Goal-Conditioned Reinforcement Learning

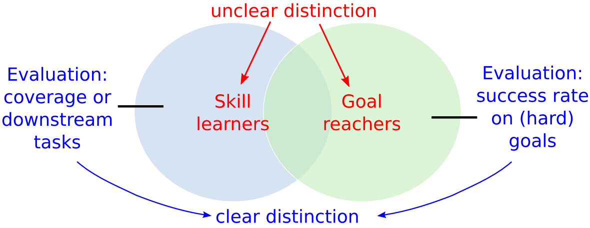
Typology of perspectives

Olivier Sigaud

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

Outline

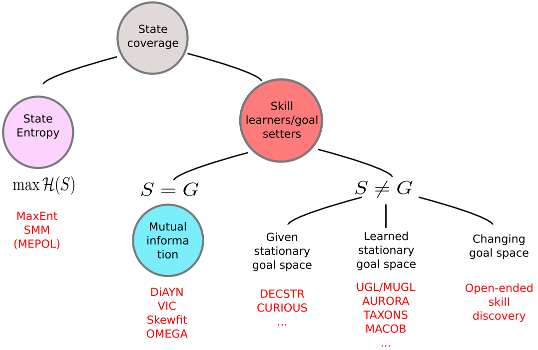
- Four perspectives:
  - Perspective 1: the skill learning (or unsupervised RL) perspective
  - Perspective 2: the setter-solver perspective
  - Perspective 3: the contextual RL perspective
  - Perspective 4: the sequential (or hierarchical) RL perspective
- Main focus: distinguishing skill learners from goal reachers



- Classification driven by the adressed problems
- Autotelic agents $\sim$ skill learner: absence of external reward
- Relation to contextual RL: a goal is a specific form of context

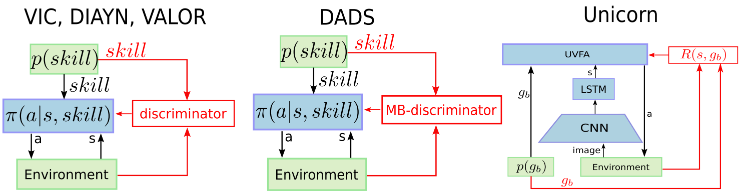
The unsupervised RL perspective

Unsupervised reinforcement learning: goal spaces



- The goal space can be absent, given, fixed and learned, or evolving
- The general objective is to cover the space of possible goals
- Downstream objective: pretrain before learning to reach specific, harder goals

Discovering a diversity of skills

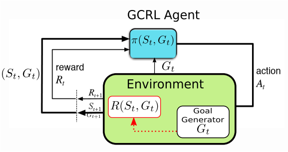


- General perspective: maximize goal space (or state space) coverage
- Or mutual information or empowerment
- Skill discovery → maximize diversity: VIC, DIAYN, VALOR, DADS
- The setter is used to generate a set of diverse trajectories

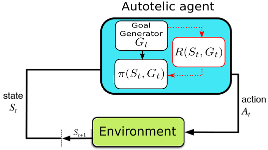
Klyubin, A. S., Polani, D., and Nehaniv, C. L. (2005) Empowerment: A universal agent-centric measure of control. In *2005 IEEE congress on evolutionary computation*, volume 1, pages 128–135. IEEE

The setter-solver perspective

Reminder: GoalEnv vs Autotelic agents



Multigoal RL agent in a GoalEnv



Autotelic agent in a non-rewarded env

- The setter-solver perspective distinguishes:
  - a goal setter, which can be the agent or the environment
  - a goal solver, which is a goal-conditioned policy learned with RL
- GoalEnv: when the setter is the environment
- Autotelic: when the setter is the agent
- The general objective is to endow an agent with fixed or open-ended goal reaching capabilities

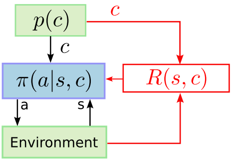
The growth and diversity of solvers

Algorithm	Reference	Solver Algo	Archi	HER
DG-LEARNING	Kaelbling (1993)	Q-LEARNING	tabular	○
UVFA	Schaul et al. (2015)	HORDE or DQN-like	not found	○
ASP	Sukhbaatar et al. (2017)	REINFORCE, TRPO	[50, 50]	○
TSCL	Matiisen et al. (2017; 2019)	PPO	CNN + LSTM	○
HPG	Rauber et al. (2017)	Policy gradient	(CNN) + [256, 256, 256]	○
Many-Goals	Veeriah et al. (2018)	A2C-like	CNN + 512 + embedding	○
GoalGAN	Florensa et al. (2018)	TRPO + GAE	[32, 32]	○
UNICORN	Mankowitz et al. (2018)	DQN-like	CNN + LSTM	○
CURIOUS	Colas et al. (2018)	DDPG	[256, 256, 256]	•
RIG	Nair et al. (2018)	TD3	not found	••
DISCERN	Warde-Farley et al. (2018)	$Q(\lambda)$, IMPALA	CNN + LSTM	•
LEAP	Nasiriany et al. (2019)	TD3	[400, 300]	•
RPL	Gupta et al. (2019)	Natural policy gradient	[256, 256]	○
CER	Liu et al. (2019)	DDPG, PPO	[256, 256, 256]	••
CWYC	Blaes et al. (2019)	SAC or DDPG + HER	[256, 256] or [256, 256, 256]	•
HGG	Ren et al. (2019)	DDPG	[256, 256, 256]	••
SKEW-FIT	Pong et al. (2019)	SAC	[400, 300]	○
SETTER-SOLVER	Racaniere et al. (2019)	IMPALA	CNN + LSTM	○
GCSL	Ghosh et al. (2019)	supervised	[400, 300]	○
EDL	Campos et al. (2020)	PPO	[128, 128]	○
AMIGO	Campero et al. (2020)	IMPALA	CNN + LSTM	○
MEGA/OMEGA	Pitis et al. (2020)	DDPG	[512, 512, 512]	•
HILBERT	Pierrot et al. (2020)	SAC	LSTM+Embedding	○
VGCRL	Choi et al. (2021)	SAC	[256, 256]	••
RIS	Chane-Sane et al. (2021)	SAC	[256, 256] or CNN	•
HIGL	Kim et al. (2021)	TD3	[300, 300]	○
UPSIDE	Kamienny et al. (2021)	SAC or TD3,	[64, 64] or [256, 256]	○
DECSTR	Akakzia et al. (2021)	SAC	DeepSets or GNNs	•
HRAC	Zhang et al. (2020; 2022)	TD3, A2C	[300, 300]	○
DCIL-II	Chenu et al. (2022)	SAC	[512, 512, 512]	•
SVGG	Castanet et al. (2023)	DDPG	[512, 512, 512]	•

- Many different solver algorithms, with growing architectures (Moore's law)
- We can recognize image-based solvers (using CNN, ...)
- HER is not so present
- Transformers and diffusion policies are coming (not shown)
- Beyond RL solvers: imitation learning, evolutionary methods...

The contextual RL perspective

Contextual RL



- In the same environment, the agent distinguishes various contexts
- The context distribution is unknown, it comes from the environment
- Instead of discrimination, the policy maximize the context-conditioned reward
- Precursor: [Kupcsik et al., 2013], Formalization: [Hallak et al., 2015]
- Recent instances: CARE [Eimer et al., 2021], SPaCE [Sodhani et al., 2021]

Kupcsik, A. G., Deisenroth, M. P., Peters, J., and Neumann, G. (2013) Data-efficient generalization of robot skills with contextual policy search. In *Twenty-Seventh AAAI Conference on Artificial Intelligence*

Hallak, A., Di Castro, D., and Mannor, S. (2015) Contextual Markov decision processes. *arXiv preprint arXiv:1502.02259*

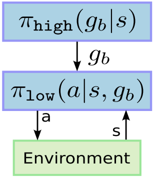
Eimer, T., Biedenkapp, A., Hutter, F., and Lindauer, M. (2021) Self-paced context evaluation for contextual reinforcement learning. In *International Conference on Machine Learning*, pages 2948–2958. PMLR

Sodhani, S., Zhang, A., and Pineau, J. (2021) Multi-task reinforcement learning with context-based representations. In *International Conference on Machine Learning*, pages 9767–9779. PMLR

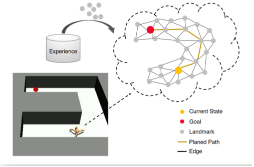
The sequential RL perspective

Sequential setters

Feudal



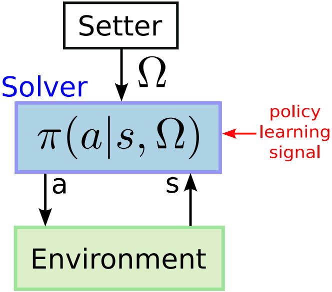
Graph-based



...

- Sequential setter: when performing a trajectory, the agent triggers a sequence of behavior goals before it reaches the final desired goal
- Hierarchical setters are the most common sequential setters
- Specificity: there is a high level policy
- Counter-examples: planning with a graph or list of goals is not truly hierarchical
- Hierarchical reinforcement learning (HRL) is the focus of another lecture
- So just a quick overview here, from the GCRL perspective

Unifying perspective: General (G)CRL template

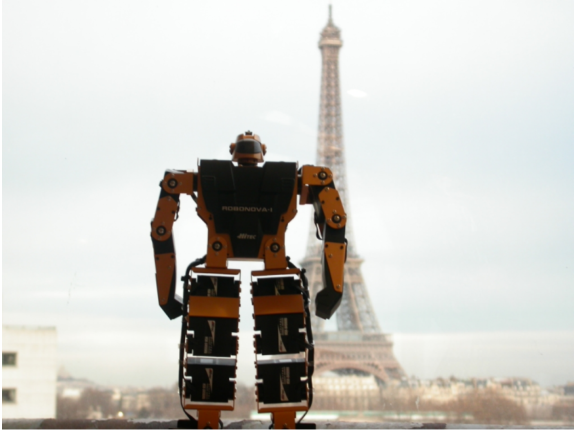


- Green is fixed, blue is learned
- The goal setter can be fixed or learned
- In the setter-solver perspective, Ω is a goal
- In the unsupervised RL perspective, Ω is a skill
- In the contextual perspective, Ω is a context
- Our main focus will be the setter-solver perspective

Upcoming classification

- Value of the setter-solver perspective: all setters of a class could be compared using an identical solver
- We ignore the differences between solvers
- We distinguish six classes of setters:
  - Non image-based skill discovery setters
  - Image-based skill discovery setters
  - Non image-based goal setters
  - Image-based goal setters
  - Non image-based sequential goal setters
  - Image-based sequential goal setters
- For each class, we mention:
  - The input type (state, object, image, or a combination)
  - Evaluation criteria
  - For goal reachers, the nature of the target goal set
  - Sub-types of setters
  - For image-based setters, the type of latent state encoder

Any question?



Send mail to: Olivier.Sigaud@upmc.fr

Eimer, T., Biedenkapp, A., Hutter, F., and Lindauer, M. (2021).
Self-paced context evaluation for contextual reinforcement learning.
In *International Conference on Machine Learning*, pages 2948–2958. PMLR.

Hallak, A., Di Castro, D., and Mannor, S. (2015).
Contextual Markov decision processes.
arXiv preprint arXiv:1502.02259.

Klyubin, A. S., Polani, D., and Nehaniv, C. L. (2005).
Empowerment: A universal agent-centric measure of control.
In *2005 ieee congress on evolutionary computation*, volume 1, pages 128–135. IEEE.

Kupcsik, A. G., Deisenroth, M. P., Peters, J., and Neumann, G. (2013).
Data-efficient generalization of robot skills with contextual policy search.
In *Twenty-Seventh AAAI Conference on Artificial Intelligence*.

Sodhani, S., Zhang, A., and Pineau, J. (2021).
Multi-task reinforcement learning with context-based representations.
In *International Conference on Machine Learning*, pages 9767–9779. PMLR.