Policy composition in reinforcement learning via multi-objective policy optimization

Shruti Mishra[1,2] Ankit Anand[2] Jordan Hoffmann[2] Nicolas Heess[2]
Martin Riedmiller[2] Abbas Abdolmaleki[2] Doina Precup[2,3,4]

[1]Sony AI [2]Google DeepMind [3]McGill University [4]Mila

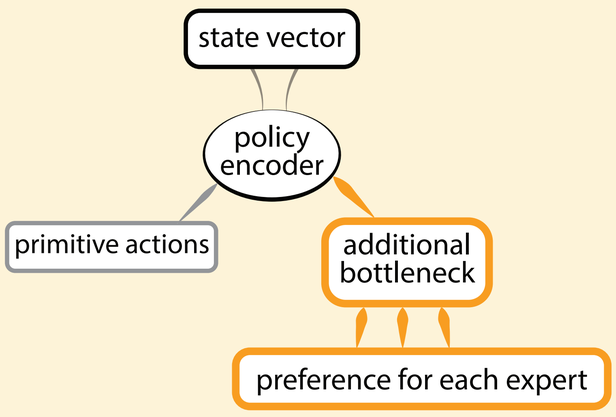
A multi-objective approach to leverage experts for locomotion

Motivation

- Pre-existing expertise can be useful for agents in locomotion domains.
- Improvements to learning efficiency can enable more general-purpose agents in environments.

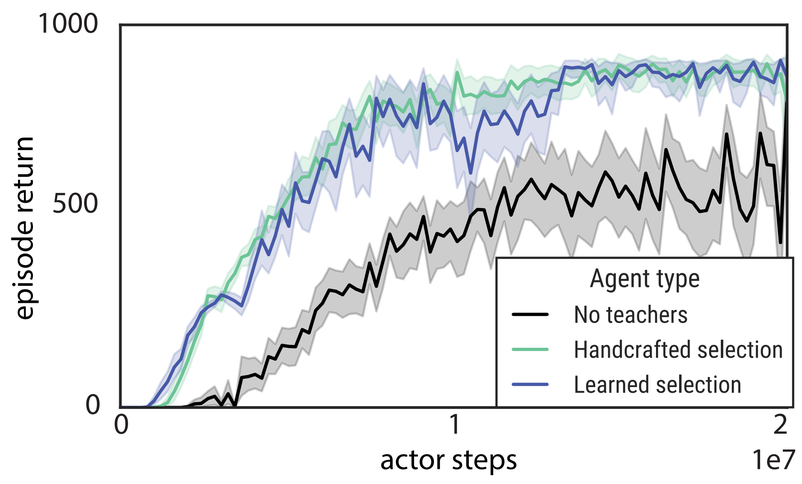
Method

1. Reinforcement learning agents in locomotion domains have access to pre-existing, sub-optimal policies.
2. Adherence to teacher policies is framed as additional objectives in a multi-objective RL setup.
3. Agents control the adherence to pre-existing policies using the task objective.

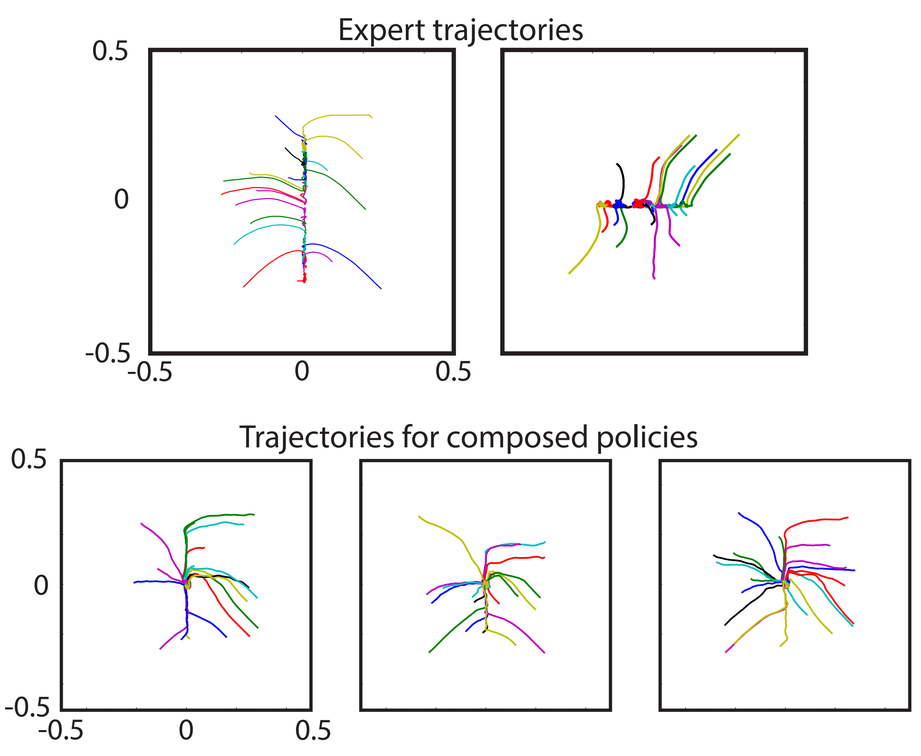


Outcomes

- Agents successfully choose how to leverage pre-existing policies.
- This can overcome limitations of reward shaping.



- Policies are shaped by experts.



contact: shruti.mishra10@alumni.imperial.ac.uk