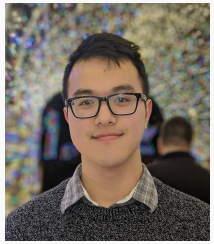
Exploration in Approximate Hyper-State Space for Meta-RL

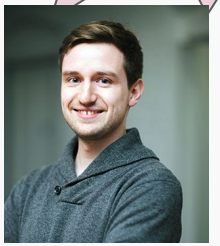
A way to solve the meta-exploration problem!

Luisa Zintgraf



Leo Feng

Cong Lu



Maximilian Igl

Kristian Hartikainen

Katja Hofmann

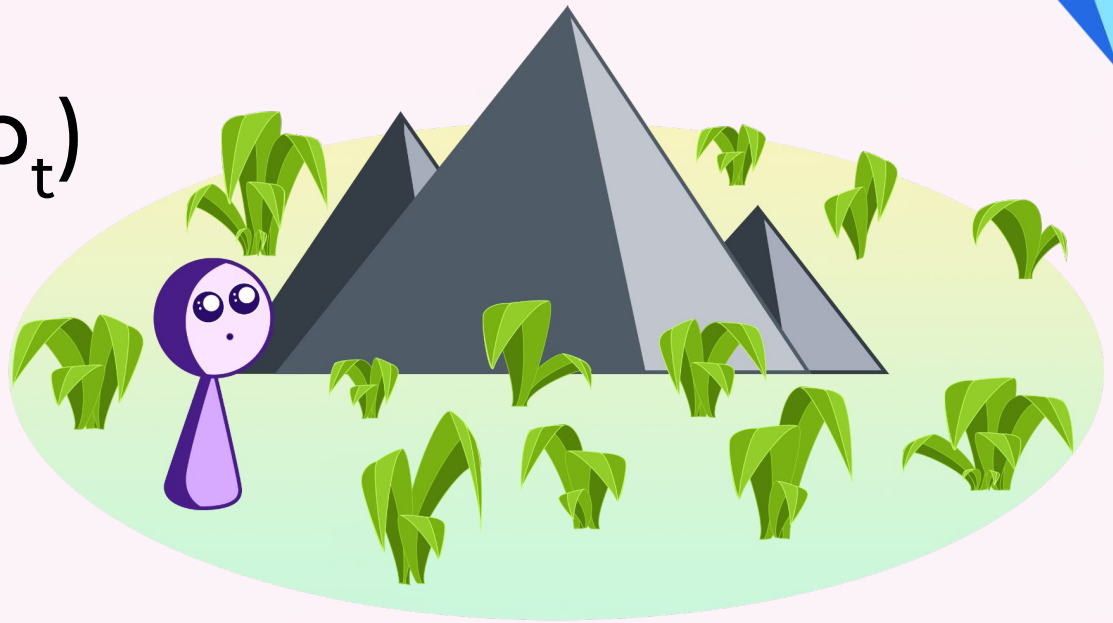
Shimon Whiteson

Meta-RL allows an agent to
improve future *learning* performance!

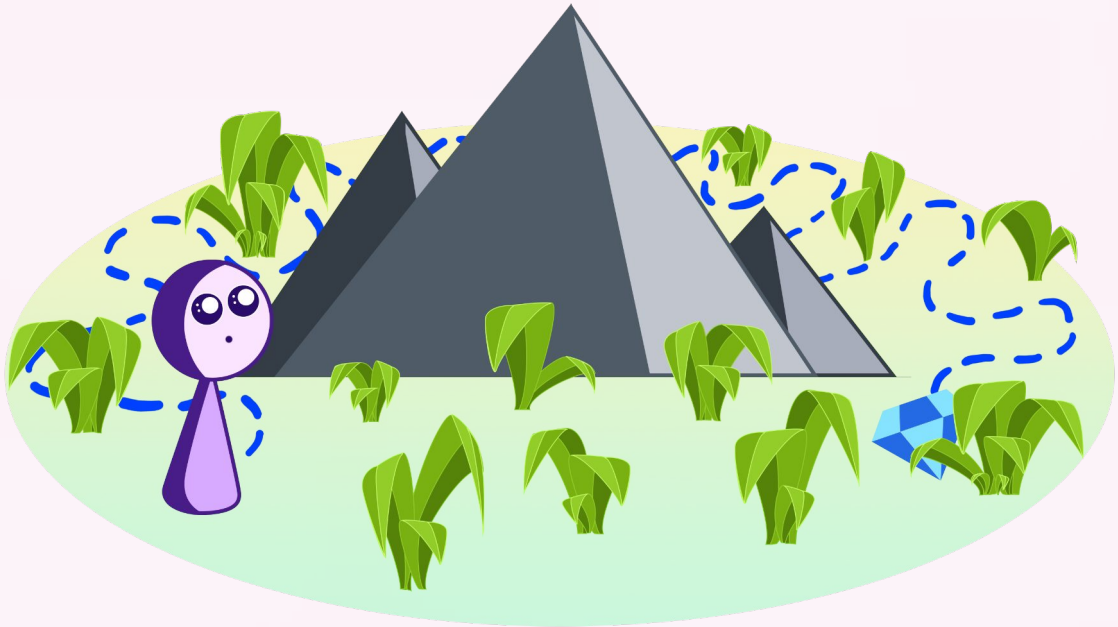
(See definition of Hospedales et al. 2020)

$b_t(x_{💎}, y_{💎} \mid \text{history}_t)$

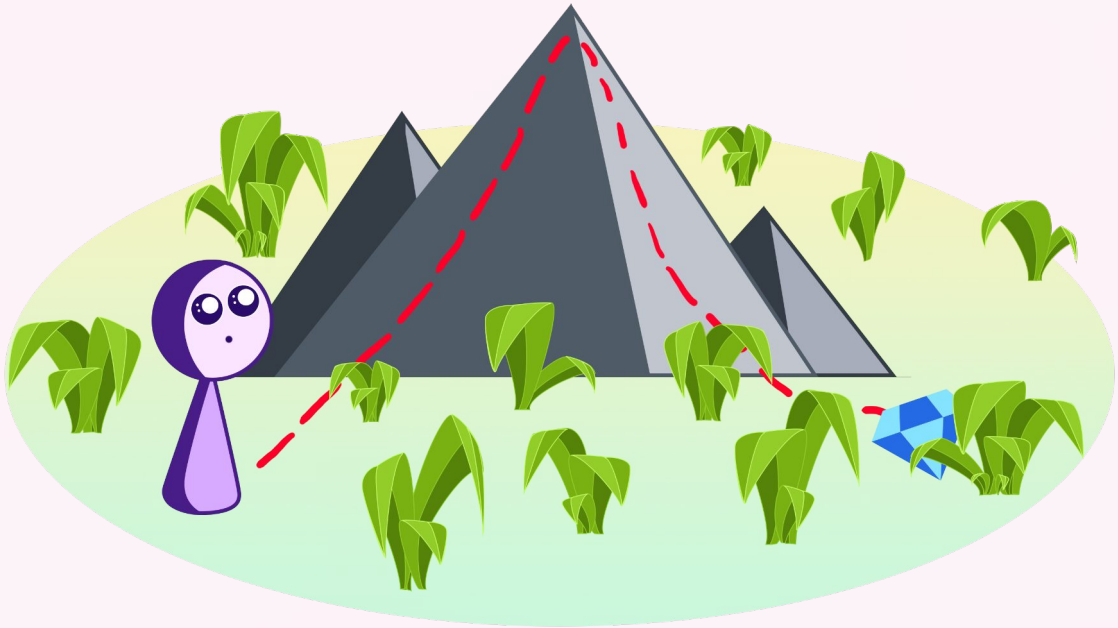
$\pi(a_t \mid s_t, b_t)$



A costly strategy

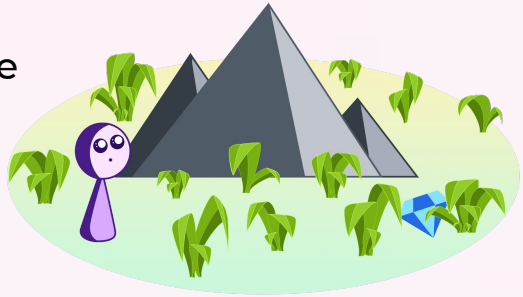


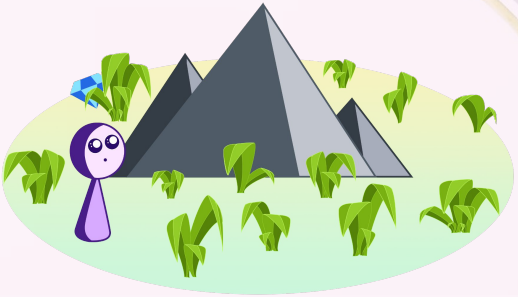
The best strategy

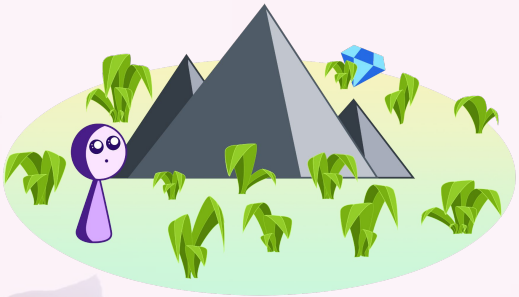


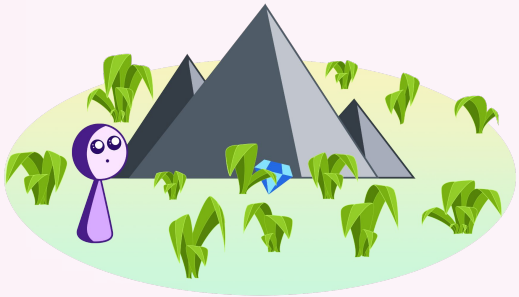
Meta-Learning

1. How to maintain a belief
2. How to use the belief to take actions with highest expected online return





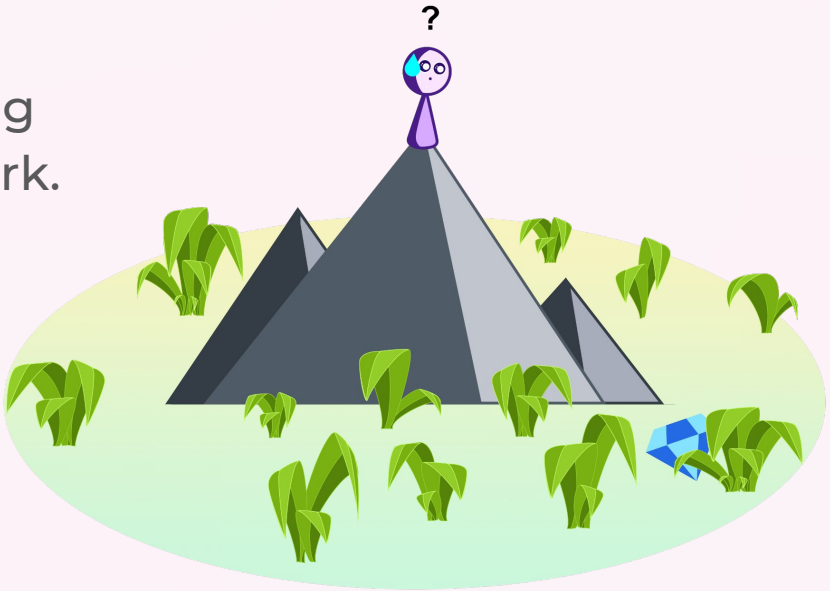




...

Hold up!

... existing meta-learning methods might not work.

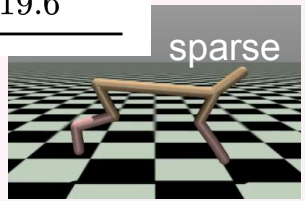


Agent doesn't know how to interpret the signal!

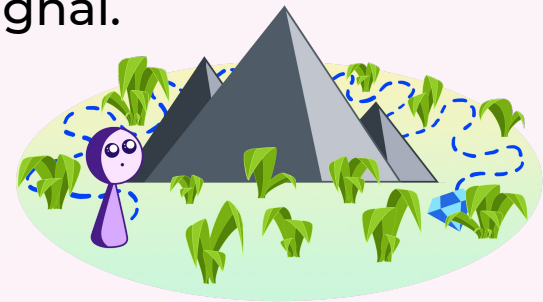
Negative rewards for climbing the mountain!

The Meta-Exploration Problem!

Method	Avg Return
VariBAD	−1.1
E-MAML	−0.4
ProMP	−0.4
Humplik et al.	−0.1
RL^2	−0.7
PEARL	−0.1
HyperX (ours)	819.6



If the agent doesn't explore enough *during* meta-training, there's not enough learning signal. Meta-learning performs poorly or fails.



HyperX: Exploration in Hyper-State Space

$$r = r_{env} + r_{error} + r_{hyper}$$

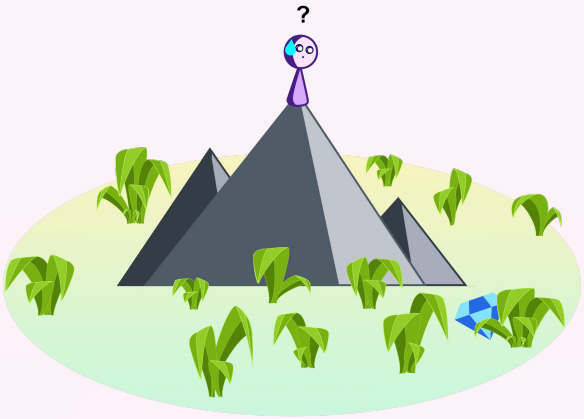
Extrinsic reward

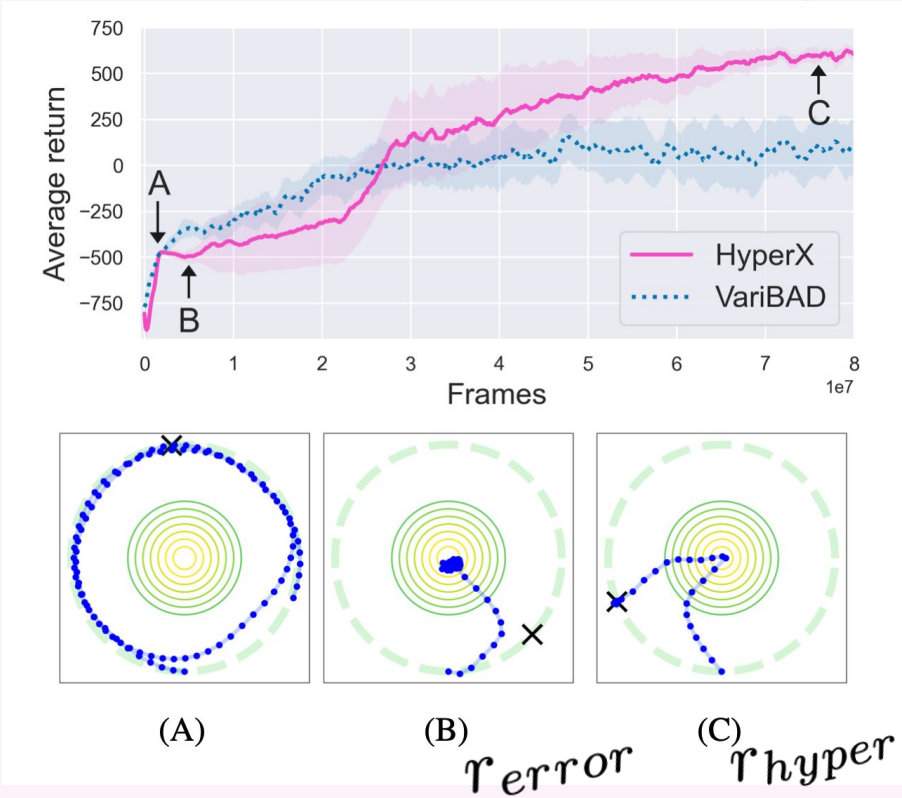
Bonus for states with wrong belief inference

→ encourages agent to **collect data to learn how to do belief inference**

Novelty bonus for hyper-states (state-belief-pairs)

→ encourages agent to **try different task-exploration strategies**


?


750
500
250
0
−250
−500
−750
Average return
A
B
C
HyperX
VariBAD
0
1
2
3
4
5
6
7
8
1e7
Frames
(A)
(B)
(C)
r_{error}
r_{hyper}

Sparse Meta-World

Sparse MuJoCo Ant-Goal, Cheetah-Dir

2D Navigation

Multi-Stage GridWorld

More experiments in the paper!

Ablations

Bonuses for different meta-learners

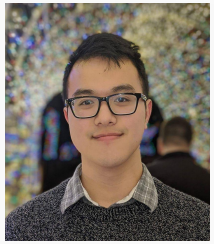
Using other bonuses

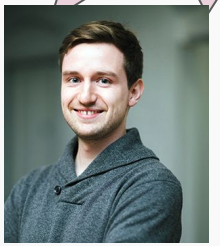
Yes, you need both bonuses!

Visualisations

Exploration in Approximate Hyper-State Space for Meta-RL

A way to solve the meta-exploration problem!




Luisa Zintgraf

Leo Feng

Cong Lu

Maximilian Igl

Kristian Hartikainen

Katja Hofmann

Shimon Whiteson

luisa.zintgraf@cs.ox.ac.uk @luisa_zintgraf