditioned RL (GCRL) is a framework to account for this richer context.
- Outline:
  - The GCRL framework
  - Application to autotelic agents

GCRL



- Universal Value Function Approximators (anterior to DQN)
- Learned with standard Q-LEARNING or ACTOR-CRITIC schemes
- Main advantage: generalization over the goal space

Schaul, T., Horgan, D., Gregor, K., & Silver, D. (2015) Universal value function approximators. In *International Conference on Machine Learning* (pp. 1312–1320)

Goal-related reward function

- A goal is the conjonction of a constraint satisfaction (goal achievement) function on the state and a reward function
- Dense reward functions: decreasing function of the distance to a goal state
- Sparse reward functions: 1 if the state is achieved, 0 otherwise (or 0/-1 to favor exploration)

Colas, C., Karch, T., Sigaud, O., and Oudeyer, P.-Y. (2022) Autotelic agents with intrinsically motivated goal-conditioned reinforcement learning: a short survey. *Journal of Artificial Intelligence Research*, 74:1159–1199

Learning from failures

- Without finding reward, an RL agent learns nothing

- Consider a learning agent whose goal is to reach a particular outcome
- In the beginning, this agent may often fail
- The failed experiment produced another outcome than the expected one
- But this can be turned into useful knowledge
- This is the essence of Hindsight Experience Replay (HER)

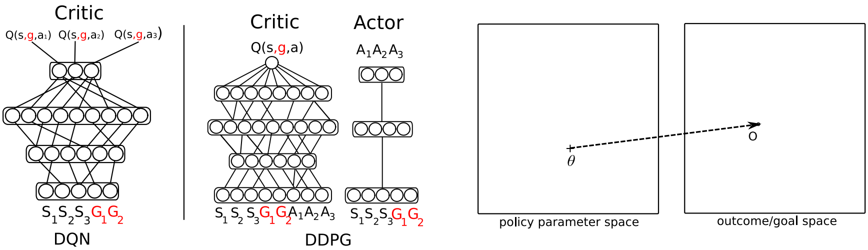
Motivation

- HER might be useful in two different contexts:

- The agent targets a difficult goal, i.e. a sparse reward RL problem
- Without a reward signal, a model-free RL agent produces an inefficient random search
- HER provides additional reward signals

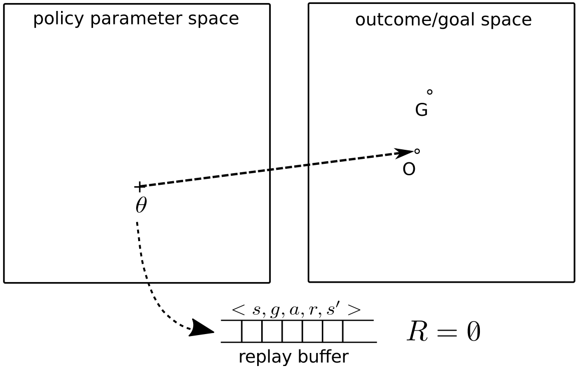
- The agent targets many goals
- Learning to achieve each goal in isolation is sample inefficient
- The HER agent learns unexpected goals through its failures

Four components



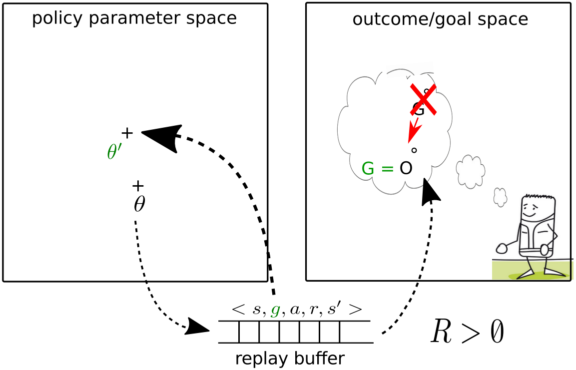
1. Goal conditioned policies
2. Mapping from policy parameter space to outcome space
3. Any RL algorithm (DQN, DDPG, TD3, PPO, SAC, ...)
4. A special replay buffer with goal substitution

General mechanism (1)



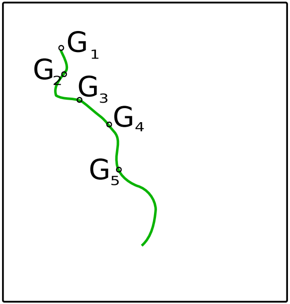
- The agent targets a goal G as outcome
- The policy π_θ produces another outcome O
- The trajectory is stored but produces no reward

General mechanism (2)



- The agent pretends it is targetting O
- HER relabels the stored trajectory with the obtained outcome
- This propagates value in the (state, action) space through generalization
- And the agent competence will increase over unseen goals

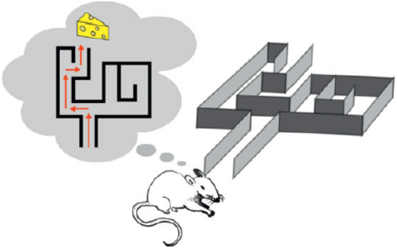
When the goal is a state



outcome/goal space

- If the goal space is the state space, HER may set as goal any state along the trajectory
- Trade-off between replaying more and trying more new actions (over-fitting to replays)

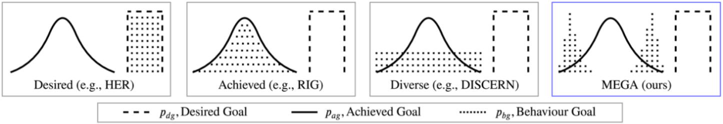
Hindsight Experience Replay: properties



- Looks like a model-based process, but without a model
- Provides an implicit form of curriculum learning
- Provides an additional reward signal
- Avoids dense reward signals

Doll, B. B., Simon, D. A., and Daw, N. D. The ubiquity of model-based reinforcement learning. *Current opinion in neurobiology*, 22(6):1075–1081, 2012

Desired Goals and achieved goals



Explain, cover the literature

- Key question 1: covering: how to set the desired goals to reach more goals
- Key question 2: performance: how to better reach the achieved goals?

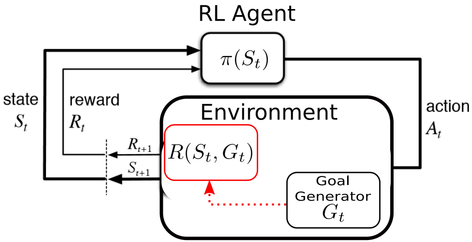
Pitis, S., Chan, H., Zhao, S., Stadie, B., and Ba, J. (2020) Maximum entropy gain exploration for long horizon multi-goal reinforcement learning. In *International Conference on Machine Learning*, pages 7750–7761. PMLR

Campos, V., Trott, A., Xiong, C., Socher, R., Giro-i Nieto, X., and Torres, J. (2020) Explore, discover and learn: Unsupervised discovery of state-covering skills. In *International Conference on Machine Learning*, pages 1317–1327. PMLR

Articulation

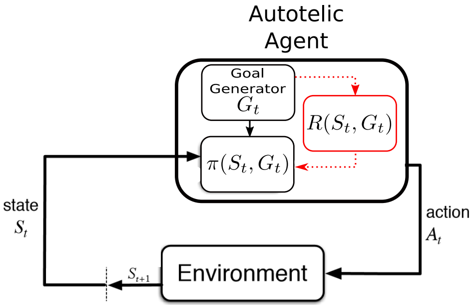
- In GCRL, the desired goal can come from the environment or from the agent
- In standard multitask RL, goals come from the environment
- In autotelic learning, the agent generates its own goals
- The reward signal becomes internal
- This becomes a specific instance of unsupervised RL

The GoalEnv view



- In OpenAI gym, and SB3 (as most librairies?) the common view of GCRL

Autotelic Agents

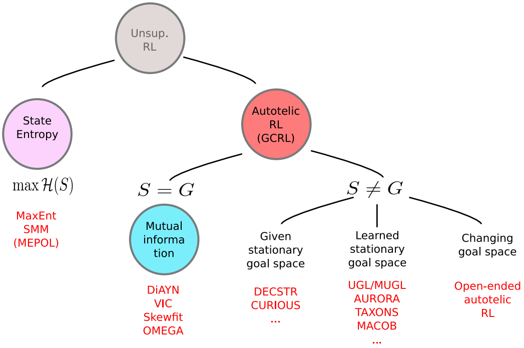


- Autotelic agents: agents equipped with forms of intrinsic motivations that enable them to represent, self-generate and pursue their own goals
- Goal generator based on: diversity, hierarchical RL, curriculum learning, social signals...

Colas, C., Oudeyer, P.-Y., Sigaud, O., Fournier, P., & Chetouani, M. (2019) CURIOUS: Intrinsically motivated multi-task, multi-goal reinforcement learning. Edité dans *International Conference on Machine Learning (ICML)*, pages 1331–1340

Colas, C., Karch, T., Sigaud, O., and Oudeyer, P.-Y. Intrinsically motivated goal-conditioned reinforcement learning: a short survey. *arXiv preprint arXiv:2012.09830*, 2020

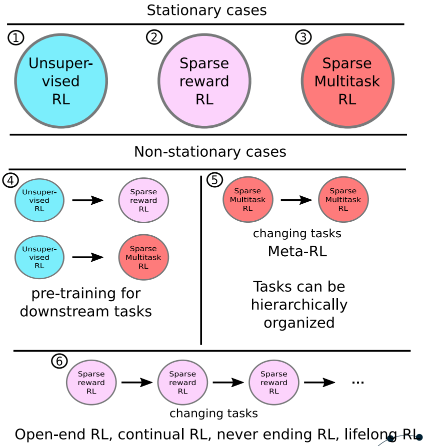
Unsupervised RL and autotelic RL



- Open-ended autotelic RL: the agent defines its own goal spaces, its own state and actions spaces in a reward free environment (the ultimate framework!)

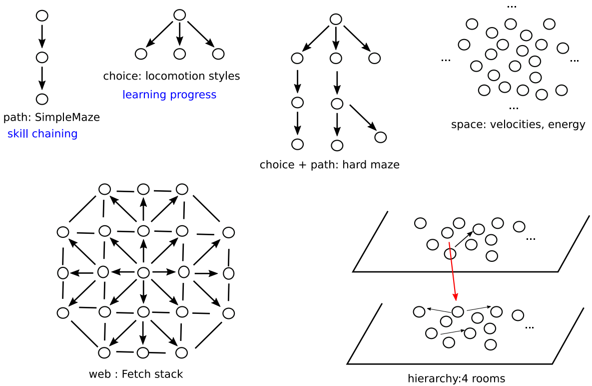
Reward-related typology

- Papers investigating Hard explorations problems may address different types of problems
- Vocabulary: unsupervised RL = task-agnostic exploration, reward-free exploration
- Specificity of Open-ended RL: learn a state and action space from lower level sensor/actuators [Doncieux et al., 2018]



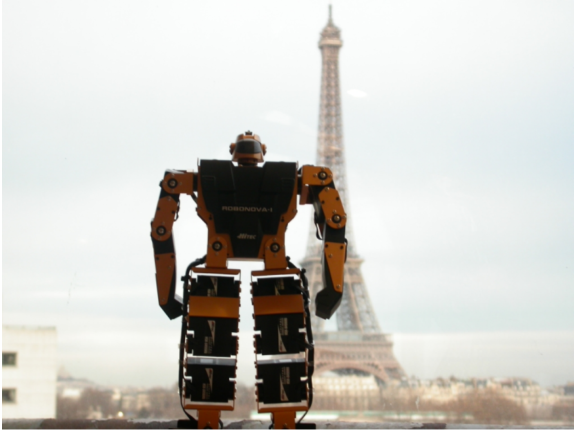
Doncieux, S., Filliat, D., Díaz-Rodríguez, N., Hospedales, T., Duro, R., Coninx, A., Roijers, D. M., Girard, B., Perrin, N., & Sigaud, O. (2018) Open-ended learning: a conceptual framework based on representational redescription. *Frontiers in Robotics and AI*, 12

Goal topologies



- Some mechanisms are topology-specific

Any question?



Send mail to: `Olivier.Sigaud@upmc.fr`

ISIR
INSTITUT
DES SYSTÈMES
INTELLIGENTS
ET DE ROBOTIQUE

Campos, V., Trott, A., Xiong, C., Socher, R., Giro-i Nieto, X., and Torres, J. (2020).
Explore, discover and learn: Unsupervised discovery of state-covering skills.
In *International Conference on Machine Learning*, pages 1317–1327. PMLR.

Colas, C., Karch, T., Sigaud, O., and Oudeyer, P.-Y. (2020).
Intrinsically motivated goal-conditioned reinforcement learning: a short survey.
arXiv preprint arXiv:2012.09830.

Colas, C., Karch, T., Sigaud, O., and Oudeyer, P.-Y. (2022).
Autotelic agents with intrinsically motivated goal-conditioned reinforcement learning: a short survey.
Journal of Artificial Intelligence Research, 74:1159–1199.

Colas, C., Oudeyer, P.-Y., Sigaud, O., Fournier, P., and Chetouani, M. (2019).
CURIOUS: Intrinsically motivated multi-task, multi-goal reinforcement learning.
In *International Conference on Machine Learning (ICML)*, pages 1331–1340.

Doll, B. B., Simon, D. A., and Daw, N. D. (2012).
The ubiquity of model-based reinforcement learning.
Current opinion in neurobiology, 22(6):1075–1081.

Doncieux, S., Filliat, D., Díaz-Rodríguez, N., Hospedales, T., Duro, R., Coninx, A., Roijers, D. M., Girard, B., Perrin, N., and Sigaud, O. (2018).
Open-ended learning: a conceptual framework based on representational redescription.
Frontiers in Robotics and AI, 12.

Pitis, S., Chan, H., Zhao, S., Stadie, B., and Ba, J. (2020).
Maximum entropy gain exploration for long horizon multi-goal reinforcement learning.
In *International Conference on Machine Learning*, pages 7750–7761. PMLR.

Schaul, T., Horgan, D., Gregor, K., and Silver, D. (2015).
Universal value function approximators.
In Bach, F. R. and Blei, D. M., editors, *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, volume 37 of *JMLR Workshop and Conference Proceedings*, pages 1312–1320. JMLR.org.