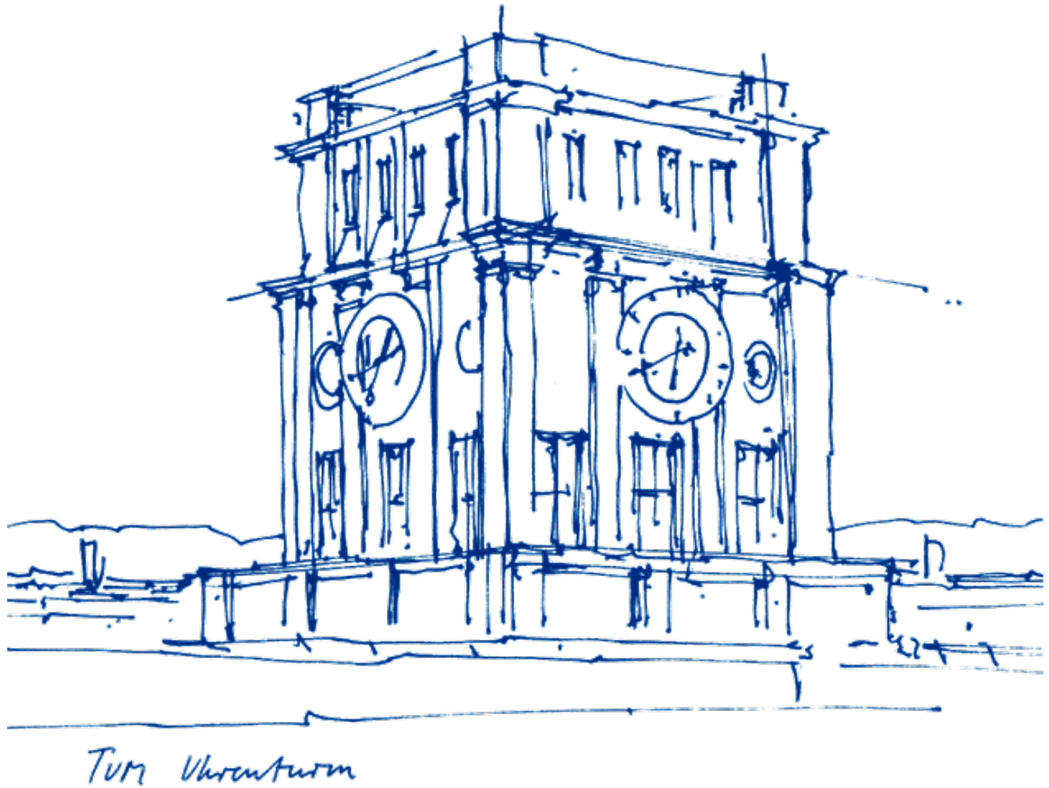
Factory Manipulation with Cooperative Multi-agent Reinforcement Learning

28. May 2024

Kirschstein, Köck



The Reward Function

Technische Universität München

Which reward should we use?

The Reward Function

Technische Universität München

Which reward should we use?

1st intention:

Reward of 1, if a block is thrown in the basket, 0 in all other cases

The Reward Function

Technische Universität München

Which reward should we use?

1st intention:

Reward of 1, if a block is thrown in the basket, 0 in all other cases

But: **Almost no learning possible**

The Reward Function

Technische Universität München

Which reward should we use?

2nd intention:

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

The Reward Function

Technische Universität München

Which reward should we use?

2nd intention:

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

Parameters to include:

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

The Reward Function

Technische Universität München

Which reward should we use?

2nd intention:

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

Parameters to include:

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

Desirable Incentives:

Possible reward function:

The Reward Function

Technische Universität München

Which reward should we use?

2nd intention:

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

Parameters to include:

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

Desirable Incentives:

- Reward vicinity to closest cube

Possible reward function:

$+ \quad 0.40 \cdot d_c$

The Reward Function

Technische Universität München

Which reward should we use?

2nd intention:

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

Parameters to include:

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

Desirable Incentives:

- Reward vicinity to closest cube
- Punish distance to other robot arms

Possible reward function:

$+ \quad 0.40 \cdot d_c$

$- \quad 0.30 \cdot d_r$

The Reward Function

Technische Universität München

Which reward should we use?

2nd intention:

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

Parameters to include:

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

Desirable Incentives:

- Reward vicinity to closest cube
- Punish distance to other robot arms
- Reward grasping while very close to cube

Possible reward function:

$$+ \quad 0.40 \cdot d_c$$
$$- \quad 0.30 \cdot d_r$$
$$+ \quad 0.20 \cdot 1_{d_c=0} \cdot (1 - g)$$

The Reward Function

Technische Universität München

Which reward should we use?

2nd intention:

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

Parameters to include:

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

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

Possible reward function:

$$+ \; 0.40 \cdot d_c$$
$$- \; 0.30 \cdot d_r$$
$$+ \; 0.20 \cdot 1_{d_c=0} \cdot (1 - g)$$
$$+ \; 0.20 \cdot 1_{d_c=0} \cdot (1 - g) \cdot d_b$$

The Reward Function

Technische Universität München

Which reward should we use?

2nd intention:

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

Parameters to include:

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

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 position

Possible reward function:

$$+ \; 0.40 \cdot d_c$$
$$- \; 0.30 \cdot d_r$$
$$+ \; 0.20 \cdot 1_{d_c=0} \cdot (1 - g)$$
$$+ \; 0.20 \cdot 1_{d_c=0} \cdot (1 - g) \cdot d_b$$
$$+ \; 0.30 \cdot 1_{d_b=0} \cdot g$$

The Reward Function

Technische Universität München

Which reward should we use?

2nd intention:

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

Parameters to include:

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

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 position
- Punish hectic motion

Possible reward function:

$$+\ 0.40 \cdot d_c$$
$$-\ 0.30 \cdot d_r$$
$$+\ 0.20 \cdot 1_{d_c=0} \cdot (1 - g)$$
$$+\ 0.20 \cdot 1_{d_c=0} \cdot (1 - g) \cdot d_b$$
$$+\ 0.30 \cdot 1_{d_b=0} \cdot g$$
$$-\ 0.10 \cdot \|a\|_2$$

The Reward Function

Technische Universität München

Which reward should we use?

2nd intention:

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

Parameters to include:

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

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 position
- Punish hectic motion

Possible reward function:

$$+\ 0.40 \cdot f_1(d_c)$$
$$-\ 0.30 \cdot f_2(d_r)$$
$$+\ 0.20 \cdot 1_{d_c=0} \cdot f_3(1-g)$$
$$+\ 0.20 \cdot 1_{d_c=0} \cdot f_4(1-g) \cdot f_5(d_b)$$
$$+\ 0.30 \cdot 1_{d_b=0} \cdot f_6(g)$$
$$-\ 0.10 \cdot f_7(\|a\|_2)$$

The Reward Function

Technische Universität München

Which reward should we use?

2nd intention:

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

Parameters to include:

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

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 position
- Punish hectic motion

Possible reward function:

$$+\ 0.40 \cdot f_1(d_c)$$
$$-\ 0.30 \cdot f_2(d_r)$$
$$+\ 0.20 \cdot 1_{d_c<\varepsilon_1} \cdot f_3(1-g)$$
$$+\ 0.20 \cdot 1_{d_c<\varepsilon_2} \cdot f_4(1-g) \cdot f_5(d_b)$$
$$+\ 0.30 \cdot 1_{d_b<\varepsilon_3} \cdot f_6(g)$$
$$-\ 0.10 \cdot f_7(\|a\|_2)$$