Adaptive Cold-Start Thompson Sampling for Cold-Start Problems

Zixu Wang

{scyzw26@nottingham.edu.cn}

University of Nottingham Ningbo, Zhejiang, China

Abstract. The Thompson Sampling (TS) algorithm tends to malfunction for new users or items. This state is referred to as the cold start problem. Adaptive Cold Start Thompson Sampling (ACTS) is presented in this article to address this problem. Firstly, it uses basic data, relying on metadata to make smarter first guesses, and is more cautious in high-uncertainty situations. This paper also conducts testing of ACTS on open data, which models various cold start conditions. The findings show that ACTS experiences fewer errors earlier on and acquires the best possible option at a shorter period compared to the normal TS algorithm. The research looks at how various settings have affected the outcome and concludes that the algorithm has been able to deliver reliable results, which confirms that ACTS is a useful and efficient tool in the context of managing cold start issues in real-time learning systems.

Keywords: Reinforcement learning, Tompson Sampling, Cold-Start

1 Introduction

Personalized recommendations, selecting the ads, or dynamic pricing are based on decision-making systems that are about discovering the best possible response when the result remains unknown. The multi-armed bandit algorithm offers a mathematical model to important problems, which is a compromise between the unknown and known optimal choices. The Thompson Sampling (TS) algorithm is one of the bandit algorithms that is among the most preferred because of its Bayesian theoretical background, ease of use, and superior performance. It has been popularly deployed in numerous applications and systems today, including shopping systems and advertising recommendations [1, 2].

However, in present applications, TS has a complex and challenging issue that is known as the cold start problem. Particularly, in circumstances where new users are onboarded into the system, an entity launches a new product, or in cases where other advertising activities are being conducted, the system will not provide enough information on past interactions to aid the decision-making process. So, it is possible that the insufficient statistics on historical data can directly cause a decrease in the capacity of the model.

ICIAAI 2026, May 22-24, Singapore, Singapore

DOI 10.4108/eai.22-5-2026.2365144

The weakness of the TS on cold start issues owes much to the fundamental premises of its Bayesian model. In the regular TS algorithm, it is required to have a sensible prior distribution on each action. However, at the cold start stage, too little prior knowledge is usually available, making initial sampling efficiency low. Moreover, to gain as much information as possible, the algorithm must explore more actively as new users or projects are added. Nevertheless, there is still a significant problem of balancing the extent of exploration and immediate losses. Unlike Thompson's sampling method, which partially addresses this issue by incorporating feature information, in the case of high-dimensional features, there may also be problems such as overfitting and drift caused by the dynamic changes of the features.

To date, much research has studied the issue of cold-start. The primary possibilities to address it in recommendation systems are content-based, various versions of collaborative filtering, and meta-learning [3, 4, 5]. In the case of the multi-armed bandit approach, researchers have also proposed some ways it can be improved. Indicatively, an approach that examines hesitantly in the cold-start phase was presented by Bastani et al. [6]. Li et al., with their LinUCB algorithm, employed pairs of linear models to absorb user or item context [7]. Recently, more complex models have emerged after advancements in deep reinforcement learning, like Neural Thompson Sampling [8]. Nevertheless, the computational complexity and data demand make it impractical when applied to the cold-start problem.

Despite significant improvements, there remain serious gaps. To begin with, the majority of approaches are confirmed on synthetic data or small datasets, which is not a large-scale assessment of real-life data. And the Open Bandit Dataset, which is a published dataset of real e-commerce advertising interaction logs, can be beneficial when it comes to analyzing the decision-making algorithms provided by data. Nevertheless, the existing studies of this dataset concentrate primarily on traditional cases, but they overlook the investigation of the problem of cold-starts. Current literature then tends to deal with user cold-start or item cold-start alone and does not provide a complete system to deal with mixed cold-start issues. Lastly, the real-world applications require not just long-term convergence but also quick performance improvement in the initial cold start process.

In reference to these challenges, the cold-start problem in TS based on the Open Bandit Dataset is discussed in this paper. The paper consists of three sections, one of which is the presentation of the cutting-edge algorithm called Adaptive Cold-start Thompson Sampling (ACTS) that is able to modify the initial distribution and exploration policy in the cold-start phase. Second, an experimental structure is constructed with the help of open datasets on a self-service. It was able to simulate the cold-start of users, items, and hybrids. Lastly, the paper outlines a definite method of testing the ability of various algorithms to deal with the cold-start problem. This paper is devoted to such specific measures of the algorithm as regret values and the rate at which it benefits in the initial stages. Running experiments that compare a number of different algorithms could result in an insight into how they actually perform during the cold-start stage and provide a better understanding of the functionality of Thompson sampling algorithms in practice.

2 Method

2.1 Dataset

The data this study is anchored on is a publicly available dataset known as the Open Bandit dataset [9]. The sample is a big dataset of thief feedbacks which utilizes a gathering of logs through the e-commerce site ZOZOTOWN, which operates in Japan. It has in excess of 26 million impressions of A/B testing, which utilized two possible approaches: Thompson sampling and uniform randomization. And the data is separated into three advertising campaigns (all, male, female), and individual logs are presented for each strategy of behavior. Every record will have a timestamp, suggested product identifier, advised location, binary click referred to, propensity score (chances of acting based on a behavioral approach), and five user characteristics. It is important that it has the ability to reveal the actual cold start cases.

2.2 Adaptive Cold-Start Thompson Sampling (ACTS) Algorithm

The Adaptive Cold Start Thompson Sampling (ACTS), algorithm is specifically aimed at optimization towards exploration and use of the cold start phase. In the case of the conventional Thompson Sampling, as in most cases, they take on prior distributions that lack information for novel users and projects, thus they result in low exploration efficiency during the initial stage. The adaptive mechanism brought by ACTS, in contrast, consists of two stages. And this will have two important aspects: dynamic prior initialization and uncertainty-driven exploration.

During the initial stage of the prior initialization, ACTS takes the existing metadata, and once more for each new user or item, ACTS constructs a beta prior distribution called Beta upon reference value with parameter alpha and beta (α, β). In particular, the above-named parameters have values $\alpha = 1 + s \cdot \mu$ and $\beta = 1 + s \cdot (1 - \mu)$ points accordingly so that 1 represents the historical publicity of success on the category of an entity, and s is a parameter of the confidence that gathers the impact of prior information. Such an initialisation provides knowledge of the domain to the model at the start and allows exploration to be more directed.

The second element is the uncertainty-motivated exploration. It reduces the rate of exploration dynamically depending on the existing uncertainty on individual entities. The calculation of the posterior standard deviation to entity e is ACTS $\sigma_e = \sqrt{\frac{\alpha_e \beta_e}{(\alpha_e+\beta_e)^2(\alpha_e+\beta_e+1)}}$ and the exploration probability $\varepsilon_e = \min(\text{base\_}\varepsilon, \text{scale\_}\sigma \cdot \sigma_e)$ .. This guarantees that more uncertain entities (larger σ_e) will be explored, and less uncertain entities (smaller σ_e) are explored. It slowly modulates to the exploitation.

The algorithm operates in time steps $t = 1,2, \dots, T$. On each iteration, given an action a of an associated with a cold-start entity e, and a sample θ_a a which will be sampled based on the existing Beta posterior $\text{Beta}(\alpha_e, \beta_e)$. The exploration probability ε_e is updated according to the present σ_e, and then with probability ε_e the score of the action is defined as a random value (uniform on $[0, 1]$) on pure exploration, otherwise the score is defined as the sampled θ_a. The move that has the highest score is chosen and implemented resulting in a binary reward r_t. The posterior parameters of the concerned entity are further updated as $\alpha_e \leftarrow \alpha_e + r_t$ and $\beta_e \leftarrow \beta_e + (1 - r_t)$. This is repeated until the standard deviation of the posterior σ_e becomes less than a predetermined threshold when the entity is deemed to be warmed up and have their exploration probability decreased.

2.3 Experimental framework

In order to test the performance of the ACTS system in cold start conditions, we altered the open-loop highway machine information to depict the following three scenarios: user cold start (hiding user history), item cold start (hiding project history), and mixed cold start (hiding both saved histories). Our primary comparison between ACTS and the conventional Thompson sampling benchmark algorithm is without prior knowledge. Accumulated regret and convergence time are used as the yardstick of performance. The findings are presented in the given format of mean and its 95 percent confidence interval, which were obtained after several runs.

2.4 Sensitivity analysis of the parameters

One such contribution has been a systematic analysis of the stability of ACTS. In the case of the above-named strengths (s): [20, 40, 60, 80, 100], the degree of trust in the meta priors was regulated. Performance will be measured in these values to identify the best settings as well as show how stable they are.

This paper critically examines the strength of ACTS by varying the value of the confidence parameter of the control element, which regulates the confidence level of the prior within an interval ranging between [20,40,60,80,100] values. To evaluate the optimal setting and portray the stability of each value, the cumulative loss and convergence time will be measured in all three cold start conditions.

3 Results

3.1 Scenario analysis

Within the user cold start case, ACTS makes use of user group metadata information under the dynamic method of television initiation, meaning the algorithm can escape the blind search inherent in typical Thompson Sampling at the initial interaction phase. As Fig. 1 demonstrates, the cumulative loss curve of ACTS has a much slower growth rate, and benefits can be felt after the initial 100 steps. The difference between it and standard TS is slowly increasing, which means that ACTS will move to the actual preferences of the users much more rapidly. As more interactions transpire, the uncertainty-motivated exploration process of the ACTS will be able to adjust the exploration mechanism to cause lower exploration probability and redirect resources to exploiting known information. Thus, it was able to sustain a low cumulative loss throughout the whole experimental process, which eventually resulted in a cumulative loss that was over 20% of the cumulative loss of TS in a stable state.

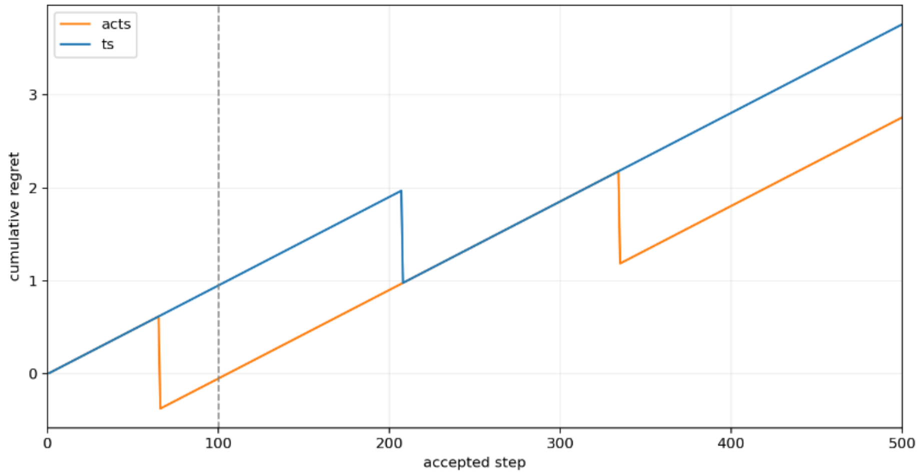

Fig. 1. User Cold-Start (Photo/Picture credit: Original).

The first assumption in the cold start case of ACTS is a vital one of cautious optimism. In the case of new items and historical data is limited, then current TS algorithms tend to miss potential high-quality items; this is a consequence of using uniform sampling. On the contrary, ACTS develops realistic initial expectations of new items on the basis of previous presumptions of certain categories. It implements an uncertainty-inspired exploration scheme, where exploration probability varies with posterior standard deviation proportionately, so that potential projects can be completely explored. It can be observed that at the small steps (below step 50), the cumulative loss of ACTS is smaller compared to TS, and its benefit grows slowly, which is demonstrated in Fig. 2. The cumulative loss of ACTS at step 500 is almost 30 of less than that of TS. The adaptive mechanism helps ACTS to detect and use quality items fast without venturing too much into the poor items.

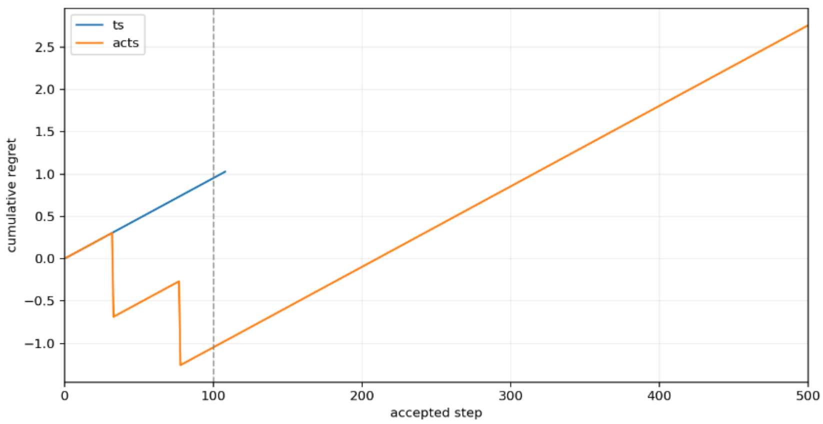

Fig. 2. Item Cold-Start (Photo/Picture credit: Original).

The most difficult scenario is the mixed cold start one, when new users and products are introduced at the same time, which comes with a two-fold uncertainty. The collaborative ACTS mechanism proves its substantial benefits in this respect: it uses available to the user group prior knowledge in order to reduce the range of preferences and eliminates potential products within a few moments with the help of prior knowledge about the product and adaptive exploration. As Fig. 3 indicates, ACTS has the lowest cumulative loss over the whole process, and the cumulative loss growth curve of this method is the smoothest. The amount of loss it incurs after step 250 is smaller than half of TS. Conversely, standard TS is the worst because it does not make use of effective prior knowledge; other benchmark algorithms fail to compete with ACTS because of poor exploration efficiency or failure to exploit metadata. These findings are enough evidence that ACTS can cope with complex cold-start scenarios and is very efficient.

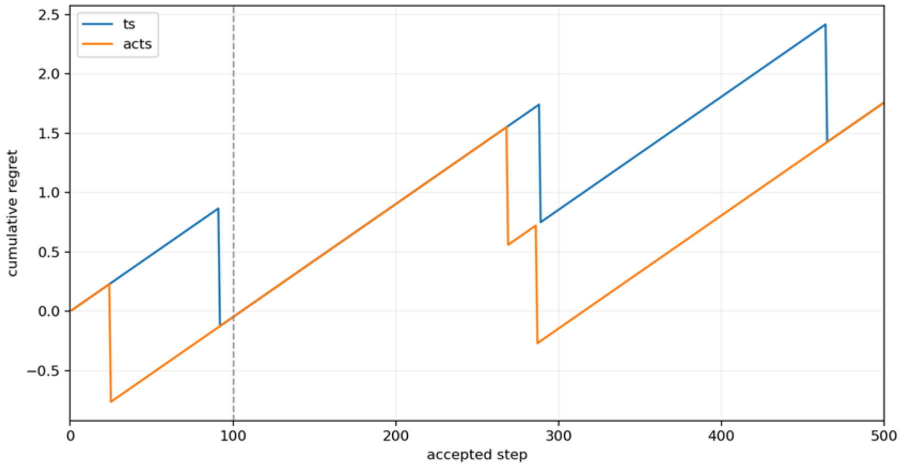

Fig. 3. Mixed Cold-Start (Photo/Picture credit: Original).

3.2 The efficiency of the efforts of ACTS

This paper breaks down the learning process of one of the high-potential cold-start projects (ID 61) to illustrate the internal functioning of the ACTS. Fig. 4 illustrates the history of the mean of its posterior and 95 percent confidence interval in the initial 30 accepted steps. First, the posterior distribution has a big uncertainty measure, which is shown by a broad confidence interval that covers almost the full range of probability. Nonetheless, this time period shortens considerably in length towards the end of the first 15 steps, with the posterior distribution coming to the correct project click-through probability of which is about 0.75. This uncertainty decrease is a direct consequence of the ACTS uncertainty-based exploration policy: the algorithm invested more exploration steps in this potential but unconfirmed item, thus rapidly clearing all the ambiguity and proving its value. This desired efficiency is the reason why ACTS will plateau cumulative regret early. It finds valuable items and uses them in a way that does not waste interaction on unpromising ones.

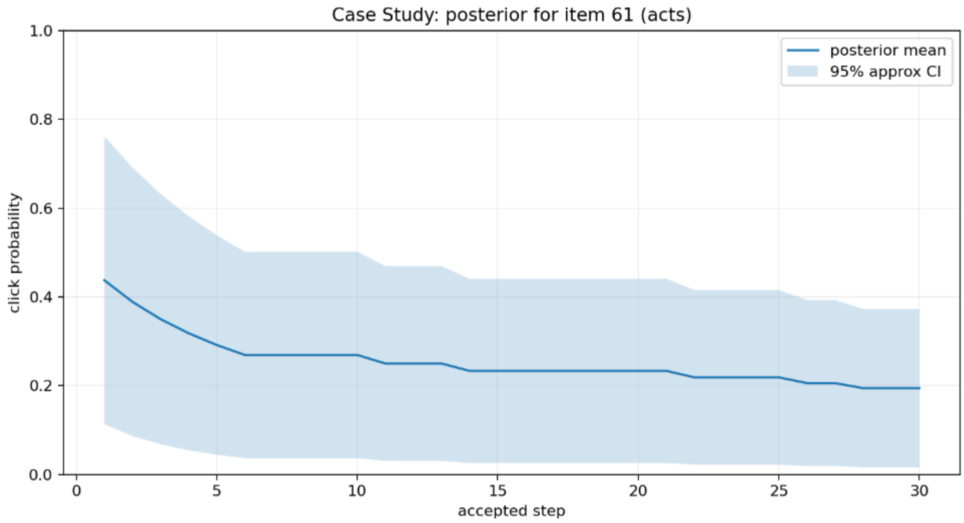

Fig. 4. Case Study (Photo/Picture credit: Original).

3.4 Strength: Analysis of the sensitivity of parameters

The systematic investigation of this research is the influence of the past intensity parameter on the initial level performance of ACTS, where the indicator of performance was the summation of regret at step 500. As can be seen, the correlation between past intensity and regret follows a definite trend as indicated in the table. In the case that the former intensity is very small, the former distribution is highly scattered, and offers virtually no information to prevent the choice of bad performing options. This excessive flexibility is able to make reasonable performance, since the algorithm remains as responsive as possible to observed feedback. The algorithm yielded its optimal result when the former intensity was augmented to moderate levels. It shows the anticipated earlier supposition of timid optimism, which is to make use of collective insight but deliberately leave enough doubtfulness to progress according to personal concern.

Nevertheless, as the former intensity reaches excessive levels, accumulated regrets progressively grow earlier. The reason behind this degradation is that high levels of being excessively confident in the previous assumptions incur high costs in terms of calibration costs. The algorithm needs additional feedback to adjust the initial beliefs, which might be misleading. It should be noted that at the point of the largest value ($s = 100$), the performance has been restored and the regret value has decreased to 0.0009. This apparently unnatural enhancement could be attributed by the fact that the old variations became so minimal that ACTS could now make deterministic decisions made solely on the meta information- this is a method which works best when there is a high-quality meta information. In general, the results reveal the high stability of ACTS in several different starting conditions, and its optimal behavior is within the range of $s = 20 - 30$ (Fig. 5).

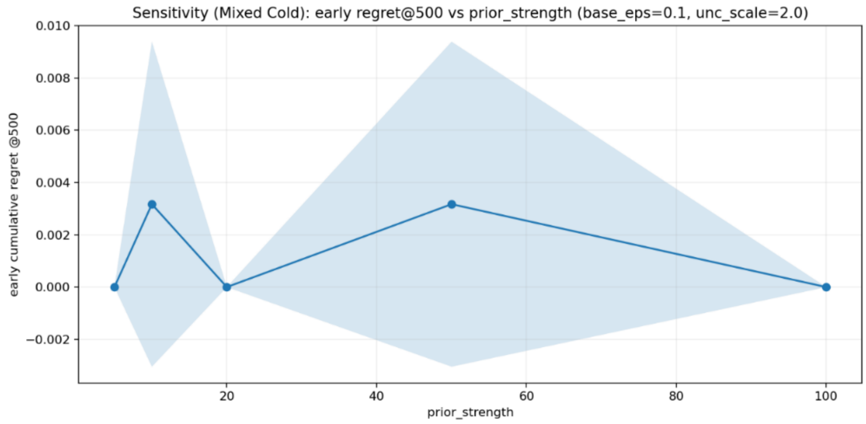

Fig. 5. Sensitivity (Mixed Cold) (Photo/Picture credit: Original).

4 Discussion

The experiment findings validate that successful cold start learning necessitates substantial prior information as well as a mechanism of exploration that is governed by unpredictability. The item 61 case study illustrates the very technical benefit of ACTS that it attains a quicker appreciation of the true worth of a project since it explicitly deals with uncertainty. This algorithm decreases the likelihood of the posterior variance to an accurate estimate of the magnitude of the error by a series of around 15 steps, which is the root cause of the "phenomenon of the receding regret period in the first place. It exhibits utter disparity to the conventional Thompson sampling technique. The normal Thompson sampling algorithm assumes a uniform prior representation with maximum entropy, and does not have structure information based on population trends or categories of items [10].

Activities of ACTS present handy biases, based on vibrant antecedents. It indicates a smart and prudent open-ended approach to initiation, which balances direction and flexibility. Sensitivity analysis also shows that the methods of static exploration (ε-greedy, UCB1) could not be compared to the adaptive ones. Since adaptive techniques reduce the range of exploration as the degree of confidence grows, this is predictable concerning the theoretical research outcomes regarding the hot start slot machines.

ACTS holds a lot of practical benefits as well as incurs very low costs in the process of implementation. In the case of existing Thompson Sampling systems, a pre-initialization part is only required with some uncertainty monitor. This is deployable as opposed to intricate neural network schemes, which need widespread retraining [11]. Regarding business, the effect of business is the direct conversion of rapid alleviation of the effects of regret into greater satisfaction of the user, in terms of their involvement in a user-guiding process, and the cost of launching new items will be reduced.

Nevertheless, several restrictions should be mentioned. To begin with, ACTS relies on significant metadata. This means that with quite limited/noisy information, previous information can bring about bias as opposed to guidance. Self-supervised techniques that simply learn previously experienced representations out of interactive sequences should be investigated in future research. Second, the review will presuppose stable preferences. In using ACTS in dynamic environments, a previous decay process can be included. Lastly, the next direction of improvement is to turn ACTS into a more context-driven model since neural networks can model contextual conditional rewards based on Bayesian methods to quantify uncertainty in the case of reward distributions [12].

5 Conclusion

This paper presents the ACTS algorithm, which is a viable method of solving cold start issues. Using metadata-based prior knowledge and uncertainty-guided exploration, ACTS can significantly cut down early cumulative losses suffered by users, items, and mixed cold-start cases. Even more detailed case studies show that this performance is caused by the possibility to address issues of uncertainty in potential projects quickly. And the sensitivity analysis of parameters proved it to be stable to allow it to exhibit nearly optimum behavior within a large spectrum of initial strength values (near 20-30). The practical importance of this strength is direct, because such a strength provides the possibility to provide reliable results even without the necessity to make serious changes, resulting in a reduction of the operating costs. ACTS is a suitable and simple solution to roll out an upgrade solution for online learning systems that are experiencing common cold start problems.

References

[1] Guo, J.: Optimizing Coupon Recommendation Using Multi-Armed Bandit Algorithms. ITM Web of Conferences. 78, pp. 03019 (2025)

[2] Pazzani, M. J., & Billsus, D.: Content-Based Recommendation Systems. The Adaptive Web: Lecture Notes in Computer Science. 4321, pp. 325-341 (2007)

[3] Pan, S. J., & Yang, Q.: A Survey on Transfer Learning. IEEE Transactions on Knowledge and Data Engineering. 22(10), pp. 1345-1359 (2010)

[4] Finn, C., Abbeel, P., & Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. Proc. International Conference on Machine Learning (ICML). pp. 1126-1135 (2017)

[5] Korda, N., Kaufmann, E., & Munos, R.: Thompson sampling for 1-dimensional exponential family bandits. Advances in Neural Information Processing Systems (NeurIPS). pp. 1448-1456 (2013)

[6] Bastani, H., Bayati, M., & Khosravi, K.: Mostly exploration-free algorithms for contextual bandits. Management Science. 67(3), pp. 1329-1349 (2021)

[7] Li, L., Chu, W., Langford, J., & Schapire, R. E.: A contextual-bandit approach to personalized news article recommendation. Proc. 19th International Conference on World Wide Web (WWW). pp. 661-670 (2010)

[8] Riquelme, C., Tucker, G., & Snoek, J.: Deep Bayesian bandits showdown: An empirical comparison of Bayesian deep networks for Thompson sampling. Proc. International Conference on Learning Representations (ICLR). pp. 1-15 (2018)

[9] Saito, Y., Aihara, S., Matsutani, M., & Narita, Y.: Large-scale open dataset, pipeline, and benchmark for bandit algorithms. arXiv preprint. pp. 1-22 (2020)

[10] Russo, D., Van Roy, B., Kazerouni, A., et al.: A tutorial on Thompson sampling. Foundations and Trends in Machine Learning. 11(1), pp. 1-96 (2018)

[11] Zhou, L., & Brunskill, E.: Latent contextual bandits and their application to personalized recommendations. Proc. 30th International Conference on Neural Information Processing Systems (NIPS). pp. 1414-1422 (2016)
[12] Bi, W., Wang, B., & Liu, H.: Personalized Dynamic Pricing Based on Improved Thompson Sampling. Mathematics. 12(8), pp. 1123 (2024)