Leveraging Dynamics Model for Single-shot Task Generalization in Reinforcement Learning

Supervisor(s):

Francesco Belardinelli (Department of Computing, francesco.belardinelli@imperial.ac.uk)

Project description:

Reinforcement learning (RL) [1] has been successfully deployed for training agents to complete tasks in diverse and challenging environments. However, the agent trained with RL typical fails a new task in the same environment, without retraining or auxiliary reward signals. A significant amount of work in RL is concerned with task transfer learning [2] and continual learning [3] – how can we leverage past experience and competence at one task to help speed up the learning process of a new task. The goal of this project is to develop an approach to task transfer learning, that facilitates either single-shot (no extra training) or few-shot (some small amount of extra training) generalization to novel tasks. Typically, the agent is first allowed some amount of (goal-free) exploration time in the environment, after which a task is given to the agent.

The student's primary objectives will include:

1) Formalizing RL tasks as Linear Temporal Logic (LTL) or similar temporal logic formula [4].
2) Develop a method for single-/few-shot task generalization using a dynamics model learned in the free exploration time.
3) Demonstrate the effectiveness of your approach empirically on at least one benchmark.

The more ambitious student may consider deep RL techniques, such as worlds models [5] or ensembles of neural networks [6]. And they may propose implement more sophisticated planning schemes (e.g., Model Predictive Control (MPC) [7]) and conduct experiments on more than one benchmark.

Timeline (tentative):

Jan 2025: literature review/preliminary experiments as well as completion of task (1).
April 2025: completion of task (2)
June 2025: completion of task (3).
July 2025: tackling any stretch goal, write up of final report.

Minimum viable thesis:

A thorough review and implementation of currently available methods. A few-shot generalization method for vanilla reachability or safety goal.

Required background & skills:

One of formal methods/logic-based languages/symbolic AI on one side, and reinforcement learning/safe AI on the other.

Representative References:

[1] Sutton, Richard S., and Andrew G. Barto. Reinforcement learning: An introduction. MIT press, 2018.
[2] León, Borja G., Murray Shanahan, and Francesco Belardinelli. "Systematic Generalisation of Temporal Tasks through Deep Reinforcement Learning."
[3] Hihn, Heinke, and Daniel A. Braun. "Hierarchically structured task-agnostic continual learning." Machine Learning 112.2 (2023): 655-686.
[4] Baier, Christel, and Joost-Pieter Katoen. Principles of model checking. MIT press, 2008.
[5] Ha, David, and Jürgen Schmidhuber. "World models." arXiv preprint arXiv:1803.10122 (2018).
[6] Janner, Michael, et al. "When to trust your model: Model-based policy optimization." Advances in neural information processing systems 32 (2019).
[7] Bharadhwaj, Homanga, Kevin Xie, and Florian Shkurti. "Model-predictive control via cross-entropy and gradient-based optimization." Learning for Dynamics and Control. PMLR, 2020.