AI SAFETY

DOMAIN SHIFTS IN RL: IDENTIFYING DISTURBANCES IN ENVIRONMENTS

TOM HAIDER , FELIPPE SCHMOELLER ROZA , DIRK EILERS , KARSTEN ROSCHER AND STEPHAN GÜNNEMAN*

Fraunhofer IKS, TU Munich*

August 19th, 2021

Fraunhofer IKS

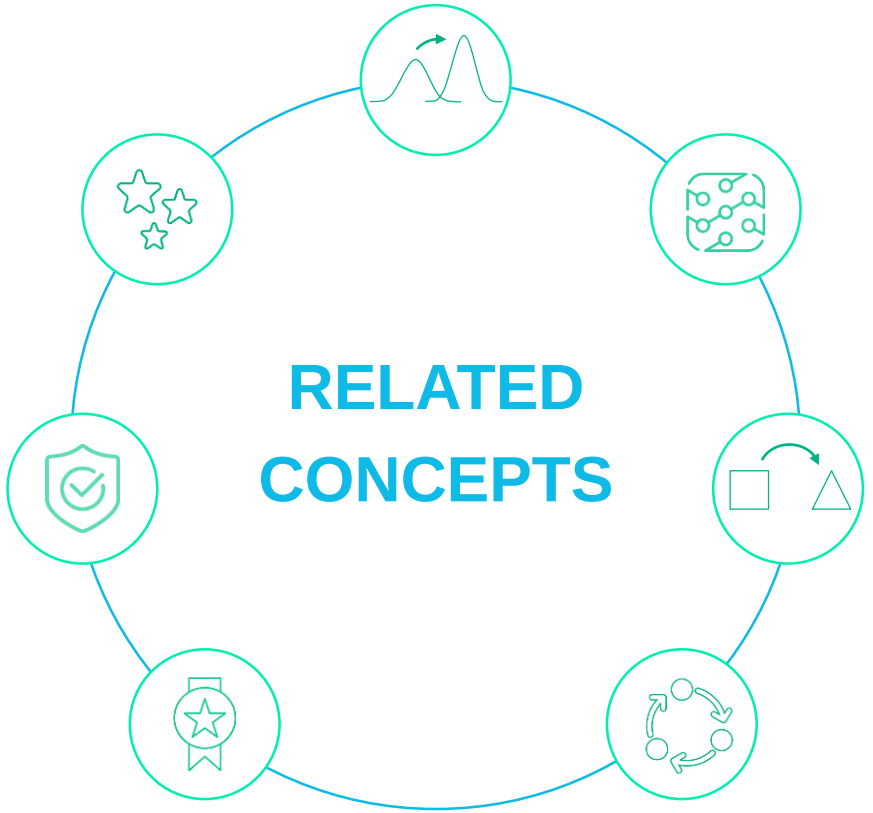
OUT OF DISTRIBUTION?

Images adopted from (Hansen, Wang; "*Generalization in Reinforcement Learning by Soft Data Augmentation*"; 2021)

Distributional Shift

Novelty

Out of Distribution



Robustness

Transfer Learning

Intrinsic Motivation

Domain Adaptation

Fraunhofer IKS

FORMALIZING DOMAIN DIFFERENCES AS ASPECTS OF THE MDP

- RL: ‘Learning from interaction to achieve a goal’ (Sutton and Barto, 1998)

- This problem can be formalized in an MDP: $\mathcal{M} := (S, A, R, P, \mu_0)$, where

  S: the set of environment states

  A: the set of available actions

  R: the reward function $R: S \times A \times S \to \mathbb{R}$

  P: the (stochastic) transition function (dynamics) P: $S \times A \times S \to [0,1]$

  μ_0:the initial state Distribution

FORMALIZING DOMAIN DIFFERENCES AS ASPECTS OF THE MDP

Two tasks, $\mathcal{M}_A$ and $\mathcal{M}_B$ can be different only in the aspects of the MDP

S (state-space): Enlargements, reductions or changes to the set of possible states

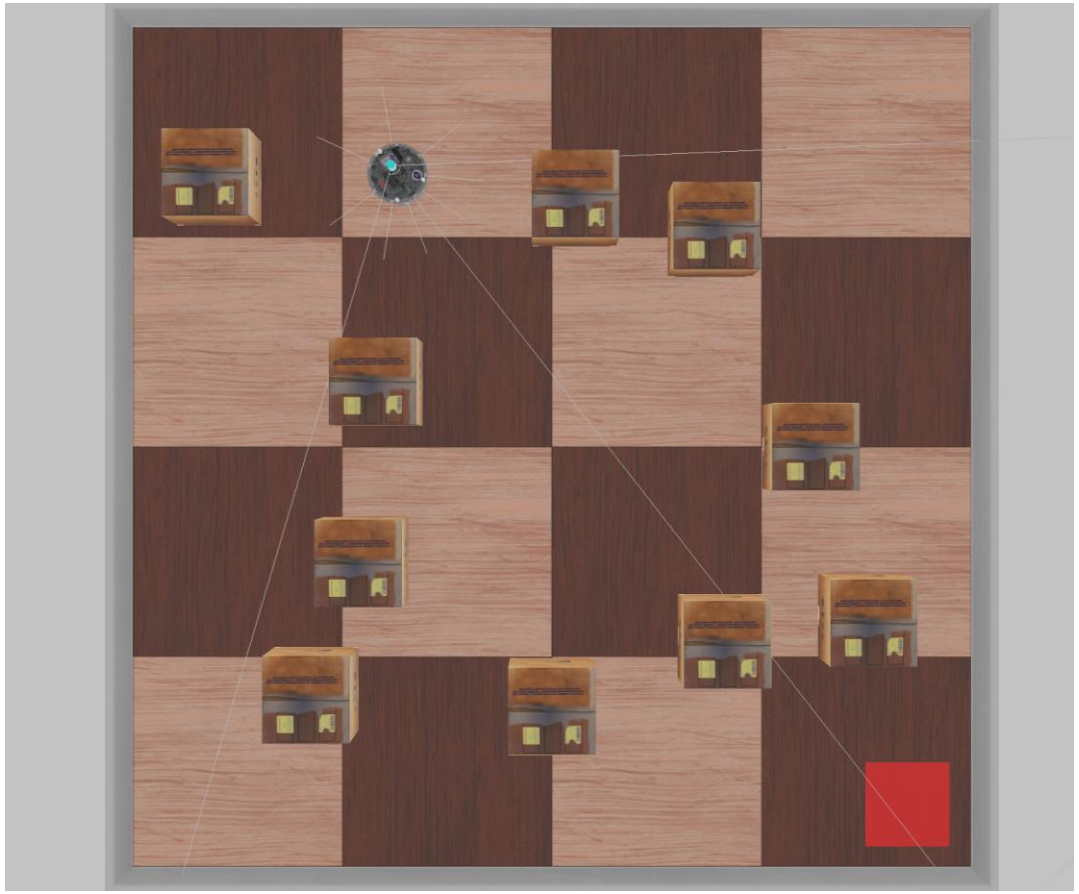
A (action-space): Additional or fewer actions, changed ranges of continuous actions

R (reward-function): Modified reward function that reinforces different behavior

P (transition dynamics): Different transition probabilities, given the same (S, A)-pair

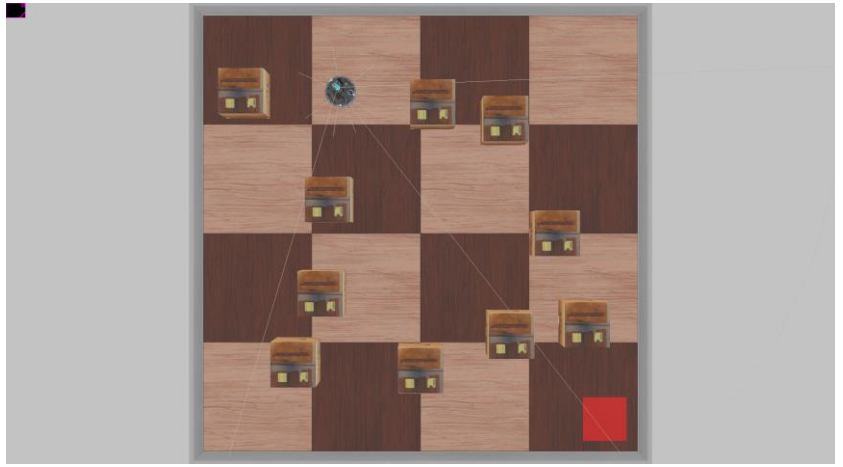
μ_0(initial states): different initial state distribution

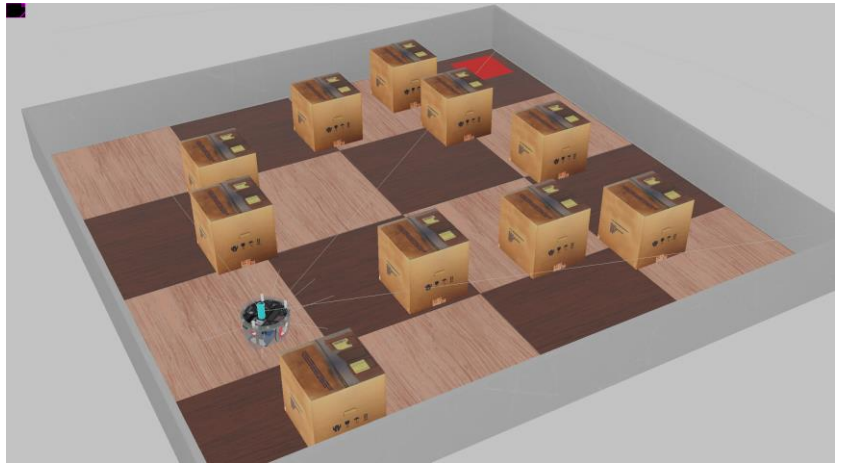
EXAMPLE OBSTACLE AVOIDANCE SCENARIO



EXAMPLE OBSTACLE AVOIDANCE SCENARIO

	S	A	P	R	μ_0
Workers interacting with the AGV	✓		✓		
Other robots interacting with the AGV	✓		✓		
Changed warehouse layout	✓				
Multiple goals				✓	
Unusual starting position					✓
Malfunctions of the AGV		(✓)	✓		
Noisy sensors	✓				





CONCLUSION

- There is a lot of existing work towards this, going under different names and making different assumptions (robustness, domain adaption/generalization, OOD, …)
- Decomposition of complex Problems into aspects of the MDP is straightforward
- Isolating individual threats into aspects of the MDP helps safety argumentation (separation of concerns) and comparability of approaches
- Isolating disturbances and safety threats is only a first step
  → detecting and handling them is still complicated → **future work**

REFERENCES (EXCERPT)

- (Hansen, Wang; "*Generalization in Reinforcement Learning by Soft Data Augmentation*"; 2021)
- (Koopman, Fratrik; "*How Many Operational Design Domains, Objects, and Events?*"; 2019)
- (Eysenbach, Levine; "*Maximum Entropy RL (Provably) Solves Some Robust RL Problems*"; 2021)
- (Tobin et al.; "*Domain randomization for transferring deep neural networks from simulation to the real world*"; 2017)
- (Taylor, Stone; "*Transfer Learning for Reinforcement Learning Domains: A Survey*"; 2009)