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

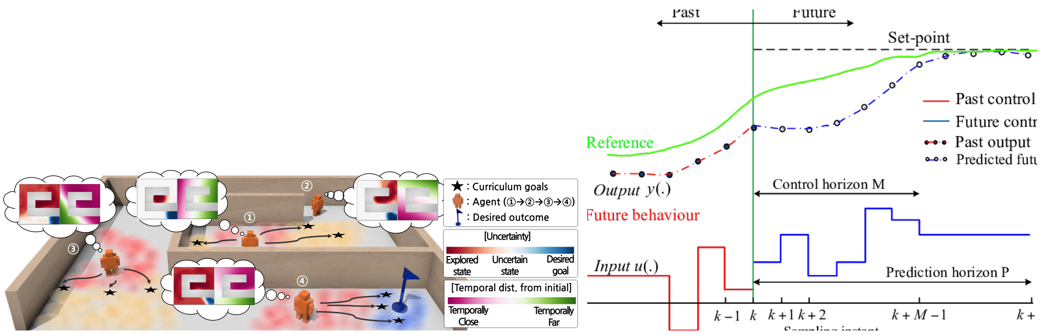
Model Predictive Controller with Model-Based RL

Master Thesis

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

Introduction and Problem Description

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 [1] 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. However, OUTPACE has limitations when applied in environments with obstacles. To address this issue, we propose an idea integrating Model Predictive Control (MPC) as an effective control approach for path planning and collision avoidance within the OUTPACE framework. MPC requires a mathematical model of the agent to predict future states over a finite time horizon. Nevertheless, it's not always feasible to establish mathematical models for complex agents. The most common way to estimate dynamics is by fitting a one-step ahead prediction model and using it to recursively propagate the predicted state distribution over long horizons. Unfortunately, this approach is known to compound even small prediction errors, making long-term predictions inaccurate. To address this challenge, we will implement a trajectory-based model [2]. This trajectory-based model takes an initial state, a future time index, and control parameters as inputs, and directly predicts the state at the future time index. Subsequently, the MPC algorithm will use the neural network-based dynamic model to generate intermediate goals to avoid obstacles. We plan to evaluate this proposed approach initially in maze environments, as illustrated in Figure 2, and later in robotic manipulation tasks.



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

(b) General concept for model predictive control

Task Description

In this thesis, your task will involve learning the Model Predictive Controller (MPC) and benchmark code.[1] To achieve this, you will:

- Initially, reproduce the results of the benchmark code.
- Next, integrate MPC into the benchmark code, using a mathematical model of a simple point-based agent.

[1] https://github.com/jayLEE0301/outpace_official

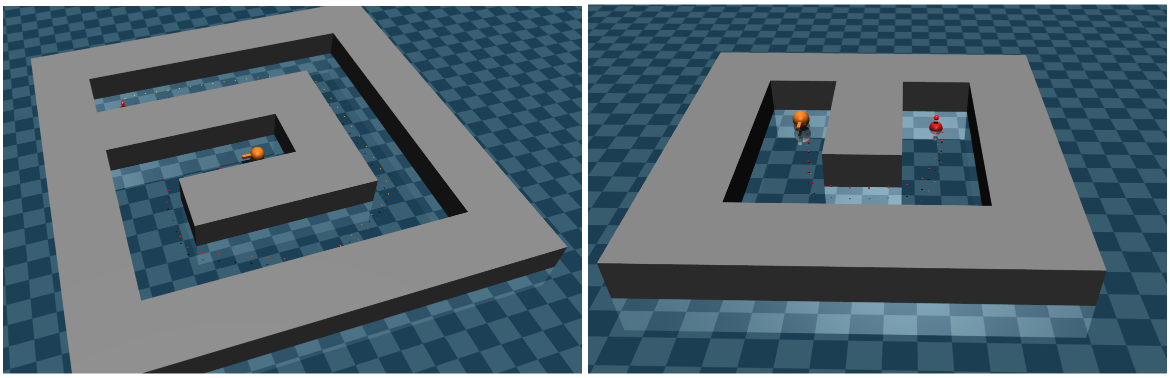


Figure 2: In maze environment examples, the big orange dot represents the agent, while the red dot represents the desired goal. The small colorful dots serve as generated curriculum points that guide the agent toward the desired goal.

- Incorporate trajectory-based neural network models to learn the dynamic model of the agent.
- Finally, instead of relying on mathematical equations as the dynamic model of the agent, MPC will be used with the neural network as a mathematical model."

Requirements

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

References

[1] 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).

[2] Lambert, N., Wilcox, A., Zhang, H., Pister, K. S., and Calandra, R. "Learning accurate long-term dynamics for model-based reinforcement learning". In: *2021 60th IEEE Conference on Decision and Control (CDC)*. IEEE. 2021, pp. 2880–2887.