Dynamic Treatment Plan Optimization for Chronic Disease Management Using Restless Multi-Armed Bandit Algorithms

Chelluri Ashish Dheer
M240342CS
Guided by: Dr. Vidhya Kamakshi

March 4, 2025

Keywords: Restless Multi-Arm Bandit (RMAB), Chronic Disease Management, Dynamic Treatment, Reinforcement Learning, Healthcare Optimization, MIMIC-III Dataset

Abstract

Chronic diseases, such as diabetes and cardiovascular disorders, are dynamic in nature and require personalized treatment strategies to manage effectively over time. Static treatment plans often fail to accommodate the changing health status of patients, leading to suboptimal outcomes. This research explores the application of Restless Multi-Armed Bandit (RMAB) algorithms to optimize dynamic treatment plans for chronic disease management. The goal is to develop an adaptive decision-making framework that personalizes treatments based on patient responses, maximizing long-term health outcomes. Using the MIMIC-III dataset, this study evaluates the effectiveness of RMAB in the allocation of healthcare resources. Through simulations and real-world data analysis, RMAB algorithms demonstrate significant potential to improve patient care and resource efficiency. This research contributes to the growing body of work in AI-driven healthcare optimization by addressing both personalization and equity in treatment planning.

1 Introduction

Chronic disease management is a critical challenge in healthcare, requiring continuous monitoring and personalized interventions. Conditions such as diabetes, cardiovascular disease, and hypertension often progress in unpredictable ways, necessitating dynamic and flexible treatment strategies. Traditional static treatment plans are not well suited to handle the complexity of chronic conditions, as they do not account for the evolving nature of patient health. As a result, there is an urgent need for adaptive, data-driven approaches that can optimize treatment decisions over time, improving patient outcomes and resource allocation in healthcare.

Recent advances in artificial intelligence (AI) have introduced powerful algorithms for optimizing sequential decision-making processes. One promising technique is the Restless Multi-Armed Bandit (RMAB) algorithm, which has been used successfully in a variety of domains, including telecommunications and marketing. In healthcare, RMABs offer the potential to dynamically adjust treatment strategies based on individual patient responses. This project aims to apply RMAB algorithms to optimize treatment plans for chronic disease management, focusing on real-world applications in healthcare using the MIMIC-III dataset. Using this large-scale critical care dataset, our aim is to validate the performance of RMAB algorithms in dynamically adapting treatment strategies to individual patients. The main goal is to create a framework that not only improves patient outcomes but also makes a more efficient use of healthcare resources, addressing the complex and dynamic nature of chronic disease management.

2 Motivation

The idea of using Multi-Armed Bandit (MAB) algorithms in healthcare management stemmed from the implementation of MABs in other fields, such as digital marketing, where these algorithms have been successfully applied to optimize ad placements based on user engagement [7]. The core principle of balancing exploration (trying new treatments) and exploitation (using known effective treatments) is relevant not only to ad placements, but also to healthcare resource allocation. This framework provides a robust method to adapt to individual patient needs and offers the potential to optimize resource use in healthcare settings.

3 Problem Definition

Managing chronic diseases is a dynamic process that requires continuously adjusting treatment strategies based on the evolving condition of a patient. The central challenge is to determine the optimal timing and nature of interventions for each patient, considering limited healthcare resources. RMAB algorithms offer a powerful solution by treating each patient as an "arm" and learning to dynamically allocate interventions over time, thus improving long-term patient outcomes. The problem addressed in this study is how to model patient states, define optimal treatment policies, and achieve the best balance between short-term and long-term health improvements using RMABs.

4 Related Work

In recent years, the application of Restless Multi-Armed Bandit (RMAB) algorithms has gained prominence in optimizing dynamic treatment strategies for chronic disease management. A notable contribution is the work by Mate et al. [2], who introduced the Streaming Bandits framework. This framework adeptly captures the dynamic nature of patient cohorts, where individuals may join or leave the system over time. Their proposed algorithms have shown significant improvements in real-world public health applications, including tuberculosis patient monitoring and maternal healthcare intervention planning [2].

Additionally, Killian et al. [3] introduced an equitable RMAB framework to optimize resource allocation in digital health programs, specifically in managing diabetes patients with limited resources. Their work demonstrated the effectiveness of RMABs in dynamically adjusting treatment strategies based on patient response, while ensuring equitable distribution of healthcare resources. The MIMIC-III dataset, a critical care dataset, has also been extensively used for predictive modeling in healthcare [4].

Liang et al.[5] presented a Bayesian approach to online learning for contextual RMABs, termed BCoR. This method effectively models complex public health scenarios characterized by contextual factors and non-stationary environments. BCoR has demonstrated superior performance in real-world applications, such as enhancing beneficiary adherence in public health programs [5]. Furthermore, Li and Varakantham addressed the integration of fairness constraints within RMAB decision-making processes. They formalized fairness definitions and developed planning and learning methods to ensure equitable resource allocation, which is crucial in public health interventions targeting diverse populations [6].

These studies highlight the versatility of RMAB algorithms in creating adaptive, personalized treatment strategies for chronic disease management. RMAB frameworks effectively handle various healthcare scenarios, from individual patient interventions to population-level resource allocation. Their ability to address the evolving nature of chronic diseases by continuously adjusting treatment plans and incorporating contextual factors like demographics and comorbidities makes them valuable for improving both short- and long-term outcomes.

5 Proposed Methodology

The proposed methodology involves applying Restless Multi-Armed Bandit (RMAB) algorithms to optimize dynamic treatment plans for chronic disease management, using real-world data from the MIMIC-III dataset. This approach aims to personalize treatment strategies based on patient health states that evolve over time, maximizing health outcomes while efficiently using healthcare resources.

5.1 Data Preprocessing

To begin, the MIMIC-III dataset will be used to extract relevant clinical data for patients with chronic conditions such as diabetes and cardiovascular disease. Key steps in the preprocessing stage include:

Normalization: Patient data, including blood pressure, cholesterol, and glucose levels, will be standardized to maintain consistency across the dataset. State Representation: Each patient's health state will be represented as a set of clinical metrics (e.g., vital signs, laboratory results), capturing their current condition and treatment history. Reward Function: A reward structure will be defined, where positive rewards correspond to improved patient outcomes (e.g., stabilized blood pressure, lower cholesterol), and negative rewards reflect deteriorating health conditions.

5.2 Restless Multi-Armed Bandit Algorithm

The core of the proposed methodology is the RMAB algorithm, which models the restless nature of chronic disease progression. Each **treatment plan** is treated as an individual "arm" of the bandit, with the algorithm determining when and how to intervene with the most appropriate treatment based on the patient's evolving condition. The RMAB will operate as follows:

- **Patient Dynamics:** A state transition model, such as a Markov Decision Process (MDP) [1], will be used to represent the evolving health status of each patient. This model helps simulate the various health states a patient may transition through over time, influenced by the treatments administered and the disease progression.

- **Policy Learning:** The algorithm will learn optimal policies for treatment allocation by observing patient responses over time. The focus will be on balancing exploration (trying new or less common treatments) and exploitation (applying treatments known to be effective based on previous patient outcomes).

- **Personalization:** Treatment decisions will be personalized based on individual patient characteristics, incorporating factors such as age, sex, comorbidities, and previous treatment responses. The algorithm will continuously adapt the treatment plan according to the patient's current state.

6 Identified Gaps

While RMAB algorithms have shown significant potential in public health and healthcare management, there are still gaps in their application to chronic disease management at the individual patient level. Existing works have largely focused on population-level health interventions, such as monitoring vital signs or optimizing limited resources for large groups of patients. However, fewer studies have explored the personalization of dynamic treatment plans for individual patients over time.

Moreover, most RMAB applications in healthcare have emphasized resource allocation and treatment fairness, but have not fully addressed the evolving nature of chronic diseases that require continuous adaptation of treatment strategies. The identified gap lies in the need for a framework that can personalize dynamic treatment decisions, taking into account both the patient's evolving health state and the long-term impact of treatment. Additionally, there is a need for more comprehensive studies on integrating real-world datasets, such as MIMIC-III, to validate the effectiveness of RMAB-based approaches in clinical settings.

7 Conclusion

This study explores the use of Restless Multi-Armed Bandit algorithms to optimize dynamic treatment plans for chronic disease management. By leveraging the MIMIC-III dataset, we demonstrate that RMABs offer a promising approach to improving healthcare outcomes by providing adaptive, personalized treatment strategies. Future research should focus on validating these findings in clinical settings and exploring the potential for RMABs in other areas of healthcare.

References

[1] Martin L Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley-Interscience, New York, 1994.

[2] Aditya Mate, Arpita Biswas, Christoph Siebenbrunner, Susobhan Ghosh, and Milind Tambe. Efficient algorithms for finite horizon and streaming restless multi-armed bandit problems, 2022.

[3] Jackson A. Killian, Manish Jain, Yugang Jia, Jonathan Amar, Erich Huang, and Milind Tambe. Equitable restless multi-armed bandits: A general framework inspired by digital health, 2023.

[4] Alistair E W Johnson, Tom J Pollard, Lu Shen, Li-Wei H Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. MIMIC-III, a freely accessible critical care database. *Sci. Data*, 3(1):160035, May 2016.

[5] Biyonka Liang, Lily Xu, Aparna Taneja, Milind Tambe, and Lucas Janson. Context in public health for underserved communities: A bayesian approach to online restless bandits, 2025.

[6] Dexun. Li and Pradeep Varakantham. Efficient resource allocation with fairness constraints in restless multi-armed bandits. In James Cussens and Kun Zhang, editors, *Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence*, volume 180 of *Proceedings of Machine Learning Research*, pages 1158–1167. PMLR, 01–05 Aug 2022.

[7] Chenyan Jiang. Optimizing online advertising with muti-armed bandit algorithms. *Applied and Computational Engineering*, 83:52–61, 09 2024.