Assessing the Impact of Context Inference Error and Partial Observability on RL Methods for Just-In-Time Adaptive Interventions

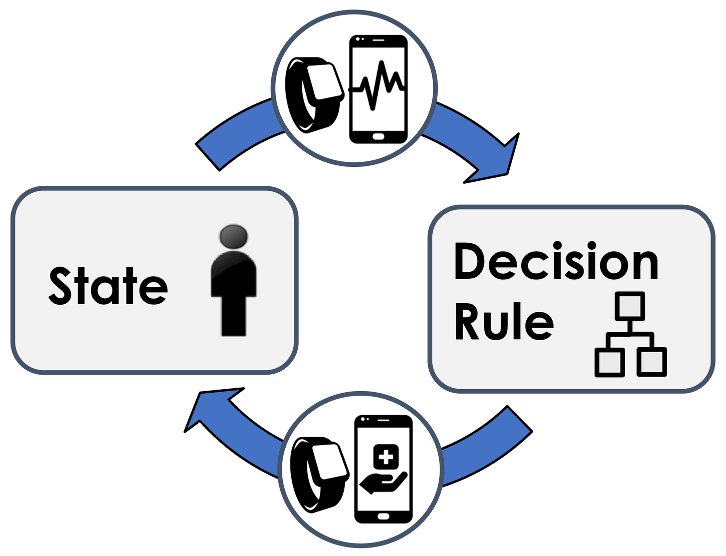
Karine[1], Pedja Klasnja[2], Susan A. Murphy[3], Benjamin M. Marlin[1]

1. University of Massachusetts Amherst, 2. University of Michigan, 3. Harvard University

SCAN ME

Introduction

- **Just-in-Time Adaptive Interventions (JITAIs)** are a type of personalized health intervention. The goal is to **select** the **best intervention** option, using some decision rules, based on the participant's state.
- However, JITAI problem domains typically contain unobserved state.



- To deal with this, JITAIs use ML-based context inferences to estimate parts of the unobserved states, but JITAIs often do not account for uncertainty in context inferences.
- In this work, we investigate the **application of RL algorithms** to JITAIs and investigate the impact for context inference error and uncertainty on the performance of learned policies.

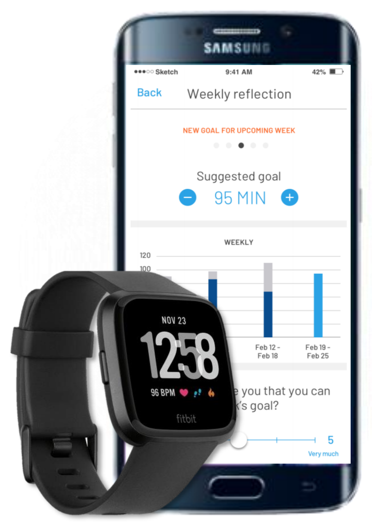
Contributions

- ✓ Introduce a new **JITAI simulation** that captures dynamics of behavior, as well as **context inference uncertainty**. This data simulator is very useful.
- ✓ Show that **policies** that use **context inference probabilities** as feature, significantly **outperform policies** that use only **most likely context** value.
- ✓ Show that under **partial observability,** the **performance** of **DQN drops much more than** the performance of **REINFORCE**.

JITAI Simulation Environment: State Variables

We develop a JITAI simulation environment inspired by interventions including the HeartSteps trial, an adaptive messaging intervention that aims to increase physical activity levels. The simulation models several key variables including:

HeartSteps



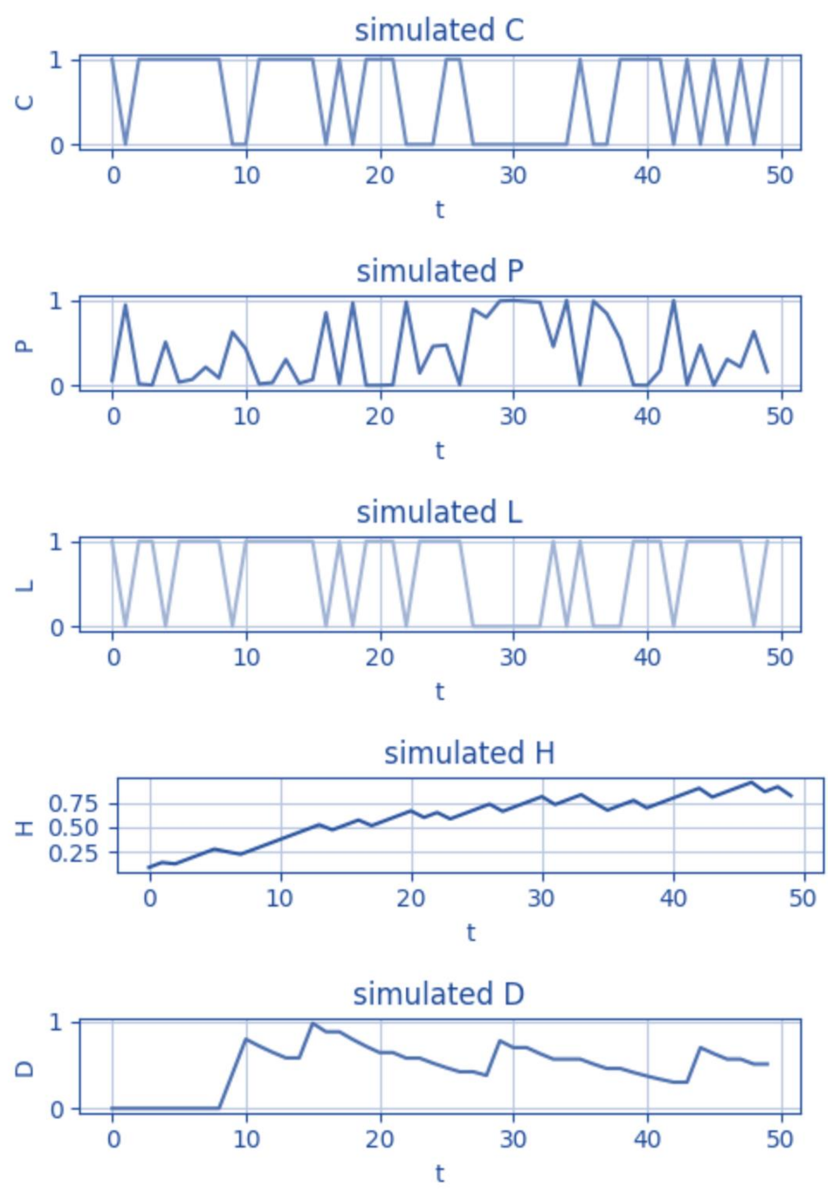
- an unobserved binary context (e.g., stress)
- disengagement risk level
- habituation level
- number of steps

Variable	Description	Values
c_t	true context	$\{0,1\}$
$\mathbf{p}_t$	context probabilities	Δ^1
l_t	most likely context	$\{0,1\}$
d_t	disengagement risk level	$[0,1]$
h_t	habituation level	$[0,1]$
s_t	number of steps	$\mathbb{N}$

Note: we assume we can not observe the true context c_t. Thus, we introduce a **context inference probability p_t** and **most likely context value: l_t**

JITAI Simulation Environment: Dynamics

State Dynamics: The dynamics of the state are driven by the interaction between the true context and the action. When the message selected does not match the context, disengagement risk increases. Sending any message increases habituation. We simulate context inference based on a class-conditional feature distribution. When the disengagement risk reaches 1, the simulated trial ends.



$$c_t \sim \text{Bernoulli}(0.5)$$
$$x_t \sim \mathcal{N}(c_t, \sigma^2)$$

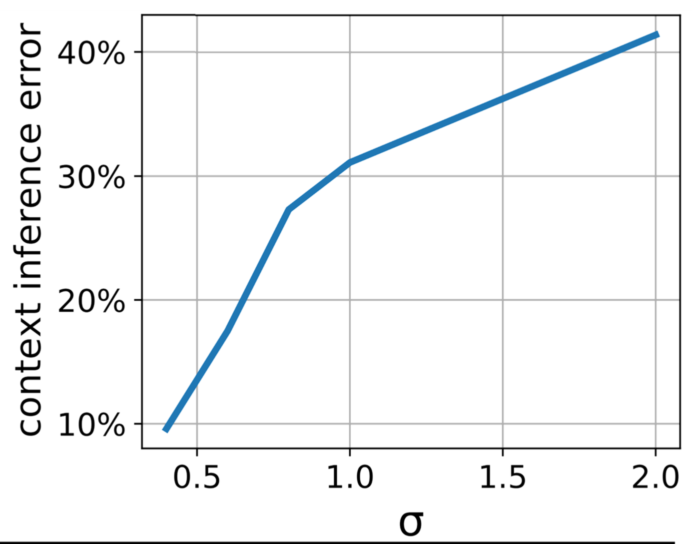
$$p_{ct} = P(C_t = c|x_t)$$

$$l_t = \text{argmax}_c \; p_{ct}$$

$$h_{t+1} = \begin{cases} (1-\delta_h)\cdot h_t & \text{if } a_t = 0 \\ \min(1, h_t + \epsilon_h) & \text{otherwise} \end{cases}$$

$$d_{t+1} = \begin{cases} d_t & \text{if } a_t = 0 \\ (1-\delta_d)\cdot d_t & \text{if } a_t = 1 \text{ or } a_t = c_t + 2 \\ \min(1, d_t + \epsilon_d) & \text{otherwise} \end{cases}$$

Parameter	Description	Value
δ_h	habituation decay	0.1
ϵ_h	habituation increment	0.05
δ_d	disengagement decay	0.1-0.4
ϵ_d	disengagement increment	0.1-0.4
ρ_1	$a_t = 1$ base reward	50.
ρ_2	$a_t = c_t + 2$ base reward	200.
σ	feature uncertainty	$\{0.4, ..., 2\}$



Actions: We model four actions including a null action (do not send a message), sending a non-tailored message, and sending a context-tailored message.

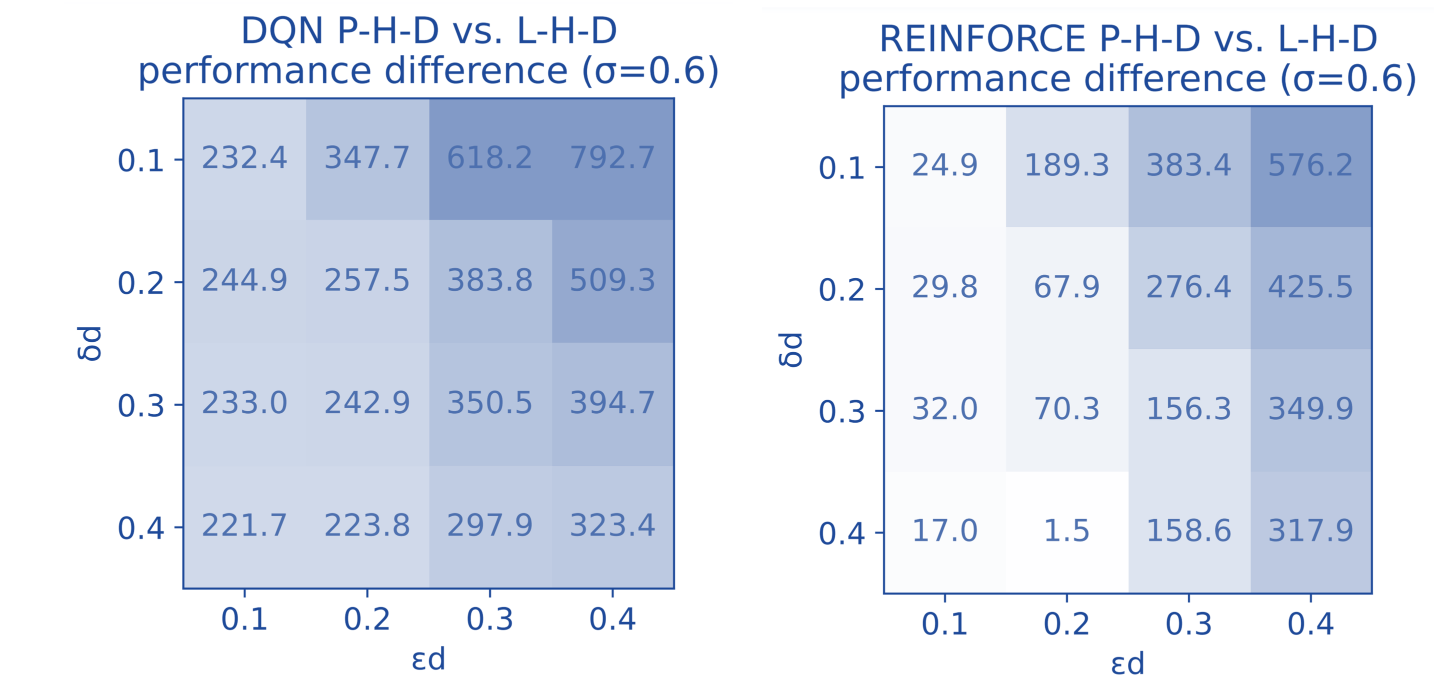
Action Value	Description
$a=0$	do not send a message
$a=1$	send a non-tailored message
$a=2$	send a message tailored to context 0
$a=3$	send a message tailored to context 1

Rewards: We model the reward as a step count that depends on the action taken, the true context, and the level of habituation to the intervention.

$$s_{t+1} = \begin{cases} \mu_{c_t} + (1-h_{t+1})\cdot\rho_1 & \text{if } a_t = 1 \\ \mu_{c_t} + (1-h_{t+1})\cdot\rho_2 & \text{if } a_t = c_t + 2 \\ \mu_{c_t} & \text{otherwise} \end{cases}$$

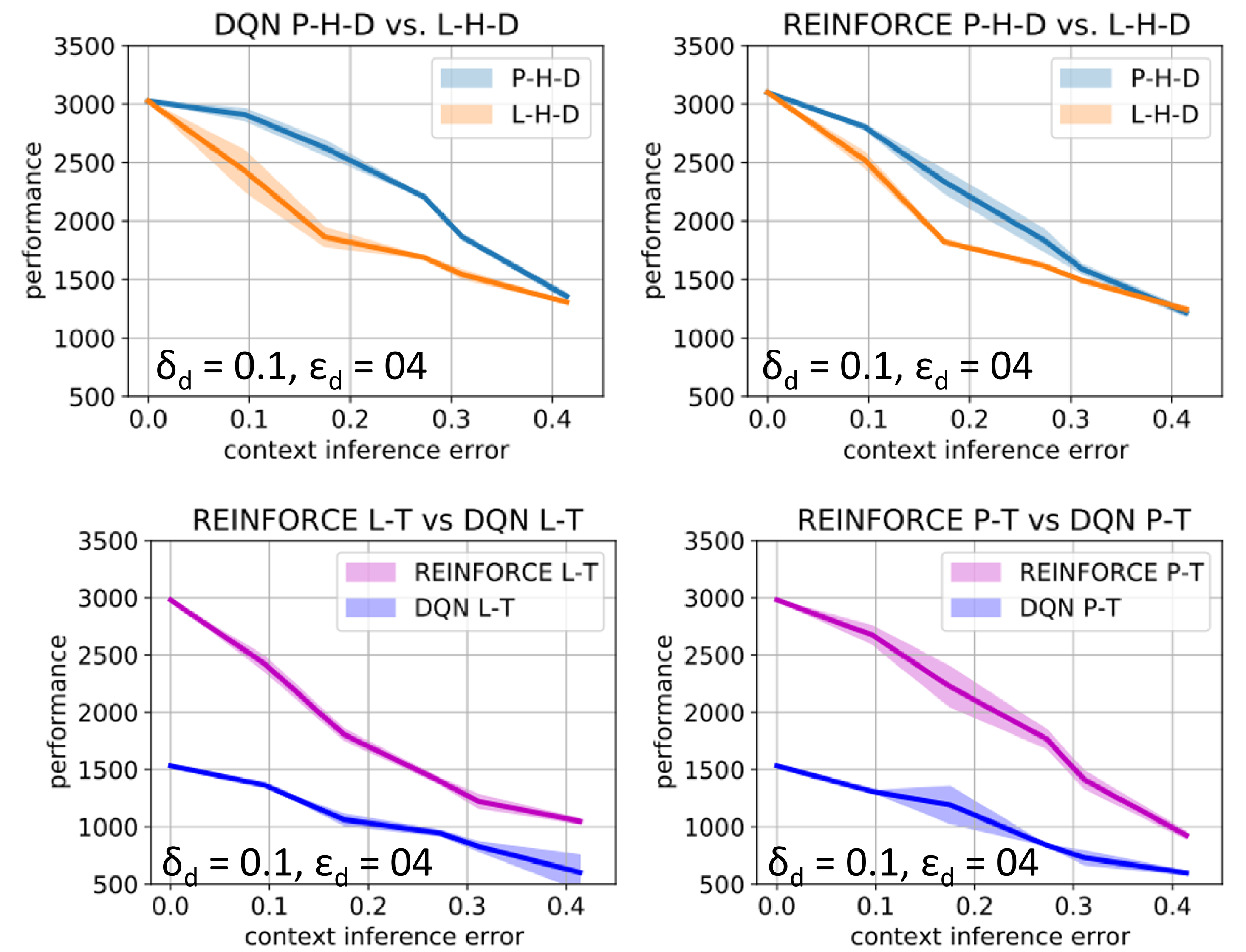
Experiments and Results

Experiments: We conduct simulations using a DQN-based method and REINFORCE. We optimize the architecture of the models used and do not constrain the number of episodes to ensure convergence to optimal representable policies. Below we compare context inference probabilities to most likely contexts (P-H-D vs L-H-D), and the performance of DQN vs the performance of REINFORCE under partial observability (L-T and P-T).



δd \ εd (DQN)	0.1	0.2	0.3	0.4
0.1	232.4	347.7	618.2	792.7
0.2	244.9	257.5	383.8	509.3
0.3	233.0	242.9	350.5	394.7
0.4	221.7	223.8	297.9	323.4

δd \ εd (REINFORCE)	0.1	0.2	0.3	0.4
0.1	24.9	189.3	383.4	576.2
0.2	29.8	67.9	276.4	425.5
0.3	32.0	70.3	156.3	349.9
0.4	17.0	1.5	158.6	317.9



Conclusions and Future Work

- Policies using context inference probabilities outperform policies using most likely context value for both DQN and REINFORCE.
- Under partial observability, DQN performance drops significantly more than REINFORCE performance.
- DQN performance may be improved via state augmentation methods to better deal with partial observability.