Introduction

Decision-making problems, where an agent must choose the best action from several alternatives at each point in time, are widely encountered in practical applications, such as when companies decide how to structure their websites to encourage shopping behavior. These problems involve a trade-off between exploration and exploitation (Lai & Robbins, 1985; Auer et al., 2002). This trade-off is typically formulated as the multi-armed bandit problem, where each possible action, or arm, has an unknown reward probability distribution, and the agent aims to maximize cumulative rewards over time by selecting the optimal arm (Bouneffouf & Rish, 2019).

In this assignment, we apply the MAB framework to a real-world dataset from Zozo, a high-traffic fashion website. Here, the arms correspond to different fashion items, and the rewards are based on user clicks on the advertisement banner. The goal is to determine which items are best to recommend and identify the most effective positions on the recommendation interface (left, center, or right) to maximize click-through rates (CTR).

To achieve this, we will first describe and analyze the dataset. Then, we will test several MAB models, including Thompson Sampling, ε-greedy, and Upper Confidence Bound (UCB), to identify the best-performing items and banner positions for maximizing clicks. Additionally, we will perform a heterogeneity analysis to examine how different user segments impact the models' performance, a batch analysis to assess how varying batch sizes affect outcomes, and a parameter analysis to improve the models.

The dataset captures user interactions with recommended fashion items and includes variables such as timestamps of impressions, item IDs, positions on the interface, and click indicator. It also contains propensity scores, which represent the probability of an item being recommended in each position, along with several user-related features such as age and gender.

Bibliography

Auer, P., Cesa-Bianchi, N., & Fischer, P. (2002). Finite-time analysis of the multiarmed bandit problem. Machine Learning, 47(3-4), 235–256. https://doi.org/10.1023/A:1013689704352
Lai, T. L., & Robbins, H. (1985). Asymptotically efficient adaptive allocation rules. Advances in Applied Mathematics, 6(1), 4-22. https://doi.org/10.1016/0196-8858(85)90002-8
Bouneffouf, D., & Rish, I. (2019). A survey on practical applications of multi-armed and contextual bandits. arXiv:1904.10040. https://doi.org/10.48550/arXiv.1904.10040