Optimizing Click-Through Rates in Online Advertising Using Thompson Sampling

Ben Zhao*

School of Information Technology, Shanghai Jian Qiao University, Shanghai, China

Abstract. Advertisers must optimize ad selection to increase click-through rates (CTR) in an unpredictable environment as a result of the quick expansion of online advertising. Despite its effectiveness, traditional A/B testing is inefficient when it comes to dynamic user behavior and an ever-changing ad pool. This study integrates user behavior and ad context data using the TS algorithm to optimize ad selection through dynamic prediction of ad click-through rates. This allows for well-informed ad choices even when data is sparse and ad performance fluctuates. Although there were some missing pieces in the dataset utilized for the experiment, TS selected the Ad5 ad about 200 times in 200 trials. This suggests that the TS method can continue to achieve high accuracy and robustness in selection while optimizing the click-through rate in the presence of missing data. As a result, TS can adjust to complex data settings and performs better in advertising optimization. The study's findings indicate that the TS algorithm can give advertisers a practical tool for ad selection, allowing them to optimize their marketing efforts in a dynamic context.

1 Introduction

Online advertising has grown in importance as a means for companies to market their goods and services in the digital age. Businesses aim to maximize the click-through rate (CTR) through targeted ad placement in order to maximize return on investment (ROI). In addition to being a clear indicator of an advertisement's efficacy, CTR is also a crucial indicator for allocating ad funds as efficiently as possible and improving user experience. But because user behavior is dynamic and complex, ad click-through rates are unpredictable and volatile, making it difficult for marketers to anticipate CTR accurately when there are several ad options [1, 2].

Even while they can assess the efficacy of various advertisements, traditional advertising techniques like A/B testing are rapidly showing their limitations. A/B testing is insufficient for the dynamic ad pool in contemporary advertising systems since it necessitates a significant amount of data and time to provide trustworthy results [3][4]. In order to manage the complicated advertising selection environment, advertisers require a more adaptable and effective decision-making tool due to the diversification of ad locations, ad formats, and user categories [5].

* Corresponding author: 2120350@stu.gench.edu.cn

Thus, the Multi-Armed Bandit (MAB) model has emerged as a crucial technique for ad selection. Because the MAB model strikes a balance between exploration and exploitation, advertisers can use continuously updated data to refine their ad placement strategies in unpredictable contexts. Its fundamental feature is that, even in cases when data is lacking or the efficacy is questionable, advertisers can continuously investigate new ads in addition to making full use of already effective ones [4, 6]. Because of this balancing mechanism, the MAB model is especially well-suited for real-time ad selection decision-making issues. It enables marketers to dynamically modify ad displays, which progressively raises the click-through rate [7].

Thompson Sampling (TS), one of the several MAB algorithms, is well-liked because of its Bayesian inference-based features. By regularly updating the probability distribution of ad click rates, the TS method can successfully handle the problems of sparse ad data and ad performance fluctuations[6]. When dealing with complex ad selection contexts, TS is more adaptable than other algorithms (such Upper Confidence Bound, UCB). It performs particularly well when the ad pool is continuously changing or when there is a shortage of ad click data [1].

The research objective of this paper is to predict ad click-through rates using the Thompson Sampling algorithm, helping advertisers identify the most promising ads and providing a reliable basis for ad selection. By combining user behavior and ad contextual information, the TS algorithm can still make efficient decisions in situations with insufficient data or high uncertainty, thereby enhancing the overall effectiveness of ad placements. Through the trade-off between exploration and exploitation, advertisers can not only optimize ad placements but also continuously adjust strategies to adapt to changes in user behavior, thereby maintaining competitiveness in a dynamic environment[4, 8].

1.1 Literature review

Many studies have been conducted on the use of the Multi-Armed Bandit (MAB) model in ad selection and recommendation systems. With the help of the MAB model, advertisers may optimize ad selection in real time by modifying their ad placement tactics. The main problem with ad selection is that marketers frequently have to make decisions based on scant information since they do not fully comprehend the click-through rate distribution of advertising in the early stages [6]. The MAB model is especially well-suited for addressing decision-making issues in online advertising because of the trade-off between exploration and exploitation. In order to improve click-through rates, advertisers can continuously gather user click data to revise their estimates of ad efficacy[9, 10].

Using probability matching, Thompson Sampling (TS), a Bayesian technique of the MAB model, can dynamically modify ad selection. In contrast to conventional deterministic algorithms, TS can nevertheless make effective decisions in situations where the ad performance is unknown or uncertain since it updates the probability distribution of ad click-through rates depending on information that is currently known [6, 11]. For instance, the TS algorithm efficiently uses the information that is already available while investigating new ads by sampling the click-through rate of advertisements following each ad display. Because of this, the TS algorithm predicts ad click-through rates remarkably well [1].

The constant changes in the ad pool and the unpredictability of ad click-through rates make ad choosing more difficult. According to research by Tang et al., the MAB model can dynamically modify ad displays based on real-time feedback by combining user behavior data. This improves ad click-through rates without adding more ad displays [12]. This implies that advertisers can use the TS algorithm to optimize ad placement tactics in unpredictable circumstances in addition to making decisions based on known ad effectiveness.

Additionally, there are only a few ad display options available in the ad pool, and marketers have to act fast to find the best advertising from the few available displays. Chakrabarti et al.'s "mortal multi-armed bandit" model tackles the problem of short ad display lifespans [1]. The algorithm must constantly look for new advertising to sustain the growth in CTR because the efficacy of ads may decrease or cease to exist as the number of impressions rises. By means of this ongoing investigation, advertisers might enhance the overall effectiveness of ad selection while avoiding an excessive dependence on underperforming advertising.

Ad selection, however, also entails the most efficient use of advertising resources. By optimizing the ad combination, the Combinatorial Multi-Armed Bandit (CMAB) can efficiently distribute ad display resources, increasing the total click-through rate, according to research by Abhishek et al. [6]. Their research offers advertisers fresh perspectives in multi-ad choice situations by showing how to take advantage of combinatorial ad display opportunities to increase click-through rates in intricate ad deployment setups.

Additionally, research by Nishimura and colleagues highlights how well advertising recommendation systems work with the MAB model. Advertisers can maximize long-term advantages by continuously optimizing the order of ad displays using the TS algorithm by tracking user click behavior on ads [4]. Nishimura notes that by dynamically updating the probability distribution of ad click rates, the TS algorithm can assist advertisers in making the best choices in settings where ad effectiveness fluctuates regularly.

Despite the widespread use of MAB and TS algorithms in ad selection and CTR prediction, there are still several issues with the current study. First, a lot of research focuses primarily on improving ad display, with little consideration given to accurately predicting ad click-through rates [6]. Second, the efficiency of current algorithms is limited when handling the sparsity of ad click data and ad performance swings, which makes it challenging to react efficiently in dynamic contexts [1].

As a result, advertisers require a system that can optimize ad selection in real-time based on user activity in addition to a tool that can dynamically modify ad displays. The TS algorithm can provide more precise predictions throughout the ad selection process by fusing user context data with past click-through rate data of advertising, giving advertisers more informed decision support. [4, 8].

There are still several glaring flaws in the current research, despite the fact that it has made great strides in forecasting ad click-through rates. First, when dealing with complicated ad display contexts, many studies do not properly account for the influence of ad display time and frequency on user click behavior [13, 14]. Second, it is challenging to accurately forecast click-through rates in dynamic ad contexts due to the restricted effectiveness of current algorithms in handling the sparsity and incompleteness of ad click data [15]. Furthermore, while current research offers a theoretical framework for ad placement optimization, there is not enough attention paid to precisely forecasting ad effectiveness [6], particularly in intricate situations when several ads are displayed at once. There is still opportunity to enhance the present prediction models.

By using the Thompson Sampling technique, this work seeks to alleviate the limitations of previous studies. In contrast to conventional advertising optimization models, the goal of this study is to help advertisers better identify the most promising ads by forecasting ad click-through rates using the TS algorithm. Even with noisy or partial data, the TS algorithm provides more precise criteria for ad selection by constantly updating the probability distribution of ad click-through rates. In order to improve the precision of click-through rate forecasts and assist advertisers in making more informed choices about ad placement in a constantly changing advertising landscape, this study will simultaneously integrate contextual data.

2 Experiment

This section will introduce the principles, design, results, and conclusions of the experiment. At the same time, the experimental design includes data processing, experimental objectives, experimental procedures, and evaluation metrics.

2.1 Experiment principle

The Multi-Armed Bandit Problem can be resolved using the well-known Thompson Sampling technique. In order to maximize the overall return, it aims to investigate several choices while progressively approaching the best option. A Multi-Armed Bandit (MAB) technique based on Bayesian probability, Thompson Sampling (TS) offers notable benefits in issues like click-through rate analysis and ad optimization. TS is more efficient at managing uncertainty and dynamically changing surroundings than other algorithms because it uses dynamically updated Bayesian probability distributions to strike a balance between exploration and exploitation. Instead than depending on a set exploration rate, TS's exploration and exploitation balance is more intelligent than the ε -Greedy algorithm, which enables it to better adjust to changes when confronted with complicated user behaviors.By bypassing the too cautious reliance on confidence intervals that the UCB (Upper Confidence Bound) algorithm relies on, TS makes judgments using a probability matching mechanism and exhibits more robustness, particularly when the reward distribution is unstable. Furthermore, unlike the Softmax approach, which requires intricate parameter adjustment for optimization, TS has a comparatively low computational complexity, making it appropriate for processing massive amounts of data and real-time updates. Because of its wide range of applications in advertising strategies and ability to guarantee a high click-through rate while minimizing resource loss, TS is the perfect option for advertising optimization scenarios.

Thompson Sampling's basic concept is to estimate each arm's reward distribution using Bayesian inference, then sample based on the estimates to choose which arm to select next. Assuming a reward distribution, initialization, sampling, arm selection, and distribution update are some of its processes. Assume that the payment for every lever follows a specific probability distribution (such as a Bernoulli distribution or a Beta distribution) at the beginning of the reward distribution assumption phase. Set each lever's reward distribution to its initial value. One can expect that the reward of a binary reward (success/failure) lever will follow a beta distribution. The unknown chance under the circumstances of a success and b failures is frequently represented by the beta distribution. Two parameters, a and b, are initialized for each lever during the initialization step. These parameters stand for the lever's success and failure rates, respectively. Typically, a=1 and b=1 are set in the starting state, which is the same as presuming that all of the levers' reward probabilities are equal. Sample a potential reward probability from each arm's Beta distribution during the sampling phase. When a lever's reward distribution is Beta(a,b), for instance, and a potential reward probability is drawn for that lever, the value represents the "potential expected value" of that lever. The lever with the biggest sample result should be chosen. In other words, the lever with the highest reward probability will be the best choice, even though all levers contain a certain degree of uncertainty. Pull the chosen lever to see the real award after the distribution is updated. Increase the value of an if pulling the lever results in success; increase the value of b if it results in failure. It's like improving the estimation of the leverage reward distribution through observation. Lastly, continue the previous steps until convergence is established or the predetermined number of times is reached.

Thompson Exploration and exploitation are naturally balanced by sampling. A wider variance in the Beta distribution of an arm is the result of increased uncertainty during exploration when there is less data available for that arm. This arm will be investigated more often since random sampling may produce dramatic results. When using it, the sampling results progressively stabilize around the expected value, the beta distribution variance of a particular lever gradually reduces as the data for that lever gradually accumulates, and the lever is used more frequently. The Bayesian inference framework, which Thompson Sampling employs, automatically takes into account past knowledge and continuously modifies distributions to accommodate evolving data.

2.2 Experimental setup

2.2.1 Data processing

These data exist in the form of CSV files, sourced from the Kaggle website, and include ad click data from a binary reward problem.The CSV file contains the click status of 10 ads, with clicks represented by 0 or 1, totaling 10,000 test data entries. During the data processing, the CSV file is first read, and the first ten columns (named Ad1 to Ad10) represent the click status of ten ads. Each column displays binary data, where 0 indicates no click and 1 indicates a click. To determine the initial click-through rate for each ad, divide the total number of clicks by the number of times the ad was displayed. No additional threshold processing is needed because the initial data is binary, consisting of 0s and 1s, which already meets the requirements for Thompson sampling. Finally, the Thompson sampling model performs posterior updates based on the observed data and uses the Beta distribution to describe the distribution of the ad click-through rate.

2.2.2 Experimental purpose

The goal of this experiment is to maximize click-through rate (CTR) and minimize ad budget waste in order to optimize ad selection using the Thompson Sampling approach. The Thompson Sampling technique allows for dynamic ad placement modifications by continuously updating its probability estimations based on real-time analysis of the click-through rates of various ads. This guarantees a more effective use of resources by allocating a larger portion of the budget to advertisements that are more likely to result in clicks. This approach minimizes financial waste by lowering the quantity of ineffective ad displays while simultaneously increasing the overall efficacy of advertising campaigns. By raising the click-through rate, this technique not only produces better results right away but also increases the return on investment, which has a long-term effect. This approach helps to optimize the efficacy of the advertising budget by making more accurate and data-driven decisions and making sure that money is spent on advertisements that have the best chance of drawing in consumers and generating conversions.

2.2.3 Experimental procedure

First, choose the dataset by looking for and downloading datasets from Kaggle that are connected to ad click-through rates. Typically, the data uses 0 and 1 to indicate if an advertisement was clicked.

The second step is to load and examine the data, preprocess it, look for outliers or missing values, and guarantee its quality. Verify that the data satisfies the TS algorithm's requirements. If not, the data must be converted into binary form via threshold processing.

In the third step, set up the model and the beta distribution. Then, utilize this method to choose the best advertisement. Take a sample from each advertisement's beta distribution in each experiment round, then choose the one with the highest value to be displayed. Adjust the advertisement's beta distribution parameters according to whether or not it was clicked.

2.2.4 Evaluation indicators

When evaluating advertisements, use these three metrics: waste amount, click-through rate, and beta distribution.

The amount that is wasted is the part of the advertising budget that does not produce effective clicks or conversions, or the costs incurred when advertisements are shown but do not bring in money. The amount wasted can be used to gauge how cost-effective the advertisement was. Low advertising efficiency is indicated by a high amount wasted, which could use up the money without providing significant returns. Analyzing the amount wasted allows for the identification of inefficient advertisements, which improves budget allocation and the ads' return on investment (ROI), hence lowering wasteful spending.

The percentage of times an advertisement is clicked as opposed to how many times it is presented is known as the click-through rate, and it typically indicates how engaged and attentive users are to the advertisement. The primary indicator of an advertisement's efficacy, the click-through rate shows how appealing it is to the intended audience. While a low click-through rate can indicate that the ad's location or content needs to be adjusted, a high click-through rate shows that the advertisement can effectively draw users. CTR serves as a foundation for further ad optimization by assisting marketers in understanding the efficacy of their advertisements.

For Bayesian updating of an advertisement's success likelihood in uncertain situations, the beta distribution is a probability distribution. The Beta distribution is used in advertising to approximate the click-through rate distribution of the ad by continuously adjusting based on click and impression data. Based on the quantity of clicks and non-clicks on the advertisement, its parameters are modified dynamically. When the success of the ads is still unclear, the Beta distribution allows for real-time adjustment of the display priority of ads, dynamically balancing ad exploration and exploitation.

2.3 Experimental results

Different beta distributions for different advertisements can be found in Figure 1. The TS method utilizes this to determine the probability distribution of each ad's success rate. The graph shows that, following several trials, Ad5 is concentrated around 0.2, suggesting that it is the ad most likely to generate the highest return.

The optimal number of incentives each advertisement received in 200 trials is shown in Figure 2. Ad5 received nearly all of the awards among them, whereas Ad4 and Ad10 and Ad9 received nearly none. This shows that Ad5 is selected as the best advertisement with the highest frequency while the algorithm is running.

The waste and click rate results for each advertisement are shown in Table 1, with Ad5 having the most waste and Ad6 having the least. Despite possibly having the biggest rewards, Ad5's high click-through rate and high exposure have led to increased placement costs and waste, as evidenced by its high waste.

Table 1. Ad waste and click-through rate.

Ad	Waste Amount($)	Click Rate (%)
Ad1	18.052	0.1703
Ad2	16.317	0.1295

Ad3	4.077	0.0728
Ad4	12.797	0.1196
Ad5	1717.254	0.2695
Ad6	0.403	0.0126
Ad7	6.672	0.1112
Ad8	101.414	0.2091
Ad9	4.379	0.0952
Ad10	1.663	0.0489

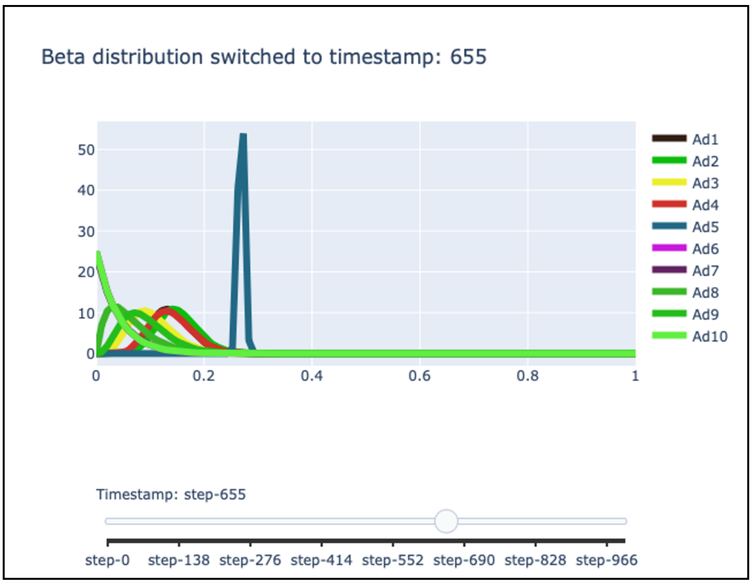

Fig. 1. Beta Distributions of Ads' Success Probabilities Over Time.

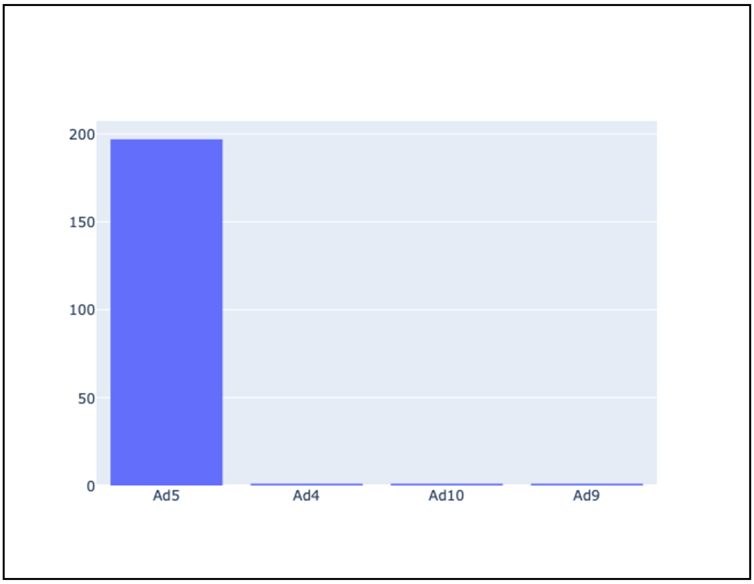

Fig. 2. Number of Selections for Each Ad.

2.4 Experimental conclusion

The multi-armed bandit problem is resolved using the heuristic approach known as Thompson Sampling. By calculating the success probability of each ad (or "arm") at each time step, it chooses the one that can provide the most reward. In the "exploration" phase, it first creates an initial estimate by examining the performance of every ad; in the "exploitation" phase, it progressively chooses more of the top-performing ads. This makes sure that plenty of new ad options are explored in advertising campaigns without sacrificing ads that have already done successfully. The system eventually focuses on the top-performing ads by updating the Bayesian posterior distribution about the ad's success rate after each ad click feedback. Through feedback from clicks and non-clicks, the TS algorithm progressively improved each advertisement's selection probability during the experiment, concluding that Ad5 had the best chance of producing large returns. Its return rate was really excellent, even if its waste cost was higher. The other advertisements perform poorly, particularly Ads 4, 10, and 9, which have low success and click-through rates. In real-world applications, raising the overall return on ad investment might be more successful if these ads are better optimized or if Ad5 is made more cost-effective.

3 Conclusions

Although the dataset utilized in this study has some missing data, the findings of applying the Thompson Sampling (TS) algorithm to the analysis of the ad click data still follow the dataset's data distribution. This suggests that one of TS's main advantages is its capacity to generate somewhat correct advertising decisions even when there is minimal or missing data. Through probability matching with sparse feedback data, TS's Bayesian update process optimizes ad selection by updating the distribution of ad click rates. This feature prevents resource waste from blind exploration, making TS more effective than other algorithms in the early phases when data is scarce. Because the TS method can maximize ad click-through rates, particularly when there is a lack of ad click data, it has been shown to perform better than other popular multi-armed bandit (MAB) algorithms in terms of ad selection accuracy. Furthermore, the TS algorithm has increased the efficacy of ad placements by fusing contextual information with user behavior. These findings suggest that in the complicated world of internet advertising, the TS algorithm can provide advertisers more intelligent and successful ad selections.

Regarding the implications for future research, this work closes several gaps in the prediction of ad click-through rates, specifically addressing the difficulties brought on by insufficient data and dynamic changes in the ad pool. Researchers studying alternative ad recommendation systems can use this study as a reference because it shows how the TS algorithm can choose advertising in real-time settings. According to research, the TS algorithm can greatly increase the effectiveness of ad placement decision-making, particularly in situations where striking a balance between exploration and exploitation is required. This provides a strong theoretical basis for the development of future advertising systems.

This study does, however, have certain drawbacks. First off, while the TS algorithm does a good job of managing variations in ad click-through rates, it may not function as effectively in settings with very sparse data or ad pools that are very dynamic. Furthermore, the performance of the TS algorithm in situations where many ads are delivered simultaneously was not thoroughly examined in this study. In order to handle more complicated ad deployment scenarios and optimize ad display methods for many ad slots, future research could investigate how to combine the TS algorithm with additional machine learning techniques.

References

1. D. Chakrabarti, R. Kumar, F. Radlinski, and E. Upfal, Mortal multi-armed bandits, Adv. Neural Inf. Process. Syst., **21**, (2008).
2. Y. Shao and Z. Fang, On multi-armed bandit with impatient arms, In Forty-first International Conference on Machine Learning.
3. S. Zhang, Utilizing reinforcement learning bandit algorithms in advertising optimization, Highlights in Science, Engineering and Technology, **94**, 195-200 (2024).
4. Y. Nishimura, Ad recommender system analysis by the multi-armed bandit problem.
5. A. Dimitrova, Ad campaign optimization—a targeted contextual bandit approach, Master's thesis, ETH Zurich, Department of Computer Science (2018).
6. K. Abhishek, Design and analysis of algorithms for combinatorial multi-arm bandit problems under complex environments, Doctoral dissertation, International Institute of Information Technology Hyderabad (2021).
7. K. Hu, Comprehensive introduction and analysis of the UCB algorithm in multi-armed bandit problems, Highlights in Science, Engineering and Technology, **94**, 247-250 (2024).
8. O. A. Hanna, L. Yang, and C. Fragouli, Solving multi-arm bandit using a few bits of communication, In International Conference on Artificial Intelligence and Statistics, pp. 11215-11236, PMLR (2022).
9. C. Zeng, Q. Wang, S. Mokhtari, and T. Li, Online context-aware recommendation with time varying multi-armed bandit, In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 2025-2034 (2016).

10. Y. Zhao, N. Behari, E. Hughes, E. Zhang, D. Nagaraj, K. Tuyls, and M. Tambe, Towards a pretrained model for restless bandits via multi-arm generalization, In Proceedings of IJCAI, pp. 321-329 (2024).
11. D. N. Hill, H. Nassif, Y. Liu, A. Iyer, and S. V. N. Vishwanathan, An efficient bandit algorithm for realtime multivariate optimization, In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1813-1821 (2017).
12. L. Tang, R. Rosales, A. Singh, and D. Agarwal, Automatic ad format selection via contextual bandits, In Proceedings of the 22nd ACM International Conference on Information & Knowledge Management, pp. 1587-1594 (2013).
13. D. Bouneffouf and I. Rish, A survey on practical applications of multi-armed and contextual bandits, arXiv preprint arXiv:1904.10040 (2019).
14. J. Zuo and C. Joe-Wong, Combinatorial multi-armed bandits for resource allocation, In 2021 55th Annual Conference on Information Sciences and Systems (CISS), pp. 1-4, IEEE (2021).
15. A. Sankararaman, A. Ganesh, and S. Shakkottai, Social learning in multi agent multi armed bandits, Proceedings of the ACM on Measurement and Analysis of Computing Systems, **3**(3), 1-35 (2019).