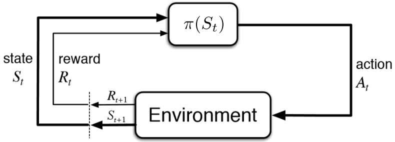
GCRL: formal frameworks and core concepts

Olivier Sigaud

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

Starting point: the RL framework

- All this lesson builds on [Colas et al., 2022]
- Goal: "a cognitive representation of a future object that the organism is committed to approach." [Elliot and Fryer, 2008]
- To define a goal, we need to emulate "an organism"
- An RL agent does so. It is "committed to approach future objects" (through the reward)
- We build on the MDP framework: $M =< S, A, T, R, \gamma >$

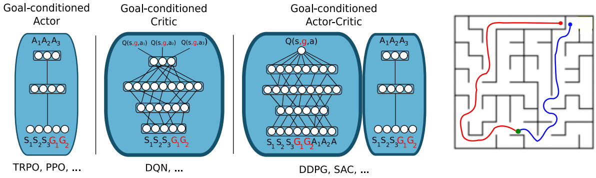


- The MDP defines a task: the problem the agent has to solve
- But we need to give the agent a cognitive representation (a goal)

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

Goal representation: basic idea

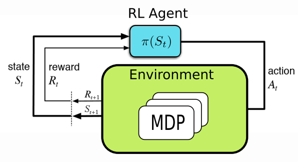
- We want to endow an agent with a goal representation
- The policy can be conditioned on a state **and a goal**



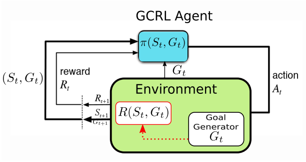
- Condition the policy and/or critic depending on the algorithm
- Main advantage: generalization over the state space **and the goal space**
- Provided some local continuity (not always present, e.g. maze example)

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

Different frameworks: multitasks vs multigoals



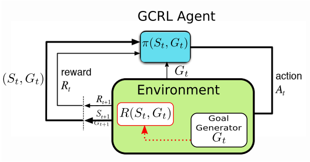
Standard RL agent in a multitask environment



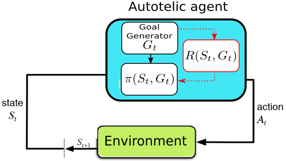
Multigoal RL agent in a GoalEnv

- In the multitask framework:
  - The agent faces a set of MDPs
  - These MDPs can differ in any MDP component $< S, A, T, R, \gamma >$
  - The agent may have a representation of which MDP it faces, or not
- In the multigoal, GoalEnv framework:
  - Goal-MDP: $M =< S, G, A, T, R, \gamma >$
  - G is the goal space, R is a goal-dependent reward function
  - The extended MDP provides the goal and the reward for solving it

Different frameworks: goalEnv vs autotelic



Multigoal RL agent in a GoalEnv



Autotelic agent in a non-rewarded env

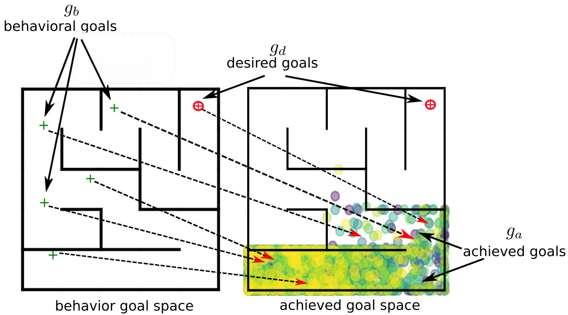
- In the multigoal, GoalEnv framework:
  - The environment provides the goal, the agent is rewarded for solving it
  - These elements are defined by the experimenter
- In the Autotelic learning framework:
  - MDP: $M =< S, A, T, R_g, \gamma >$
  - There is a single task, corresponding to the underlying MDP
  - The goal g is not provided by the environment, but set by the agent
  - The goal-dependent reward function R_g defines the corresponding reward
  - R_g is often experimenter-defined, but not always (see VLMs)
  - If the goal space is pre-defined, an intermediate framework $M =< S, G, A, T, R_g, \gamma >$ is possible

Goals and goal spaces

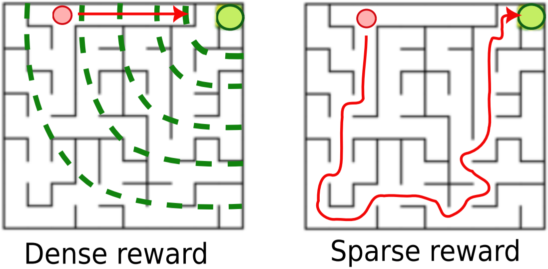

- A goal is a point in a goal space, or a member of a discrete set of goals
- A goal space can be given, or learned (e.g. as the output space of a neural network, or as an embedding)
- To determine which goal was achieved, one needs a goal achievement function $g = Ach(\tau)$
- Can be a function of the current state, or of the full trajectory (more general)
- The goal space is often the state space
- If goal space = state space, $Ach(.)$ is the identity, often with a tolerance ϵ
- Defining $g = Ach(\tau)$ can be as hard as defining reward functions

Desired, behavioral and achieved goals

- We need to distinguish three types of goals:
  - desired goals g_d: goals we ultimately want to achieve
  - behavioral goals g_b: goals we input to the policy
  - achieved goals g_a: goals given by $g_a = Ach(\tau)$

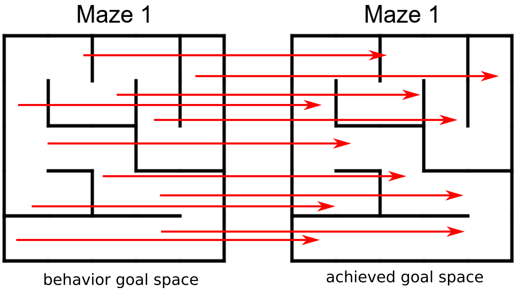


Goal-dependent reward function



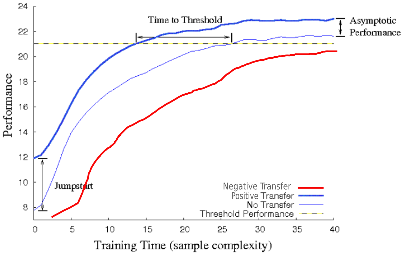
- Goal-dependent reward function: given a behavior goal g_b
  - Sparse reward functions: 1 if the goal is achieved, 0 otherwise (or 0/-1 to favor exploration)
  - Dense reward functions: decreasing function of the distance between the state and g_b (assumes projecting the two in the same space)
  - Research in autotelic agents often uses sparse rewards
  - As they are simpler to define and less prone to deceptive gradients

Goal-conditioned learning: a distributional perspective



- Desired goals could be represented as a distribution $p(g_d)$
- If uniform over the goal space (coverage objective), can be ignored
- Behavioral goals could be sampled from a distribution $p(g_b)$
- Before the agent gets expert, the achieved goal is not the behavioral goal
- One perspective on GCRL is to try to get them equal (learn identity mapping between behavioral and achieved goals distributions)

Transfer learning and catastrophic forgetting

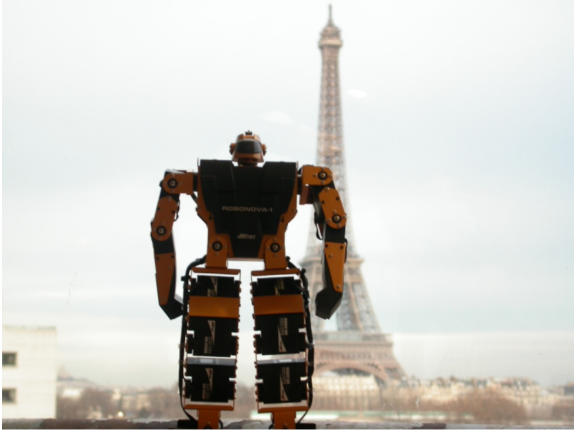


- Positive transfer: dark blue is above light blue
- Different measures of transfer efficiency
- Negative transfer affects performance on the next task
- Catastrophic forgetting affects performance on the previous task
- Continual learning: leverage positive transfer and mitigate catastrophic forgetting

Taylor, M. E. and Stone, P. (2009) Transfer learning for reinforcement learning domains: A survey. *Journal of Machine Learning Research*, 10(7)

Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., and Wermter, S. (2019) Continual lifelong learning with neural networks: A review. *Neural networks*, 113:54–71

Any question?



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

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.

Elliot, A. J. and Fryer, J. W. (2008).
The goal construct in psychology.
Handbook of motivation science, 18:235–250.

Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., and Wermter, S. (2019).
Continual lifelong learning with neural networks: A review.
Neural networks, 113:54–71.

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

Taylor, M. E. and Stone, P. (2009).
Transfer learning for reinforcement learning domains: A survey.
Journal of Machine Learning Research, 10(7).