Max Planck Institute for Intelligent Systems – Autonomous Learning Group

MAX-PLANCK-GESELLSCHAFT

imprs-is

MARIE CURIE ACTIONS

Control What You Can: Intrinsically Motivated Task-Planning Agent

Sebastian Blaes, Marin Vlastelica Pogančić, Jia-Jie Zhu, Georg Martius

Abstract

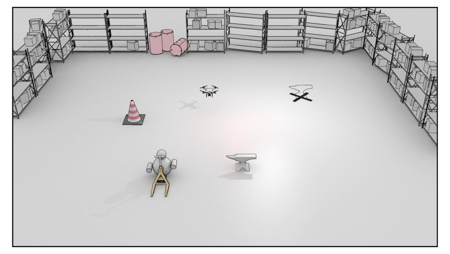
We present an **intrinsically motivated agent** that **learns how to control the environment** in the fastest possible manner by optimizing learning progress.
It learns what can be controlled, how to allocate time and attention and the relations between objects using surprise based motivation. The effectiveness of our method is demonstrated in a synthetic as well as a robotic manipulation environment yielding considerably improved performance and smaller sample complexity.
In a nutshell, **our work combines several task-level planning agent structures** (backtracking search on task graph, probabilistic road-maps, allocation of search efforts) **with intrinsic motivation to achieve learning from scratch**.

Contributions

Autonomously learning to solve challenging control problems

- using **intrinsic motivations** (IM):
  - **maximizing controllability**
  - **learning progress**
  - **surprise**
- using several **task-level planning ideas**:
  - sub-task graph
  - backtracking search on task graph
  - goal regression
  - allocation of search efforts

Result: Learning from scratch to control environment and to acquire skills



Setup and Environments

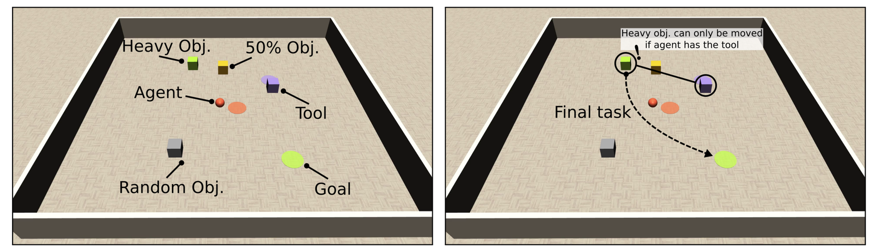
Purpose/goal of agent: Gain control over coordinates of observation space.
Controllability:: Reaching arbitrary points / goal states

Basic tool-use/object manipulation environment

Observations: Object/agent positions.

Self-posed Tasks: Manipulating coordinates of observations to a goal.

Example: Change position of tool to a goal position



Objects:
Tool can be picked up immediately,
50% object can be picked up only in 50% of the trials
Heavy object can only picked up when in possession of the tool
Random object can not be manipulated by the agent, moves randomly

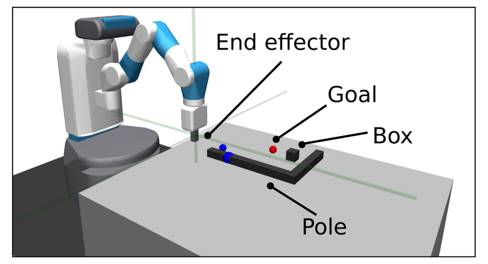
Robotic object manipulation environment

Observations: Position of gripper, hook and box.

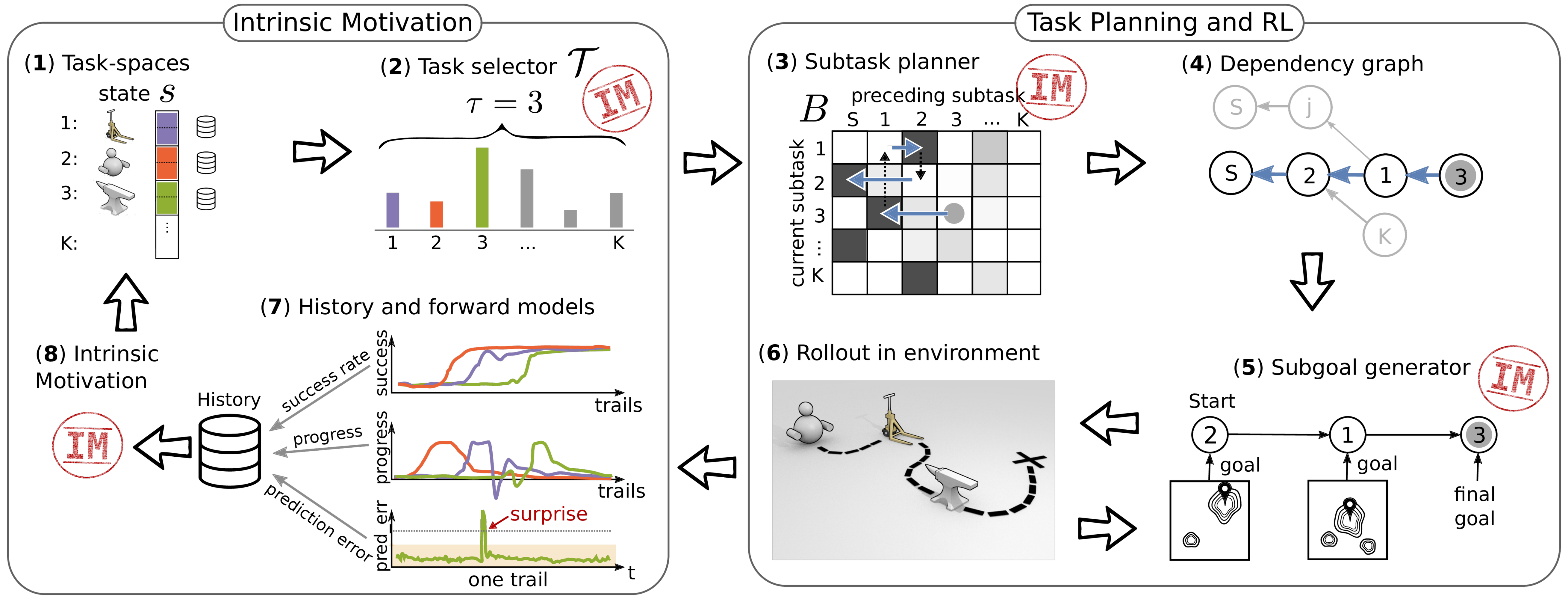
Self-posed Tasks: Manipulating coordinates of observations to a goal

Example: Change position of box to a goal position

Box Can only be moved with the help of the hook



Intrinsic Motivation / Task Planning and RL



Math

(2) Task selector

$Q_i(t) = Q_i(t-1) + \alpha \cdot (r_i(t) - Q_i(t-1))$, for all tasks i

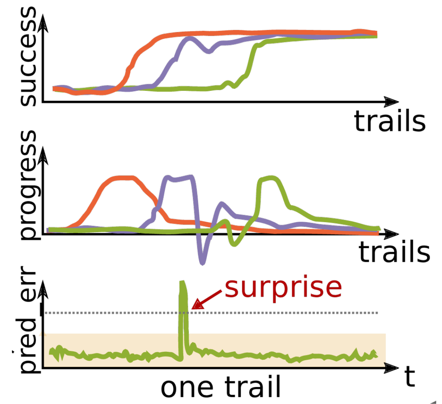
with $r_i = |\rho_i| + \beta \max_t(\text{surprise}_i(t))$

(3) Task Planner

$B_{i,j} = Q_{i,j} / \sum_k Q_{i,k}$

with $Q_{i,j} = \langle 1 - T_{i,j}/T^{\text{max}} + \beta \max_t(\text{surprise}_i(t)) \rangle$

$T_{i,j}$ is the runtime for solving task i by doing task j before, T^{max} is the maximum number of steps



(5) Sub-goal generator

$\mathcal{L}_{i,j} = \min_\omega \sum_{k=1}^n \|G_{i,j}(\omega, s_k) - r_{i,j}(s_k)\|^2$

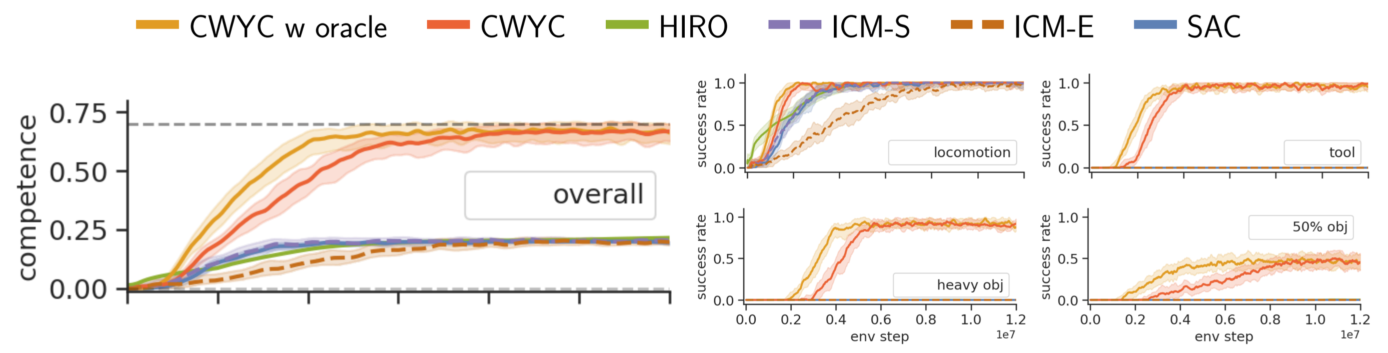
with $G_{i,j}(s) = \exp\left(-\gamma \sum_{k=1}^n \sum_{l=k+1}^n \|\omega_{kl}^1 s_k + \omega_{kl}^2 s_l + \omega_{kl}^3\|^2\right)$

and $r_{i,j}(s_t) = \min\left(1, \text{succ}_i \cdot \Gamma_{i,j}(s_t) + \text{surprise}_i(t)\right)$

$\Gamma_{i,j}(s)$ is 1 if the agent decides to switch from task j to i in state s and zero otherwise

Results

Basic tool-use/object manipulation environment



CWYC w oracle: Upper baseline for our method with hand-crafted (optimal) task planner and sub-goal generators

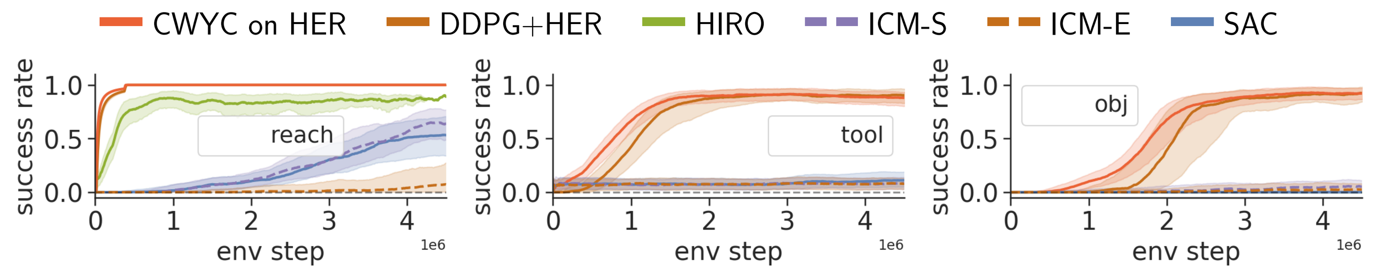
HIRO: Hierarchical RL baseline

ICM-(S/E): Intrinsically motivated RL agent baselines

SAC: Vanilla soft actor-critic algorithm

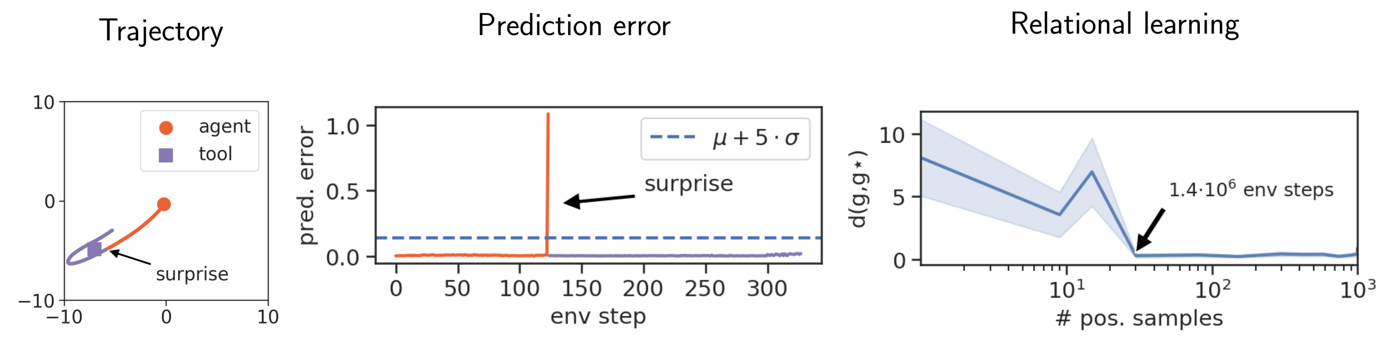
See the related work & references box for additional information regarding the hierarchical and intrinsically motivated baselines.

Robotic object manipulation environment



DDPG+HER: RL baseline without intrinsic motivation or hierarchical structure, the other baselines are the same as in the upper plots.

Surprise / Intrinsic Motivation



(left) **Trajectory** of the agent picking up the tool. (middle) Picking up the tool causes large **prediction error (surprise)** in agent's internal forward model. (right) **Task and object relations** can already be inferred from a handful of surprising events and successful task transitions (cf. learned task graph and learned object relations)

Analyzing the Components

All components are learned from data!

Related Work & References

	Intrinsic motivation	Computational methods
CWYC Ours	**learning progress + surprise**	**task-level planning, relational attention**
h-DQN [4]	reaching subgoals	HRL, DQN
IMGEP [3]	learning progress	memory-based
CURIOUS [1]	learning progress	DDPG, HER, E-UVFA
SAC-X [6]	auxiliary task	HRL, (DDPG-like) PI
Relational RL [8]	-	relation net, IMPALA
ICM [5]	prediction error	A3C, ICM
Goal GAN [2]	adversarial goal	GAN, TRPO
Asymmetric self-play [7]	self-play	Alice/Bob, TRPO, REINFORCE

References

[1] Cédric Colas, Pierre-Yves Oudeyer, Olivier Sigaud, Pierre Fournier, and Mohamed Chetouani. CURIOUS: intrinsically motivated modular multi-goal reinforcement learning. In *International Conference on Machine Learning (ICML'19)*, pages 1331–1340, 2019.

[2] Carlos Florensa, David Held, Xinyang Geng, and Pieter Abbeel. Automatic goal generation for reinforcement learning agents. In *International Conference on Machine Learning (IMCL'18)*, pages 1514–1523, 2018.

[3] Sébastien Forestier, Yoan Mollard, and Pierre-Yves Oudeyer. Intrinsically motivated goal exploration processes with automatic curriculum learning. *arXiv:1708.02190*, 2017.

[4] Tejas D Kulkarni, Karthik Narasimhan, Ardavan Saeedi, and Josh Tenenbaum. Hierarchical deep reinforcement learning: Integrating temporal abstraction and intrinsic motivation. In *Advances in Neural Information Processing Systems (NIPS'16)*, pages 3675–3683, 2016.

[5] Deepak Pathak, Pulkit Agrawal, Alexei A Efros, and Trevor Darrell. Curiosity-driven exploration by self-supervised prediction. In *International Conference on Machine Learning (ICML'17)*, volume 2017, 2017.

[6] Martin Riedmiller, Roland Hafner, Thomas Lampe, Michael Neunert, Jonas Degrave, Tom van de Wiele, Vlad Mnih, Nicolas Heess, and Jost Tobias Springenberg. Learning by playing solving sparse reward tasks from scratch. In *International Conference on Machine Learning (ICML'18)*, volume 80, pages 4344–4353, 2018.

[7] Sainbayar Sukhbaatar, Zeming Lin, Ilya Kostrikov, Gabriel Synnaeve, Arthur Szlam, and Rob Fergus. Intrinsic motivation and automatic curricula via asymmetric self-play. In *International Conference on Learning Representations, ICLR 2018, Conference Track*, 2018.

[8] Vinicius Zambaldi, David Raposo, Adam Santoro, Victor Bapst, Yujia Li, Igor Babuschkin, Karl Tuyls, David Reichert, Timothy Lillicrap, Edward Lockhart, et al. Relational deep reinforcement learning. *arXiv:1806.01830*, 2018.

{sebastian.blaes,marin.vlastelica,jzhu,georg.martius}@tue.mpg.de

Paper Code al.is.mpg.de