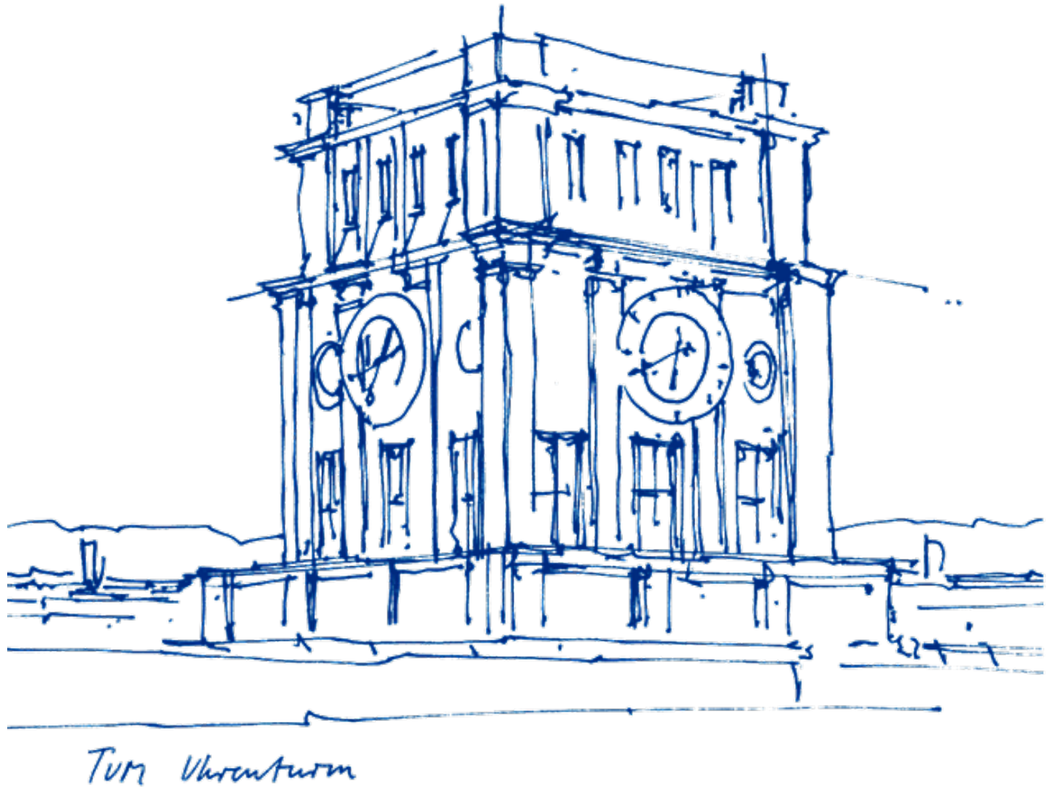
Factory Manipulation with Cooperative Multi-agent Reinforcement Learning

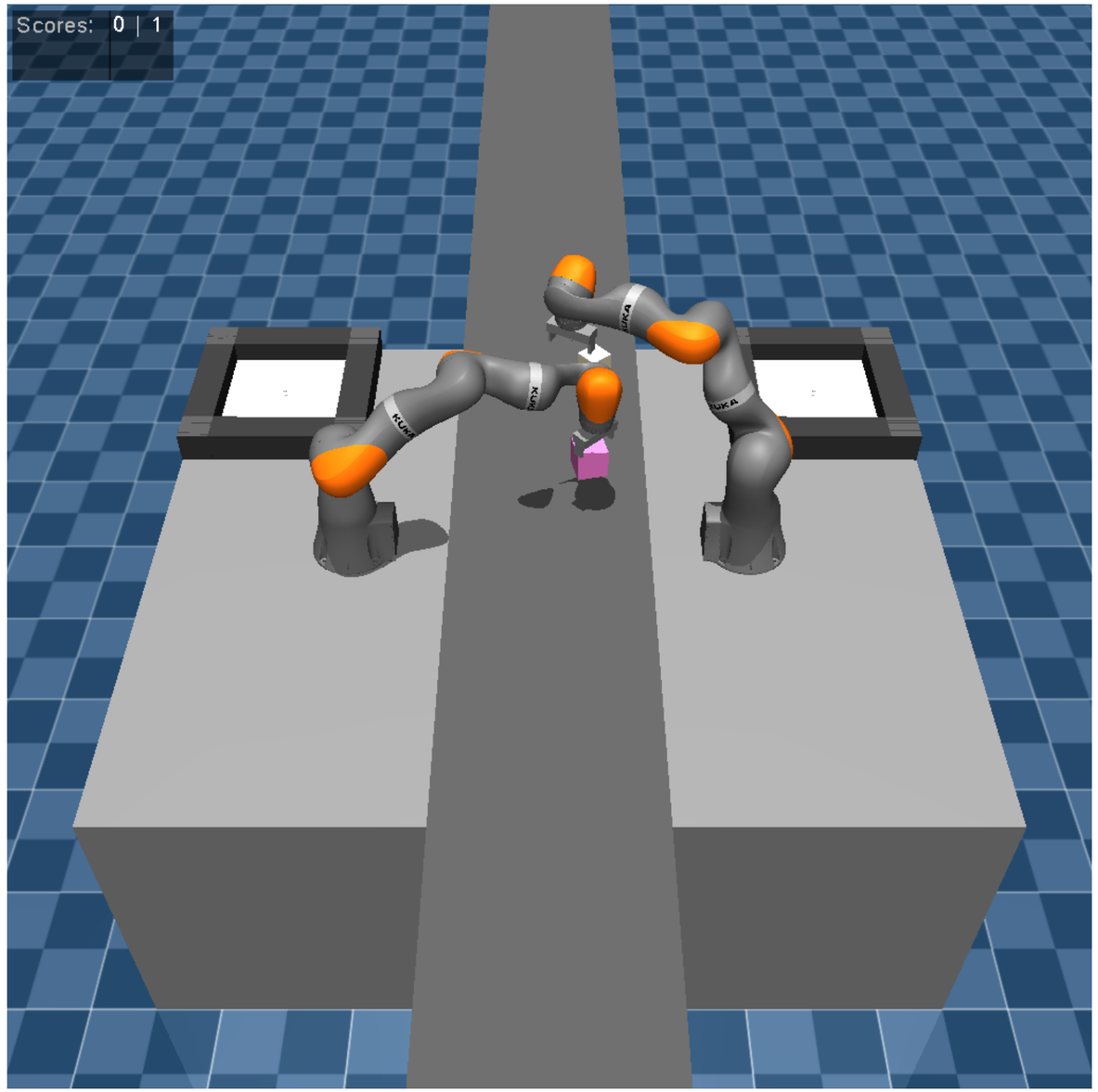


28. May 2024

Kirschstein, Köck

The Setting

- 2 or more robot arms
- basket in reach for each arm
- conveyor belt (with increasing speed)
- cubes transported on the conveyor belt
  $\rightarrow$ shall be put into the basket
- arms must not hit each other
- arms must not hit the environment
- 8 degrees of freedom
- Arms use RL to learn movements



The Reward Function

Which reward should we use?

The Reward Function

Which reward should we use?

1st intention:

Reward of 1, if a block is thrown in the basket

The Reward Function

Which reward should we use?

1st intention:

Reward of 1, if a block is thrown in the basket

Sparse reward

- $\rightarrow$ **Almost no learning possible**

HierkönnteihrFotostehen

The Reward Function

Which reward should we use?

2nd intention:

Function **increases monotonously with *progress to target***

The Reward Function

Which reward should we use?

2nd intention:

Function **increases monotonously with *progress to target***

Parameters to include:

- N: Number of cubes in basket
- d_c: Distance to closest cube
- d_r: Distance to closest robot arm
- d_b: Distance to basket
- g: Gripper state
- $a \in [-1, 1]^{\mathtt{DOF}}$: action vector

The Reward Function

Which reward should we use?

2nd intention:

Function **increases monotonously with *progress to target***

Desirable Incentives:

- Reward vicinity to closest cube
- Punish distance to other robot arms
- Reward grasping while very close to cube
- Reward vicinity to basket with grasped cube
- Reward relaxing grasp over basket
- Punish hectic motion

The Reward Function

Which reward should we use?

2nd intention:

Function **increases monotonously with *progress to target***

Possible reward function with Incentives $I_0, \ldots, I_6$:

$$r = \sum_{i=0}^{6} \varepsilon_i I_i$$

with ε_i scaling factor as hyperparameter

Start with equal ε_i, goal: $\varepsilon_0 >> \varepsilon_i$, ideally most $\varepsilon_i = 0$