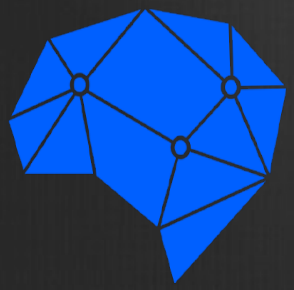
Distributionally Robust Q-Learning

Zhengyuan Zhou (joint with)

Qinxun Bai, Jose Blanchet, Perry Dong, Zijian Liu, Wei Xu and Zhengqing Zhou



Horizon Robotics

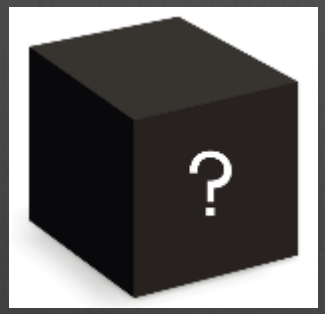
Stanford University

Reinforcement Learning

Historical data

or

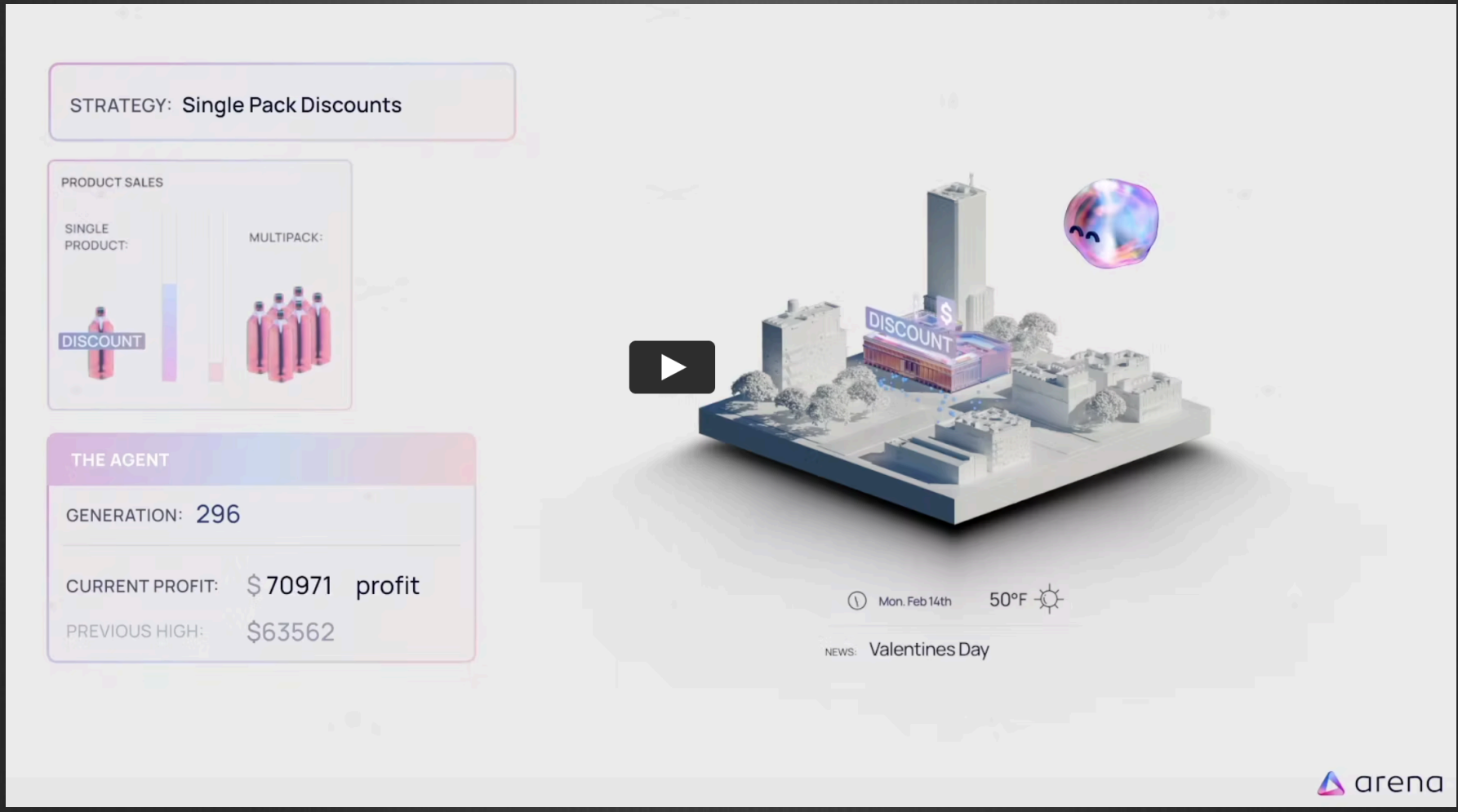
Simulator



policy

Deployment

Real Environment Applications



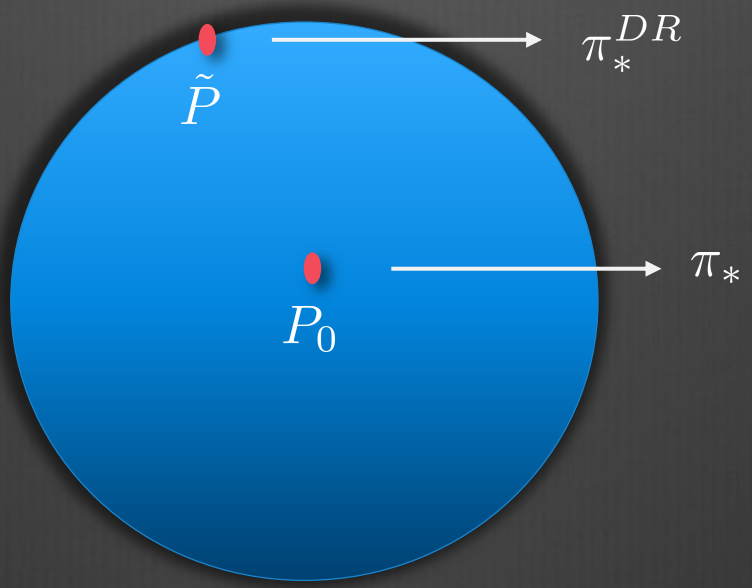
Motivation

Fragile

- Simulator discrepancy
- Environment shifts

How to be robust?

How to be Robust?



A Distributionally Robust Approach

Problem

$$V^{\pi}(s) := \mathbb{E}\left[\sum_{t=1}^{\infty} \gamma^{t-1} r(s_t, a_t) \,\middle|\, s_1 = s\right]$$

$$V_{\delta}^{\mathrm{rob},\pi}(s) := \inf_{\mathbf{p}\in\mathcal{P}(\delta),\mathbf{r}\in\mathcal{R}(\delta)} \mathbb{E}_{\mathbf{p},\mathbf{r}}\left[\sum_{t=1}^{\infty} \gamma^{t-1} r(s_t, a_t) \,\middle|\, s_1 = s\right]$$

$$V_{\delta}^{\mathrm{rob},*}(s) := \max_{\pi\in\Pi} V_{\delta}^{\mathrm{rob},\pi}(s), \quad \forall s \in S$$

Recent Work

Distributionally robust contextual bandits:

Nian, Zhang, Zhou and Blanchet, ICML 2020

Model-based distributionally robust RL:

Zhou et al. AISTATS 2021

Yang et al. arXiv: 2105.03863.2021

Kishan and Kalathil, arXiv: 2112.01506, 2021

Kido, arXiv: 2205.04637, 2022

What about model-free distributionally robust learning algorithms?

Main Question

- Can we design a distributionally robust Q-Learning algorithm?

- Come check us out:

  Hall E # 932

  Wed (today): 6:30pm – 8:30pm