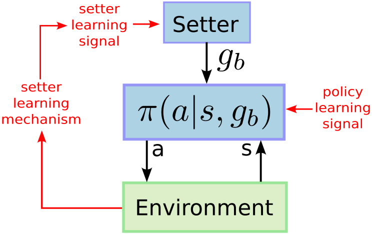
Goal-Conditioned Reinforcement Learning

State-based goal reaching setters

Olivier Sigaud

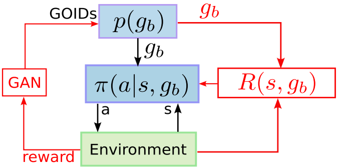
Sorbonne Université
http://people.isir.upmc.fr/sigaud

Goal reaching setters



- By contrast with the skill discovery family, most GCRL methods evolve the behavioral goal distribution
- They have a curriculum: set behavioral goals so that achieved goals finally reach desired goals
- They often use HER
- Examples: Goal GAN, Skew-Fit, setter-solver, MEGA, SVGG, ...

Goal GAN



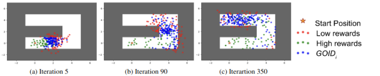


Figure 3. Goals that, at iterations i, our algorithm trains on - 200 sampled from Goal GAN, 100 from replay. Green goals satisfy $\bar{R}^g(\pi_i) \geq R_{\max}$. Blue ones have appropriate difficulty for the current policy $R_{\min} \leq \bar{R}^g(\pi_i) \leq R_{\max}$. The red ones have $R_{\min} \geq \bar{R}^g(\pi_i)$.

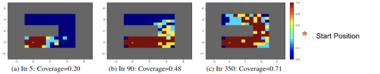
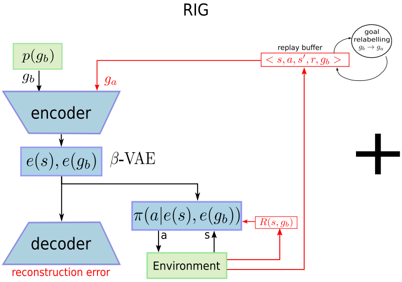


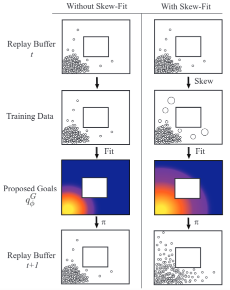
Figure 4. Visualization of the policy performance (same policy training as in Fig. 3). For illustration purposes, each grid cell is colored according to the expected return achieved when fixing its center as goal: Red indicates 100% success; blue indicates 0% success.

- The GAN discriminates between easy and hard goals (with fixed thresholds...)
- GOIDs: Goals of Intermediate Difficulty
- The solver uses TRPO + GAE with 2 hidden layers of size 32
- GANs are known to be unstable, expensive and hard to tune...
- Still does not use HER

Florensa, C., Held, D., Geng, X., and Abbeel, P. (2018) Automatic goal generation for reinforcement learning agents. In *International conference on machine learning*, pages 1515–1528. PMLR

Skew-Fit

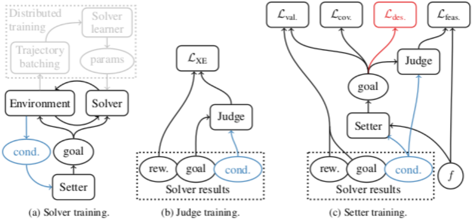




- Based on RIG (see image-based setters, works with images)
- In RIG, behavioral goals try to fit the distribution of achieved goals
- Lacks a mechanism to expand the distribution of achieved goals
- Skew-Fit increases the probability to target rare states as goals

Pong, V. H., Dalal, M., Lin, S., Nair, A., Bahl, S., and Levine, S. (2019) Skew-fit: State-covering self-supervised reinforcement learning. *arXiv preprint arXiv:1903.03698*

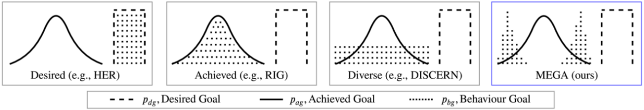
SETTER-SOLVER

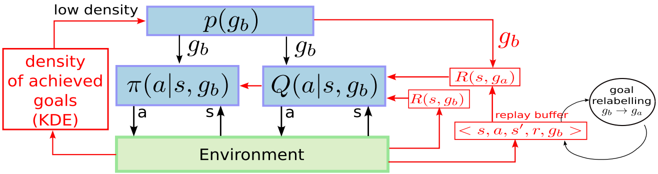


- Builds on Goal GAN, addresses dynamic environments
- The goal setter ensures validity, diversity and feasability of goals
- Chooses behavioral goals close to achieved goals
- The judge predicts the probability that the agent reaches the goal
- Relies on an invertible network (RNVP) to map from the latent to the goal space
- Limits the modeling power and problematic in case of discontinuities

Racaniere, S., Lampinen, A., Santoro, A., Reichert, D., Firoiu, V., and Lillicrap, T. (2019) Automated curriculum generation through setter-solver interactions. In *International Conference on Learning Representations*

MEGA

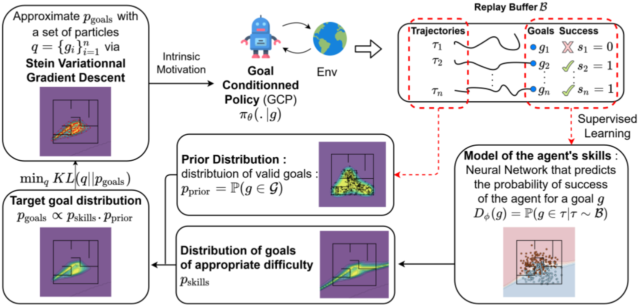




- Instead of the GAN, uses a KDE model of the density of achieved goals
- Goals are sampled from regions of low density
- Needs to avoid invalid goals (ugly hacks)
- Reaches many goals, but does not master them (catastrophic forgetting)

Pitis, S., Chan, H., Zhao, S., Stadie, B., and Ba, J. (2020) Maximum entropy gain exploration for long horizon multi-goal reinforcement learning. In *Proceedings of the 37th International Conference on Machine Learning*, pages 7750–7761

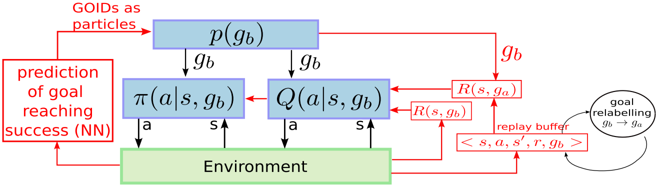
SVGG



- Learns a predictive model of the capability to reach goals
- Models the set of most unpredictable goals with SVGD
- Combines with a learned model of valid goals
- Target goals are a set of particles which repulse each other
- Endowed with a good recovery property

Castanet, N., Lamprier, S., and Sigaud, O. (2023) Stein variational goal generation for reinforcement learning in hard exploration problems. *ICML*

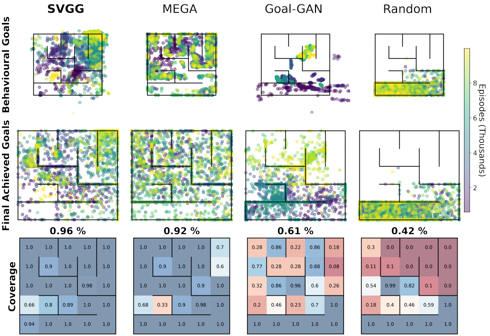
SVGG architecture



- Uses
  - The GOIDs idea
  - SVGD to approximate the goal distribution with particles
  - DDPG as solver
  - HER to accelerate solver convergence

Liu, Q. and Wang, D. (2016) Stein variational gradient descent: A general purpose Bayesian inference algorithm. *arXiv preprint arXiv:1608.04471*

SVGG, GoalGAN, MEGA and others



- Random goal sampling remains stuck (a good curriculum is needed)
- Achieved goal are more spread than behavioral goals
- SVGG optimizes success coverage: reaching behavioral goals everywhere (blue)

Evaluation criteria and types of setters

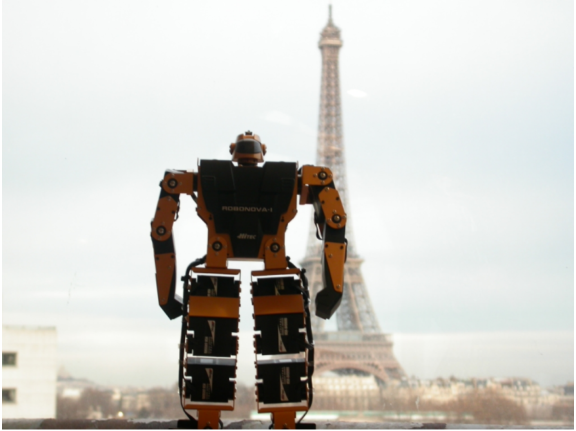
- Evaluation: which set of goals is the agent evaluated on?
  - **Single**: the agent must reach one desired goal, generally hard to reach
  - **Final distrib.**: the agent must reach a fixed distribution of desired goal
  - **Current distrib.**: the agent must reach a distribution of temporary goals
  - **Coverage**: the agent must reach as many goal from the goal space as possible
- For skill discovery agents, often coverage, but evaluation relies more on downstream tasks
- Types: which setter mechanism?
  - **Achiever**: expand the behavioral goal distribution by going beyond the current achieved goal distribution
  - **HER-based**: uses HER rather than a curriculum (or both)
  - **LP-based**: samples goals based on learning progress
  - **Particle-based**: evolves the goal sampling distribution as a set of particles

Selected algorithms

Algorithm	Reference	Input type	Goal set	Type of Setter
DG-LEARNING	[Kaelbling, 1993]	tabular states	Coverage	All
UVFAs	[Schaul et al., 2015]	states	Final distrib.	Uniform
Goal GAN	[Florensa et al., 2018]	states	Coverage	Achiever
CURIOUS	[Colas et al., 2018]	objects	Coverage	LP-based
TSCL	[Matiisen et al., 2019]	task label	Single	Tutor-based
CER	[Liu et al., 2019]	states	Single	HER-based
DHER	[Fang et al., 2019a]	states	Single	HER-based
CHER	[Fang et al., 2019b]	states	Final distrib.	HER-based + Acheiver
HGG	[Ren et al., 2019]	object	Final distrib.	HER-based
GCSL	[Ghosh et al., 2019]	states	Single	Fixed
MEGA/OMEGA	[Pitis et al., 2020]	states, objects	Final distrib.	Achiever
DESCTR	[Akakzia et al., 2021]	objects	Coverage	LP-based
(not named)	[Yang et al., 2021]	objects	Single	Achiever
SVGG	[Castanet et al., 2023]	states	Coverage	Particle-based

- Could be updated with more recent works
- See the references below

Any question?



Send mail to: Olivier.Sigaud@upmc.fr

Akakzia, A., Colas, C., Oudeyer, P.-Y., Chetouani, M., and Sigaud, O. (2021).
Grounding language to autonomously-acquired skills via goal generation.
In *ICLR 2021*.

Castanet, N., Lamprier, S., and Sigaud, O. (2022).
Stein variational goal generation for reinforcement learning in hard exploration problems.
International Conference in Machine Learning (ICML).
arXiv preprint arXiv:2206.06719.

Castanet, N., Lamprier, S., and Sigaud, O. (2023).
Stein variational goal generation for reinforcement learning in hard exploration problems.
In *Proceedings of the 40th International Conference on Machine Learning*.

Colas, C., Fournier, P., Sigaud, O., and Oudeyer, P.-Y. (2018).
CURIOUS: intrinsically motivated multi-task multi-goal reinforcement learning.
In *ICML*.

Fang, M., Zhou, C., Shi, B., Gong, B., Xu, J., and Zhang, T. (2019a).
DHER: Hindsight experience replay for dynamic goals.
In *International Conference on Learning Representations*.

Fang, M., Zhou, T., Du, Y., Han, L., and Zhang, Z. (2019b).
Curriculum-guided hindsight experience replay.
Advances in neural information processing systems, 32.

Florensa, C., Held, D., Geng, X., and Abbeel, P. (2018).
Automatic goal generation for reinforcement learning agents.
In *International conference on machine learning*, pages 1515–1528. PMLR.

Ghosh, D., Gupta, A., Reddy, A., Fu, J., Devin, C., Eysenbach, B., and Levine, S. (2019).
Learning to reach goals via iterated supervised learning.
arXiv preprint arXiv:1912.06088.

Kaelbling, L. P. (1993).

Learning to achieve goals.
In *IJCAI*, pages 1094–1099. Citeseer.

Liu, H., Trott, A., Socher, R., and Xiong, C. (2019).
Competitive experience replay.
arXiv preprint arXiv:1902.00528.

Liu, Q. and Wang, D. (2016).
Stein variational gradient descent: A general purpose Bayesian inference algorithm.
arXiv preprint arXiv:1608.04471.

Matiisen, T., Oliver, A., Cohen, T., and Schulman, J. (2019).
Teacher–student curriculum learning.
IEEE transactions on neural networks and learning systems, 31(9):3732–3740.

Pitis, S., Chan, H., Zhao, S., Stadie, B., and Ba, J. (2020).
Maximum entropy gain exploration for long horizon multi-goal reinforcement learning.
In III, H. D. and Singh, A., editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 7750–7761. PMLR.

Pong, V. H., Dalal, M., Lin, S., Nair, A., Bahl, S., and Levine, S. (2019).
Skew-fit: State-covering self-supervised reinforcement learning.
arXiv preprint arXiv:1903.03698.

Racaniere, S., Lampinen, A., Santoro, A., Reichert, D., Firoiu, V., and Lillicrap, T. (2019).
Automated curriculum generation through setter-solver interactions.
In *International Conference on Learning Representations*.

Ren, Z., Dong, K., Zhou, Y., Liu, Q., and Peng, J. (2019).
Exploration via hindsight goal generation.
Advances in Neural Information Processing Systems, 32.

Schaul, T., Horgan, D., Gregor, K., and Silver, D. (2015).
Universal value function approximators.

In *International Conference on Machine Learning*, pages 1312–1320. PMLR.

Yang, D., Zhang, H., Lan, X., and Ding, J. (2021).
Density-based curriculum for multi-goal reinforcement learning with sparse rewards.
CoRR, abs/2109.08903.