Dynamic Treatment Plan Optimization for Chronic Disease Management

A Research Overview Using Contextual Restless Bandit Models

Chelluri Ashish Dheer (M240342CS)

Guided by: Dr. Vidhya Kamakshi
Project Progress and Future Work

August 21, 2025

Agenda

The Challenge: Managing Chronic Disease

The Problem with Static Plans

- Chronic diseases are the leading cause of death and disability in the U.S., with **6 in 10 adults** having at least one condition.[1]
- Static, "one-size-fits-all" treatment plans are fundamentally ill-suited for the dynamic, evolving nature of these illnesses.[2]
- This is especially true for patients with Multiple Chronic Conditions (MCCs), leading to care fragmentation, medical errors, and increased costs.[3]

Our Goal: Adaptive, Personalized Care

- To move beyond static plans by creating an adaptive framework that personalizes treatment in real-time.
- This requires a model that can learn from a patient's evolving state to optimize long-term health outcomes. This is the core idea of **Dynamic Treatment Regimes**

The Core Dilemma: Exploration vs. Exploitation

The Multi-Armed Bandit (MAB) Analogy

- Imagine facing multiple slot machines ("bandits"), each with an unknown payout rate.[5]
- To maximize winnings, you must balance two actions:
  - **Exploitation:** Pulling the arm with the best results so far.
  - **Exploration:** Trying other arms to see if they might be better.
- This tradeoff is fundamental in medical decision-making, where a doctor must balance using the best-known treatment with exploring new options for future patients.

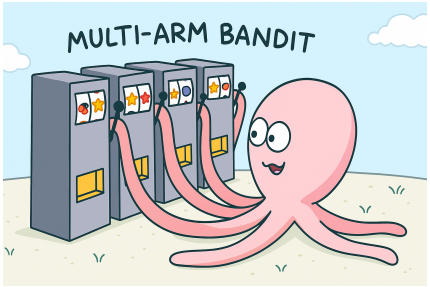

Figure 1: Which machine should you play?

Formalizing the Dilemma: The UCB1 Algorithm

The UCB1 Algorithm Equation

The Upper Confidence Bound (UCB1) algorithm formalizes this tradeoff. At each step t, it selects the arm a that maximizes:

$$\text{action}_t = \arg\max_a \left(\underbrace{\hat{\mu}_a}_{\text{Exploit}} + \underbrace{\sqrt{\frac{2 \ln t}{n_a}}}_{\text{Explore}} \right)$$

Where $\hat{\mu}_a$ is the current average reward for arm a, t is the total number of plays, and n_a is the number of times arm a has been played.[6]

- The first term encourages **exploiting** the arm with the highest known average reward.
- The second term encourages **exploring** arms with high uncertainty.

From MAB to RMAB: Modeling the "Restless" Patient

The Standard MAB Assumption is Flawed for Healthcare

A standard MAB assumes the slot machines' payout rates are **fixed**. They don't change over time.

The Reality: Patients are "Restless" face-with-medical-mask

- A patient's health state is **not fixed**. It can change on its own, even without an intervention.[4]
- This is what "Restless" means in a **Restless Multi-Armed Bandit (RMAB)**. The state of an arm (patient) can evolve even when it is not actively pulled.[7]
- This makes RMABs ideal for modeling public health scenarios where resources must be allocated to a population of patients with evolving health states.

Adding Patient Data: The Contextual Bandit

The Missing Piece: Personalization

Treating every patient the same is not optimal. We need to personalize the treatment based on their unique characteristics. This is where "context" comes in.

The Linear Reward Assumption

Contextual bandits like LinUCB often assume the expected reward is a linear function of the context features:

$$E[r_{t,a}|x_{t,a}] = x_{t,a}^{\top}\theta_a^*$$

Where $x_{t,a}$ is the feature vector (context) for action a at time t, and θ_a^* is an unknown weight vector for that action. The algorithm's goal is to learn these weights.[8]

*"What is the best treatment for a patient with **this specific context**?"*

Why Simpler Variants Are Insufficient

Two Core Requirements for Chronic Disease Modeling

Any effective model must address both:

- **Temporal Dynamics:** How a patient's health evolves over time, even passively.
- **Patient Heterogeneity:** How patients differ from one another.

Limitations of Standard Bandit Models

- **Standard MAB:** Fails because it assumes a static environment.
- **Contextual MAB:** Captures patient differences but assumes their underlying health state is static.
- **Restless MAB (RMAB):** Models the dynamic evolution of a patient's state but fails to personalize decisions.

Our Approach: The Contextual Restless Bandit (CRB)

The Right Foundational Model

- The **Contextual Restless Bandit (CRB)** is the first framework to formally synthesize the two essential dimensions: the internal "restless" state and the external "context".[9]
- It provides the most appropriate and complete model that directly maps to the clinical problem.

Focus on Core Complexity First

- More advanced hybrid models, such as those combining CRBs with Factorial or Combinatorial structures, address exponentially complex action spaces and introduce significant computational challenges.[10]
- Our current focus is to establish a proof-of-concept with the core CRB model. Exploring these more complex hybrids is a logical next step for future work.

Our Research Question

How can we leverage the Contextual Restless Bandit framework to learn optimal, personalized, and dynamic treatment regimes from observational EHR data, thereby overcoming the limitations of static care models for chronic disease management?

Problem Formulation and Dataset

Mapping our Problem to a CRB

- **Arms:** Each patient with a chronic condition.
- **Actions:** Applying a specific treatment plan.
- **Internal State:** Patient-specific factors like medication adherence.
- **Global Context:** External factors like time of day.
- **Reward:** A measure of improvement in the patient's health.

Dataset: MIMIC-III

- We will use the **MIMIC-III database**, a large, de-identified critical care database.[11]
- It contains comprehensive, longitudinal data including demographics, vitals, lab results, and medications, making it ideal for this research.

Proposed Implementation Pipeline

Our plan is to build a complete CRB pipeline:

1. **Data Preprocessing & Feature Engineering:**
   - Select a relevant patient cohort from MIMIC-III.
   - Clean and normalize the data: handle irregular time-series, impute missing values, and aggregate features into robust clinical concepts.
   - Utilize open-source tools like **MIMIC-Extract** to ensure reproducibility.[12]
2. **Reward Function Engineering:**
   - This is a critical challenge: defining a scalar reward that captures complex clinical goals.[13]
   - We will explore outcomes like in-hospital mortality (binary) or SOFA score changes (continuous).
3. **Algorithm Implementation Evaluation:**
   - Implement a CRB algorithm and train it on the preprocessed MIMIC-III data.
   - Since we cannot test on live patients, we will use **Off-Policy Evaluation (OPE)** methods to estimate our policy's performance.[14]

Summary and Expected Contribution

Summary

- Chronic disease requires dynamic, personalized treatment, which static models fail to provide.
- The Contextual Restless Bandit (CRB) framework is perfectly suited to model this complex, real-world problem.
- We will use the rich MIMIC-III dataset to develop and validate our approach.

Expected Contribution

- A novel framework for optimizing **individual-level** dynamic treatment plans.
- A validated proof-of-concept demonstrating the potential of AI to improve patient care and resource efficiency.

Challenges and Future Work

Key Challenges and Research Frontiers

While promising, this work must address several critical challenges that represent active areas of research:

- **Equity and Fairness:** Standard optimization can lead to "arm starvation," where certain patient groups are neglected. Future work must explore **Equitable RMABs (ERMABs)**.[7]
- **Human-in-the-Loop Systems:** Models should not ignore clinical expertise. Frameworks like **Recourse Bandits** create a collaborative partnership between the AI and human experts.[15]
- **Interpretability and Trust:** For clinical adoption, "black box" models are insufficient. We need interpretable models and evaluation metrics that provide clear rationales for AI-driven recommendations.

References I

[1] Centers for Disease Control and Prevention. *Living With a Chronic Disease*. https://www.cdc.gov/chronic-disease/living-with/index.html. Accessed: August 21, 2025. 2023.

[2] Wil M van der Aalst, Mathias Weske, and Akhil Kumar. "The current approach to care for patients with chronic diseases: a literature review". In: *Journal of healthcare engineering* 2015 (2015).

[3] ChartSpan. *5 Challenges of Managing Patients with Multiple Chronic Conditions*. https://www.chartspan.com/blog/managing-multiple-chronic-conditions/. Accessed: August 21, 2025. 2022.

[4] Bibhas Chakraborty and Susan A Murphy. "Dynamic treatment regimes". In: *Annual review of statistics and its application* 1 (2014), pp. 447–464.

References II

[5] Wikipedia. *Multi-armed bandit.* https://en.wikipedia.org/wiki/Multi-armed_bandit. Accessed: August 21, 2025. 2024.

[6] Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. "Finite-time analysis of the multiarmed bandit problem". In: *Machine Learning*. Vol. 47. Springer. 2002, pp. 235–256.

[7] Dexun Li, Pradeep Varakantham, and Feng Wu. "Fairness in Restless Multi-Armed Bandits". In: *International Conference on Machine Learning*. PMLR. 2022, pp. 12896–12915.

[8] Wei Chu et al. "Contextual Bandits with Linear Payoff Functions". In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. 2011, pp. 208–214.

References III

[9] Jian Li et al. “Contextual Restless Bandits”. In: *arXiv preprint arXiv:2403.15640* (2024).

[10] T Killian, A Mate, and M Tambe. “Reinforcement learning with combinatorial actions for coupled restless bandits”. In: *International Conference on Learning Representations*. 2022.

[11] Alistair EW Johnson et al. “MIMIC-III, a freely accessible critical care database”. In: *Scientific data* 3.1 (2016), pp. 1–9.

[12] S Wang, M Ghassemi, and T Naumann. “Mimic-extract: A data extraction, preprocessing, and representation pipeline for mimic-iii”. In: *Proceedings of the ACM Conference on Health, Inference, and Learning*. 2020, pp. 1–11.

[13] N Prasad et al. “A deep reinforcement learning framework for dynamic treatment regimes”. In: *Conference on Health, Inference, and Learning*. 2017, pp. 1–10.

References IV

[14] J P Hanna, P Stone, and S Niekum. “UnO: A Universal Off-Policy Evaluation Estimator”. In: *Advances in Neural Information Processing Systems*. Vol. 34. 2021, pp. 2303–2315.

[15] M Cao et al. “HR-Bandit: Human-AI Collaborated Linear Recourse Bandit”. In: *arXiv preprint arXiv:2403.15640* (2024).

Thank you!

Questions?

Chelluri Ashish Dheer

M240342CS