Chair of Robotics, Artificial Intelligence and Real-time Systems
TUM School of Computation, Information and Technology
Technical University of Munich

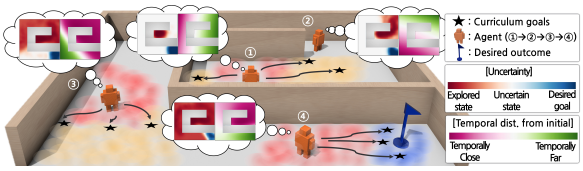
Generating curriculum via Decision Transformer in Maze and Robotics environments

Master Thesis

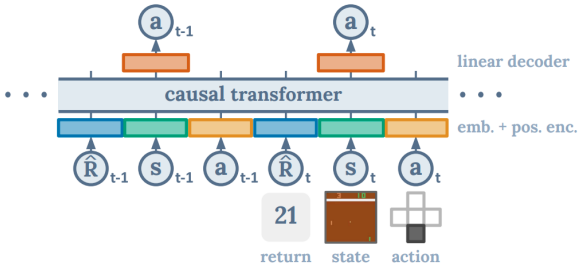
Advisor: Erdi Sayar *(erdi.sayar@tum.de)*
Supervisor: Prof. Alois Knoll *(erdi.sayar@tum.de)*

Introduction and Problem Description

Hindsight Experience Replay (HER) has been proposed to allow off-policy RL algorithms to perform effective learning in solving goal-based tasks with sparse/binary rewards, such as the manipulation of robotic arms [1]. HER takes advantage of failed trajectories by replacing desired goals with the achieved goals. However, it cannot solve the tasks If desired goal is far away from the initial states. Curriculum-based RL approaches decompose complex tasks into sequences of gradually more difficult tasks, by relying on heuristics that guide the agent to explore the environment more efficiently. For example, OUTPACE [3] generates distance-aware curriculum RL with intrinsic rewards based on the classifier by Conditional Normalized Maximum Likelihood (CNML) and Wasserstein distance, as shown in Figure 1a. Recently, Decision Transformer(DT) [2] as shown in Figure 1b has been presented for solving RL problems by conditioning it on the desired return, past states, and past actions, and it could generate optimum future actions that achieve the desired return. In this work, CNML will be replaced by Decision Transformers to generate future curriculum goals and will be evaluated in maze and robotic manipulation environments as illustrated in Figure 2.



(a) OUTPACE generates uncertainty and temporal distance-aware curriculum goals.



(b) Decision Transformer architecture

Task Description

In this thesis, your task will be learning Decision Transformer implementation on the Reinforcement Learning tasks and adapt it to generate curriculum goals. Specifically-

- You will first learn basic knowledge of reinforcement learning.
- You will reproduce the results from Decision Transformer RL paper [2]. By doing this, you will have a deep understanding of the state-of-the-art research result.
- You will adapt the Decision Transformer to generate future curriculum goals instead of generating future actions.
- Franka robotic environments will also be integrated into the code and tested with the Decision Transformer, as well as other baseline approaches.

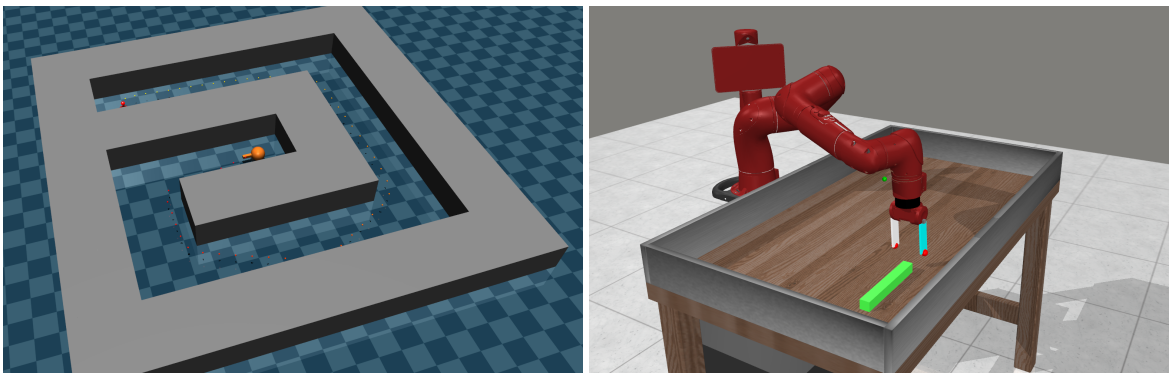


Figure 2: In maze environment examples, the big orange dot represents the agent, while the red dot represents the desired goal. The small colorful dots serve as generated curriculum points that guide the agent toward the desired goal. In the robotic environment, the objective is to transport the green object to the green dot.

Requirements

- High self-motivation;
- Experience or knowledge from related courses
- Python programming experience

References

[1] Andrychowicz, M., Wolski, F., Ray, A., Schneider, J., Fong, R., Welinder, P., McGrew, B., Tobin, J., Pieter Abbeel, O., and Zaremba, W. "Hindsight experience replay". In: *Advances in neural information processing systems* 30 (2017).

[2] Chen, L., Lu, K., Rajeswaran, A., Lee, K., Grover, A., Laskin, M., Abbeel, P., Srinivas, A., and Mordatch, I. "Decision transformer: Reinforcement learning via sequence modeling". In: *Advances in neural information processing systems* 34 (2021), pp. 15084–15097.

[3] Cho, D., Lee, S., and Kim, H. J. "Outcome-directed Reinforcement Learning by Uncertainty & Temporal Distance-Aware Curriculum Goal Generation". In: *arXiv preprint arXiv:2301.11741* (2023).