Comparative Study of Multi-Armed Bandit Algorithms in Clinical Trials

Wuyue Huang[1], Wenling Wang[2], Yudong Wu[3], Chuheng Xi[1,4,*]

[1]Universtiy of California San Diego, 9500 Gilman Dr, La Jolla, CA 92093
[2]Qingdao Hengxing University of Science and Technology, 588 Jiushui East Street, Qingdao, Shandong, China
[3]University of Washington, 1410 NE Campus Pkwy, Seattle, WA 98195

[4]cxi@ucsd.edu
*corresponding author

Abstract. In recent years, with the rapid development of the information age, the influence of Multi Armed Bandit Algorithms (MAB) models in clinical trials for disease prevention has been increasing. In this study, based on Python programming language, Multi-Armed Bandit Algorithms (MAB) algorithm, Upper Confidence Bound (UCB) algorithm, Adaptive Epsilon-Greedy Algorithm, and Thompson Sampling (TS) algorithms to validate the idea of preventing, controlling and predicting the occurrence of diseases. The results show that the MAB model can effectively solve various decision-making problems in clinical trials, improve the efficiency of access to medical care, save doctors 'diagnosis time, and at the same time achieve the prevention and treatment of diseases while minimising patients' pain. This study is dedicated to proposing a more effective decision-making method and verifies that the method has a wide range of applications and great potential for development today.

Keywords: Multi-Armed Bandit Algorithms, Python, Clinical Trial.

1. Introduction

According to the table1 below, HIV is a very serious disease across the United States. Now in the information age, clinical trials need to be enabled with new technologies to monitor and treat such conditions. The rapid development of the information age has markedly increased the influence and application of Multi-Armed Bandit (MAB) algorithms across a range of sectors. Initially applied in clinical settings, MAB algorithms such as Upper Confidence Bound (UCB), Adaptive Epsilon-Greedy, and Thompson Sampling (TS) have proven effective in disease prevention trials. Employed to validate methods for preventing, controlling, and predicting disease occurrences, these algorithms have facilitated significant advancements in medical care by reducing diagnostic times and improving treatment accessibility. The versatility of MAB algorithms extends beyond clinical applications, demonstrating substantial potential in various non-clinical domains.

Table 1. HIV Diagnoses in 2022 by Region in USA.

	Percentage of HIV Diagnoses (2022)	Number of HIV Diagnoses (2022)
South	52	19793
West	21	7848
Midwest	13	4891
Northwest	13	5069
US Territories and Freely Associated States	1	380

In educational settings, Rafferty et al.demonstrate how MAB approaches can dynamically adjust experimental conditions, significantly enhancing learning outcomes compared to traditional methods. This ability to respond to real-time data underscores the algorithms' adaptability and potential for broader applications, including those outside the medical field [1].

Further extending the scope of MAB algorithms, Fouché et al. introduce the Scaling Multi-Armed Bandit (S-MAB) model, designed to dynamically determine the number of arms to play while balancing reward maximisation with selection costs. Their model performs well in both static and dynamic environments, making it suitable for predictive maintenance and data stream monitoring, areas critical to the operational efficiencies in various industries [2].

Additionally, the work of Avadhanula et al.showcases the application of stochastic bandits in optimising budget allocation across multiple advertising platforms. Their approach not only outperforms common benchmarks but also emphasises the practical impact of MAB algorithms in managing complex budget optimization tasks in real-world scenarios [3]. Zhang offers a comprehensive review of real-world applications of MAB algorithms,covering areas such as healthcare, content recommendation, and machine learning optimization.

The paper emphasises the versatility and ongoing evolution of MAB algorithms, showcasing them

potential to address increasingly complex decision-making problems across various domains [4]. All the theoretical findings are corroborated by simulation experiments showing how effective the suggested algorithms are in practical settings, including clinical trials in the following parts.

Together, these examples illustrate the broad applicability and ongoing evolution of MAB algorithms. As these methods continue to address increasingly complex decision-making problems, their potential to transform various aspects of society becomes ever more apparent, highlighting their development potential in today’s rapidly advancing technological landscape.

2. Introduction to the MAB model and its benefits in clinical trial design

To address the exploration-exploitation trade-off inherent in MAB problems, we can use several algorithms:

2.1. Adaptive Epsilon-Greedy Algorithm

This algorithm uses an adaptive parameter ε to balance exploration (trying out different treatments) and exploitation (choosing the treatment with the highest observed success rate). The adapt ability of ε is crucial in clinical settings to ensure that patients receive the best possible treatment while still allowing for the discovery of potentially better options.

2.2. Upper Confidence Bound (UCB)

The UCB algorithm selects the treatment with the highest upper confidence bound on the expected reward, thus addressing the need for both exploration and exploitation. UCB is particularly useful when there is uncertainty regarding the effectiveness of treatments.

2.3. Thompson Sampling (TS)

Thompson Sampling selects treatments based on the probability that each treatment is the best, given the current data. This Bayesian approach is well-suited for clinical trials where outcomes are binary, and it naturally incorporates uncertainty into decision-making.

2.4. Contextual Bandits (LinUCB)

The LinUCB algorithm extends the UCB framework by incorporating contextual information such as patient demographics or medical history. This allows for personalised treatment decisions, improving the overall efficacy of the trials.

3. Multi-Armed Bandit Algorithms in Clinical and Non-Clinical Applications

Multi-armed bandit (MAB) algorithms have been widely studied and applied across various domains due to their effectiveness in optimising decision-making processes under uncertainty. This section provides a comprehensive review of recent experimental studies, distinguishing between clinical and non-clinical applications and includes detailed analysis for each reference.

3.1. Clinical Applications

In the clinical domain, MAB algorithms significantly enhance the design and efficacy of clinical trials. Aziz et al.explore the application of Thompson Sampling in dose-finding trials.Their study reveals that Thompson Sampling is highly effective in identifying the Maximum Tolerated Dose (MTD) and balancing efficacy and toxicity in phase I and phase I/II trials. This approach minimises the number of sub-optimal dose selections compared to traditional methods, thus improving patient safety and trial outcomes[5]. Their analysis highlights the benefits of adaptive designs over static methods, particularly in optimising clinical trial efficiency and safety. Bulucu investigates contextual MAB algorithms for personalised medicine. The study focuses on reducing the curse of dimensionality by identifying relevant patient features and treatment options. The proposed algorithms, including CMAB-RL and CGP-UCB with relevance learning, show substantial improvements in treatment outcomes for conditions such as diabetes.

This work underscores the importance of individualised treatment strategies and the potential of MAB algorithms to tailor interventions based on patient-specific data [6]. Chen et al. address safety constraints in clinical trials through their novel MAB strategies. They introduce algorithms that use doubly optimistic indices for both rewards and risks,ensuring that the number of unsafe arms played is logarithmic relative to the number of trials. This approach effectively balances reward optimization with safety, providing a robust framework for managing risks in clinical settings [7]. The study's theoretical and empirical results demonstrate the practical utility of incorporating safety considerations into MAB algorithms.Shin et al. examine the bias in sample means within MAB problems. Their analysis reveals that optimistic sampling tends to introduce negative bias, while optimistic stopping and choosing leads to positive bias. This nuanced understanding of bias is crucial for accurately estimating treatment effects and ensuring reliable decision-making in clinical trials. Their theoretical results and simulations provide valuable insights into how bias affects MAB algorithms in practice [8].

3.2. Non-Clinical Applications

In non-clinical domains, MAB algorithms are applied to a variety of complex decision-making problems. Huang et al. focus on heavy-tailed loss environments, which are prevalent in online advertising and other applications with unbounded variance in loss distributions. Their algorithms, HTINF and AdaTINF, achieve near-optimal regret bounds and are designed to perform well even with unknown distribution parameters. This work highlights the importance of adapting MAB algorithms to handle heavy-tailed losses effectively [9]. Shi et al. propose a framework for dealing with delayed feedback in MAB algorithms.Their delay-adjusted inverse propensity weighting (DAIPW) estimator provides theoretical guarantees for consistency and asymptotic normality despite feedback delays. This framework is particularly relevant for applications where feedback is not immediate, such as in e-commerce and healthcare settings, ensuring reliable inference and decision-making [10]. Bouneffouf and Rish provide a thorough survey of practical applications of MAB and contextual bandits. Their review covers diverse fields such as dynamic pricing, anomaly detection, and machine learning optimization. The paper highlights trends and future directions, underscoring the broad applicability and continued advancements in MAB algorithms [11]. Chen introduces a variant of the MAB problem requiring monotone arm sequences,applicable to dynamic pricing and clinical trials where actions must follow a monotonic order. The proposed algorithm achieves theoretical regret bounds, offering insights into handling monotonic constraints in MAB problems. This study emphasises the trade-offs involved in incorporating suchconstraints [12].

Overall, these studies illustrate the diverse applications and ongoing advancements in MAB algorithms, reflecting their significant impact on optimising decision-making across both clinical and non-clinical domains

4. Case Study and Analysis for AIDS Clinical Trial

To further validate the application of Multi-Armed Bandit (MAB) algorithms in clinical trials,

We conducted a case study using data from AIDS clinical trials. The dataset, collected in 1996, includes data from 2,139 patients and 24 variables, focusing on the effectiveness and safety of various AIDS treatments

4.1. Overview of the Dataset

The dataset used in this study provides a comprehensive view of the outcomes of monotherapy versus combination therapies in HIV-infected patients with CD4 counts between 200 to 500

cells/mm3. The key research objectives were:

- To assess the impact of different treatment regimens on disease progression.
- To understand which treatments are more effective in preventing mortality within a set timeframe

4.2. Simulation Setup and Evaluation Criteria

- Initialization: Each algorithm began with no prior knowledge of treatment efficacy to ensure unbiased learning and fair comparison.
- Decision Process: Algorithms selected treatments for new patients based on the accumulated data from previous outcomes, with each new reward informing subsequent decisions.
- Reward Definition: Rewards were based on improvements in CD4 count and survival rates, directly reflecting treatment efficacy.

4.3. Metrics for Evaluation

- Cumulative Reward: Measures the total reward gathered over all trials, indicating the effectiveness of the treatment strategies.
- Regret: Calculates the difference between the obtained rewards and the optimal rewards achievable in hindsight, assessing the learning efficiency of the algorithms.

This setup provides a comprehensive framework for evaluating how well each MAB algorithm adapts and optimises treatment choices in a simulated clinical trial environment.

4.4. Visualisation and Analysis

We employed several MAB strategies, including Thompson Sampling, Upper Confidence Bound

(UCB), Linear UCB (LinUCB), and Epsilon-Greedy. According to the result, as with many past experiments, TS and UCB are much better than ETC. However, as shown in The Exploration-Then-

Commitment (ETC) strategy demonstrated superior cumulative rewards over 2,000 trials by first exploring all options and then committing to the best one.

Firstly, for Efficiency in Clinical Trials, ETC is particularly effective in clinical trials where minimizing time spent on ineffective treatments is crucial. By committing to the best option early, ETC can lead to better patient outcomes and more efficient resource use. After that, for Clear Decision-Making, ETC's rapid ascent in the reward curve indicates its ability to quickly identify and stick with the best treatment, avoiding unnecessary trials with less effective options. Furthermore, for High-Stakes Environments, ETC's structured "explore first, then commit" approach is ideal for scenarios where quick, decisive action is needed, outperforming more adaptive but riskier strategies like Thompson Sampling or UCB.

Despite ETC's strong results, there are potential drawbacks. At first, it has Exploration Phase Dependence. Success relies on a well-executed exploration phase; if too short, it may commit to a suboptimal treatment, and if too long, it may prolong exposure to ineffective options. The second one is Rigidity, ETC's inflexibility after committing could be a disadvantage in dynamic environments where treatment effectiveness may change over time. Lastly, in Context-Specific Performance, ETC's effectiveness might not generalize across all clinical trials, as its performance can depend on trial-specific factors like the number of treatments and patient variability.

In summary, ETC shows promise but should be applied with consideration of its context and limitations in clinical trials' Figure 1, the Exploration Then-Commitment (ETC) strategy outperformed the others in cumulative reward over 2,000 trials.

This steep ascent on the plot indicates that ETC is particularly effective in efficiently identifying and committing to the most effective treatment option, leading to superior results in clinical trial settings.

4.5. Further Development in MAB Algorithms

MAB Algorithms need improvement. As MAB algorithms continue to evolve, recent advancements in algorithm design offer new possibilities for clinical trials. Future directions include:

- Integration of MAB algorithms with modern machine learning and AI techniques, such as deep learning.
- Development of hybrid models that combine MAB with other learning paradigms to enhance the robustness and adaptability of the algorithms.
- Exploration of real-time adaptation in clinical settings to better respond to ongoing trials and patient responses.

These developments are expected to significantly enhance the precision and efficacy of clinical trials, potentially transforming how new treatments are tested and implemented.

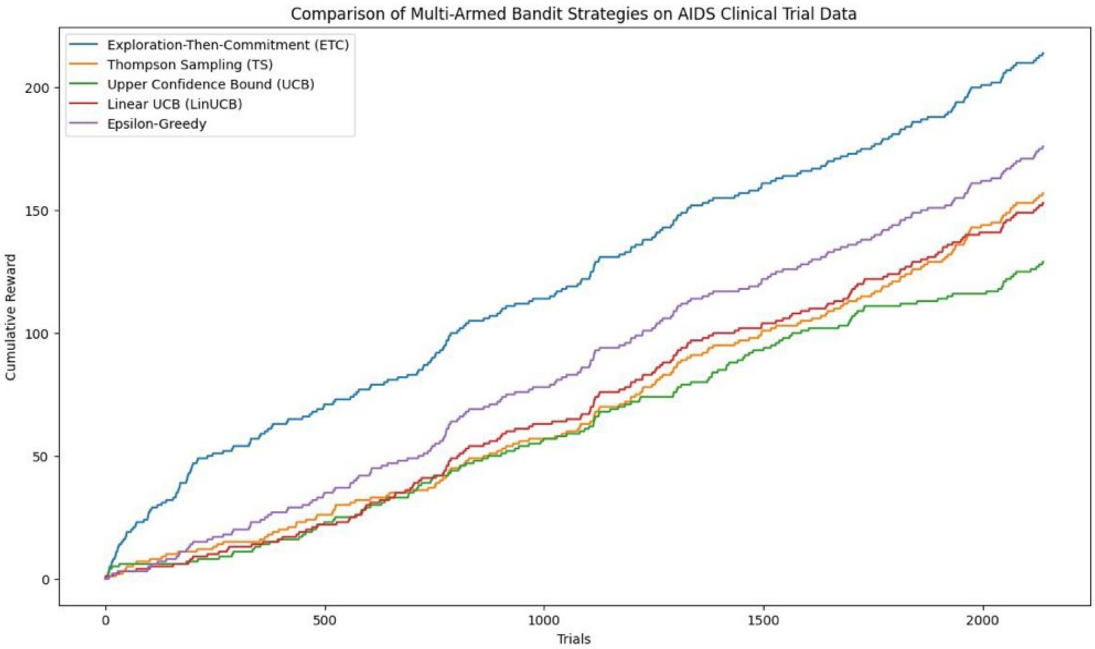

Figure 1. Comparison of MAB algorithms performance (cumulative reward) for strategies on AIDS treatment

5. Conclusion

To sum up, with the rapid development of the information age, the application of dobby slot Machine algorithms in clinical trials have attracted much attention. This study aims to compare the effect of multi-arm slot machine algorithms in clinical trials, using Thompson sampling method and multi-arm slot machine algorithms, aiming to diagnose the disease while reducing the pain of the patient, and to contribute more to the clinical trials to promote the health of the body. The results of the study show that the multi-armed slot machine algorithm, as a simple yet powerful reinforcement learning framework, helps to balance the problem between exploration and exploitation, optimise resource allocation, improve efficiency and achieve optimal solutions. The effectiveness of the algorithm in clinical trials was verified through multiple sets of tests, demonstrating its great value and positive impact in the context of the era of rapid information development. The algorithm is expected to maximise benefits and bring important development opportunities to the field of clinical trials.

This study fills the research gap in the application of MAB algorithms in clinical trials and provides new ideas and methods for future related research. Exploring the application of multi-armed slot machine algorithms in clinical trials provides new tools and methods for healthcare practitioners and researchers, which can help improve the patient treatment experience,increase the efficiency of clinical trials and promote the development of the healthcare field.

This paper focuses on solutions in different environments to delve into the decision-making process in reinforcement learning to optimise decisions with intelligent learning algorithms to maximise rewards in the long term. However, the study still needs to ensure dataset accuracy and avoid fuzzy inferences. Challenges remain in the application of algorithms that need to be further optimised to improve computational efficiency. Future research should focus on data accuracy and delve deeper into topics in this area.

Athors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Rafferty, A., Ying, H., & Williams, J. (2019). "Statistical consequences of using multi-armed bandits to conduct adaptive educational experiments". Journal of Educational Data Mining, 11(1), 47-79.

[2] Fouché, E., Komiyama, J., & Böhm, K. (2019). "Scaling multi-armed bandit algorithms". Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.

[3] Avadhanula, V. et al. (2021). "Stochastic Bandits for Multi-Platform Budget Optimization in Online Advertising". Proceedings of the Web Conference 2021. pp. 1472-1483. URL: https://dl.acm.org/doi/abs/10.1145/3442381.3450074.

[4] Zhang, Q. (2024). "Real-world Applications of Bandit Algorithms: Insights and Innovations". Transactions on Computer Science and Intelligent Systems Research, 5, 753-758.

[5] Aziz, M., Kaufmann, E., & Riviere, M.-K. (2021). "On multi-armed bandit designs for dose-finding trials". Journal of Machine Learning Research, 22(14), 1-38.

[6] Bulucu, C. (2019). "Personalizing treatments via contextual multi-armed bandits by identifying relevance". MS thesis. Bilkent Universitesi (Turkey).

[7] Chen, T., Gangrade, A., & Saligrama, V. (2022). "Strategies for safe multi-armed bandits with logarithmic regret and risk". International Conference on Machine Learning. PMLR, pp. 3198-3210.

[8] Shin, J., Ramdas, A., & Rinaldo, A. (2019). "Are sample means in multi-armed bandits positively or negatively biased?". Advances in Neural Information Processing Systems, Vol. 32.

[9] Huang, J., Dai, Y., & Huang, L. (2022). "Adaptive best-of-both-worlds algorithm for heavy-tailed multi-armed bandits". International Conference on Machine Learning. PMLR.

[10] Shi, L., Wang, J., & Wu, T. (2023). "Statistical inference on multi-armed bandits with delayed feedback". International Conference on Machine Learning. PMLR.

[11] Bouneffouf, D., & Rish, I. (2019). "A survey on practical applications of multi-armed and contextual bandits". arXiv preprint arXiv:1904.10040.

[12] Chen, N. (2021). "Multi-armed bandit requiring monotone arm sequences". Advances in Neural Information Processing Systems, Vol. 34, 16093-16103.