Reinforcement Learning and Applications to Mobile Health

Manchester Interdisciplinary Mathematics Undergraduate Conference (MIMUC)

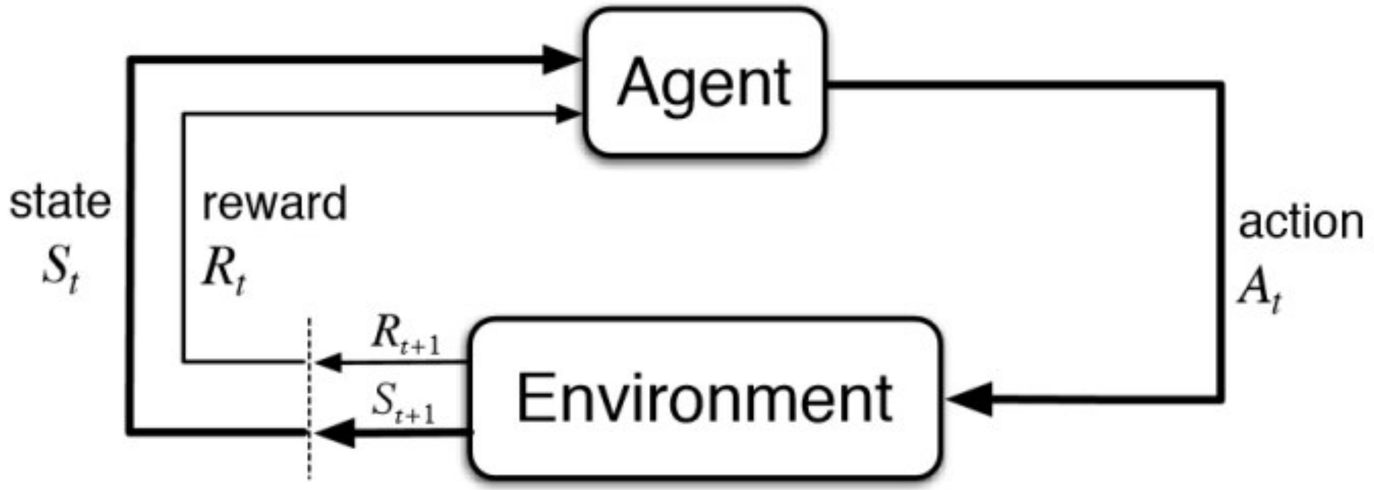
Lluís Salvat Niell

Principles of Reinforcement Learning

What is Reinforcement Learning?

The science of sequential decision-making

- What is intelligence?
- Computer Science + Psychology + Economics + Mathematics + Engineering
- Goal: maximise expected cumulative reward
- Applications: large language models (LLMs) and recommendation systems
- RL ⊂ ML ⊂ AI



Markov Decision Processes (MDPs)

A world model for RL, $\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma \rangle$

- State space: states are Markovian, $\mathbb{P}[S_{t+1}|S_t] = \mathbb{P}[S_{t+1}|S_1, \ldots, S_t]$
- Action space
- State transition matrix: $\mathcal{P}^a_{ss'} = \mathbb{P}[S_{t+1} = s'|S_t = s, A_t = a]$
- Reward function: $\mathcal{R}^a_s = \mathbb{E}[R_{t+1}|S_t = s, A_t = a]$
- Discount factor: $\gamma \in [0, 1]$

Reinforcement Learning Agent

The decision-maker

- Policy: $\pi : \mathcal{S} \rightarrow \mathcal{A}$
- Value Function: $v_\pi(s) = \mathbb{E}_\pi[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \ldots | S_t = s]$
- Q-Function: $q_\pi(s, a) = \mathbb{E}_\pi[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \ldots | S_t = s, A_t = a]$
- Value and Q-functions can be approximated via neural networks — Deep RL!

Reinforcement Learning in Mobile Health and DyadicRL

What is Mobile Health?

Improving health outcomes with technology

- Enhance health outcomes by delivering digital interventions to individuals
- Examples: medication adherence, physical activity, dental hygiene

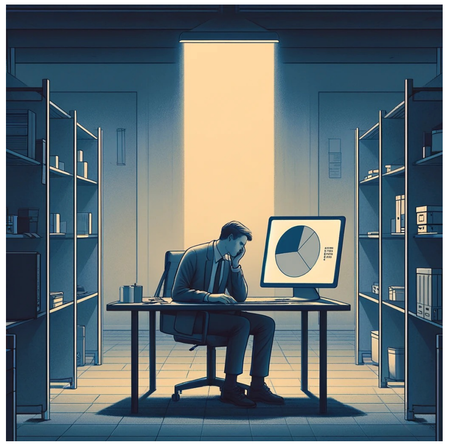
What are the Challenges in Mobile Health?

Domain-specific problems

User Heterogeneity

Intervention Burden

High-noise Environment



Sample Efficiency

Reward Sparsity

User Non-stationarity

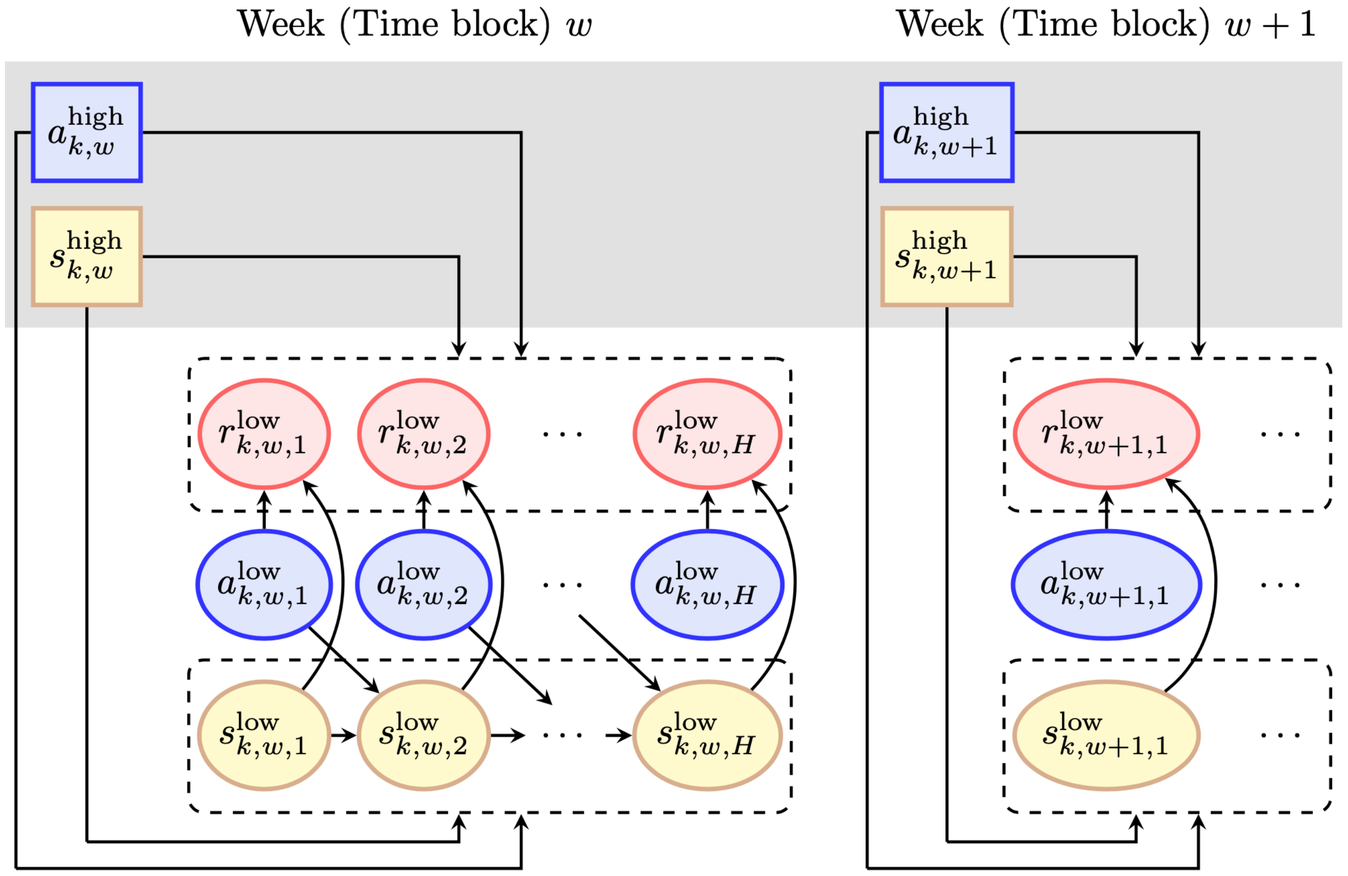
Dyadic Reinforcement Learning

A Bayesian RL algorithm for mobile health and beyond

- Support networks are crucial for recovery
- Dyad = Target Person + Care Partner, ADAPTS HCT
- Perhaps delivering multiple interventions at different time scales increases reward
- Hierarchical structure

Dyadic Reinforcement Learning

A Bayesian RL algorithm for mobile health and beyond

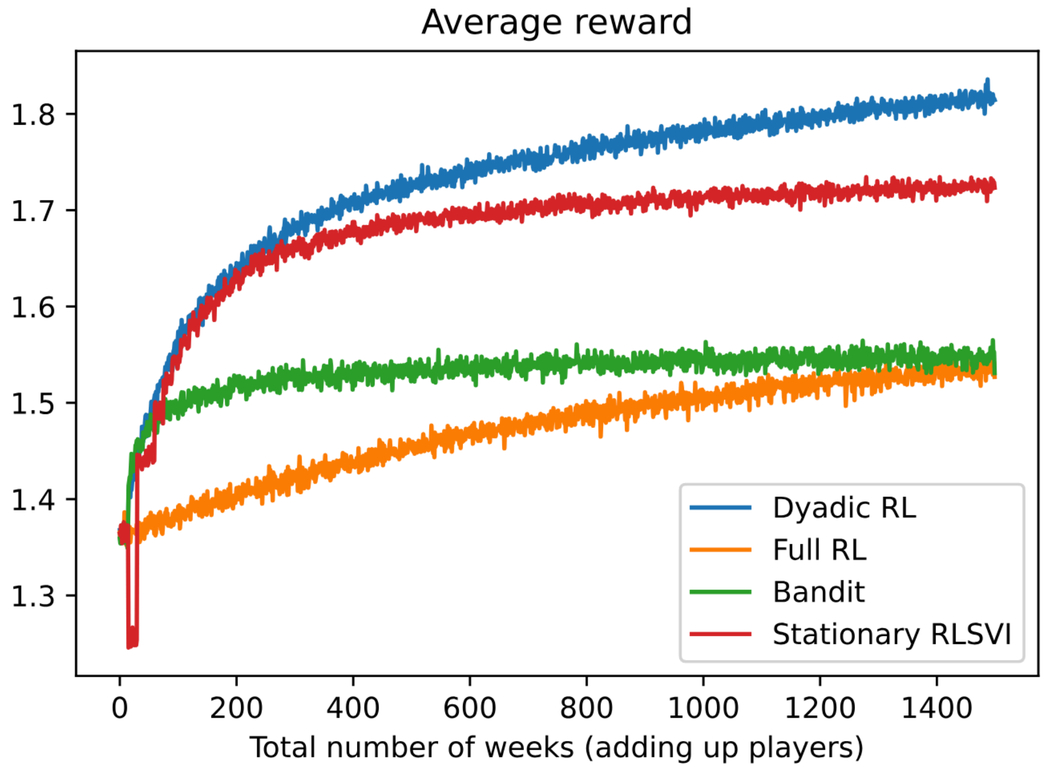


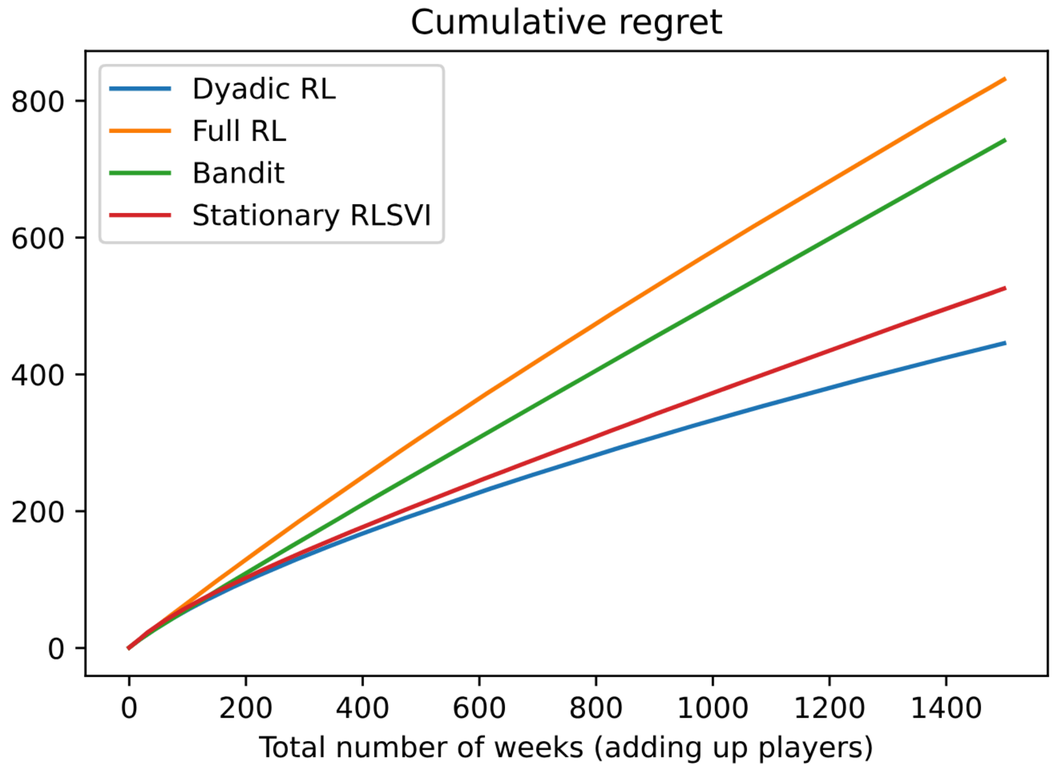
Algorithm Dyadic Reinforcement Learning

input feature mappings and parameters
for episode $k = 1, \ldots, K$ **do**
  for week $w = 1, \ldots, W$ **do**
    observe high-level state
    get $\tilde{\beta}_{w,h}$ from bandit RLSVI $(H = 1)$
    sample high-level action $\text{argmax}_\alpha \tilde{\beta}^\top_{k,w} \psi$
    get $\{\tilde{\theta}_{k,w,h}\}_{h=1}^H$ from MDP RLSVI
    for day $h = 1, \ldots, H$ **do**
      observe low-level state
      sample low-level action $\text{argmax}_\alpha \tilde{\theta}^\top_{k,w,h} \phi_h$
      observe low-level reward
    construct high-level reward $\max_\alpha \tilde{\theta}^\top_{k,w,1} \phi_1$

Dyadic Reinforcement Learning

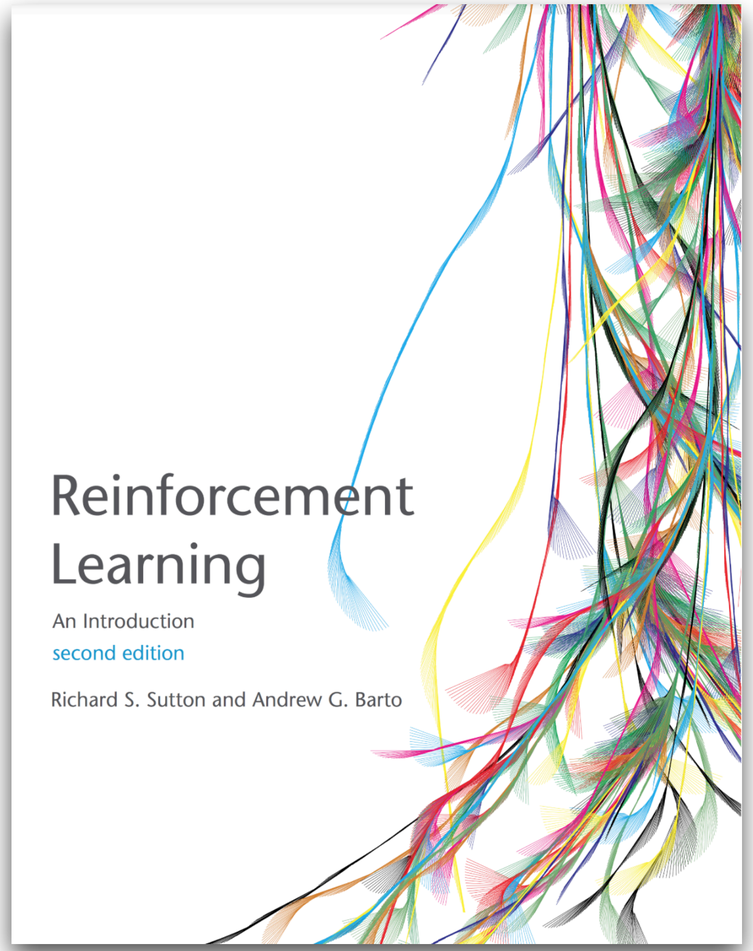
A Bayesian RL algorithm for mobile health and beyond





Further Reading

Resources for deeper understanding



http://incompleteideas.net/book/the-book-2nd.html

A Tutorial on Thompson Sampling

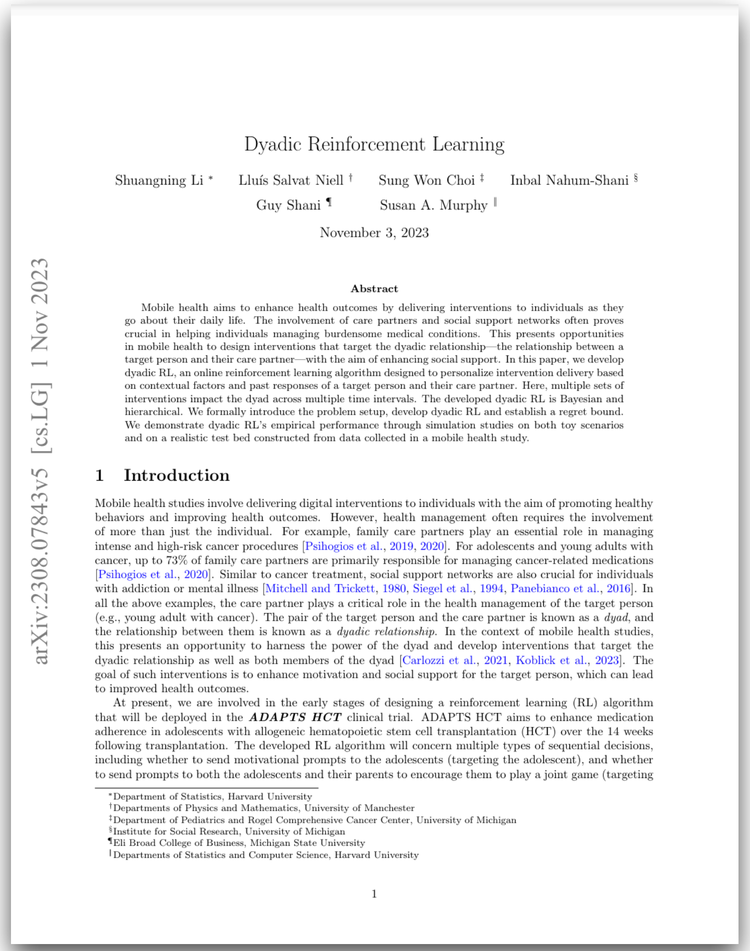
Daniel J. Russo[1], Benjamin Van Roy[2], Abbas Kazerouni[2], Ian Osband[3] and Zheng Wen[4]

[1]*Columbia University*
[2]*Stanford University*
[3]*Google DeepMind*
[4]*Adobe Research*

ABSTRACT

Thompson sampling is an algorithm for online decision problems where actions are taken sequentially in a manner that must balance between exploiting what is known to maximize immediate performance and investing to accumulate new information that may improve future performance. The algorithm addresses a broad range of problems in a computationally efficient manner and is therefore enjoying wide use. This tutorial covers the algorithm and its application, illustrating concepts through a range of examples, including Bernoulli bandit problems, shortest path problems, product recommendation, assortment, active learning with neural networks, and reinforcement learning in Markov decision processes. Most of these problems involve complex information structures, where information revealed by taking an action informs beliefs about other actions. We will also discuss when and why Thompson sampling is or is not effective and relations to alternative algorithms.

https://arxiv.org/abs/1707.02038



Dyadic Reinforcement Learning

Shuangning Li * Lluís Salvat Niell † Sung Won Choi ‡ Inbal Nahum-Shani §
Guy Shani ¶ Susan A. Murphy ‖

November 3, 2023

Abstract

Mobile health aims to enhance health outcomes by delivering interventions to individuals as they go about their daily life. The involvement of care partners and social support networks often proves crucial in helping individuals managing burdensome medical conditions. This presents opportunities in mobile health to design interventions that target the dyadic relationship—the relationship between a target person and their care partner—with the aim of enhancing social support. In this paper, we develop dyadic RL, an online reinforcement learning algorithm designed to personalize intervention delivery based on contextual factors and past responses of a target person and their care partner. Here, multiple sets of interventions impact the dyad across multiple time intervals. The developed dyadic RL is Bayesian and hierarchical. We formally introduce the problem setup, develop dyadic RL and establish a regret bound. We demonstrate dyadic RL's empirical performance through simulation studies on both toy scenarios and on a realistic test bed constructed from data collected in a mobile health study.

1 Introduction

Mobile health studies involve delivering digital interventions to individuals with the aim of promoting healthy behaviors and improving health outcomes. However, health management often requires the involvement of more than just the individual. For example, family care partners play an essential role in managing intense and high-risk cancer procedures [Psihogios et al., 2019, 2020]. For adolescents and young adults with cancer, up to 73% of family care partners are primarily responsible for managing cancer-related medications [Psihogios et al., 2020]. Similar to cancer treatment, social support networks are also crucial for individuals with addiction or mental illness [Mitchell and Trickett, 1980, Siegel et al., 1994, Panebianco et al., 2016]. In all the above examples, the care partner plays a critical role in the health management of the target person (e.g., young adult with cancer). The pair of the target person and the care partner is known as a *dyad*, and the relationship between them is known as a *dyadic relationship*. In the context of mobile health studies, this presents an opportunity to harness the power of the dyad and develop interventions that target the dyadic relationship as well as both members of the dyad [Carlozzi et al., 2021, Koblick et al., 2023]. The goal of such interventions is to enhance motivation and social support for the target person, which can lead to improved health outcomes.

At present, we are involved in the early stages of designing a reinforcement learning (RL) algorithm that will be deployed in the ***ADAPTS HCT*** clinical trial. ADAPTS HCT aims to enhance medication adherence in adolescents with allogeneic hematopoietic stem cell transplantation (HCT) over the 14 weeks following transplantation. The developed RL algorithm will concern multiple types of sequential decisions, including whether to send motivational prompts to the adolescents (targeting the adolescent), and whether to send prompts to both the adolescents and their parents to encourage them to play a joint game (targeting

*Department of Statistics, Harvard University
†Departments of Physics and Mathematics, University of Manchester
‡Department of Pediatrics and Rogel Comprehensive Cancer Center, University of Michigan
§Institute for Social Research, University of Michigan
¶Eli Broad College of Business, Michigan State University
‖Departments of Statistics and Computer Science, Harvard University

https://arxiv.org/abs/2308.07843