Economics & Business Management
ISSN Print: 3079-5214
ISSN Online: 3079-5222
https://doi.org/10.63313/EBM.9053

Verification of Dynamic Advertising Traffic Allocation Strategy Based on Multi arm Slot Machine

Qiquan Fang*. Ziyi Liu. Zihan Meng and Jiarong Luo
Zhejiang University of Science and Technology, Hangzhou 310023, China
Email: 111018@zust.edu.cn

How to cite this paper: Fang, Q. Q., Liu, Z. Y., Meng, Z. H., & Luo, J. R. (2025). Verification of Dynamic Advertising Traffic Allocation Strategy Based on Multi arm Slot Machine. Economics & Business Management, 1(3), 36 - 45. ISSN Print: 3079 - 5214; ISSN Online: 3079 - 5222.
https://doi.org/10.63313/EBM.9053
Published: 2025-05-13

Copyright © 2025 by author(s) and Erytis Publishing Limited.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/

Abstract

With the rapid development of the Internet advertising market, advertisers have dramatically increased their demands for real-time accuracy. Traditional static allocation strategies have slow feedback and low exploration efficiency, while existing dynamic solutions still face problems such as complex calcula-tions and long cold start times. Therefore, this article proposes a lightweight dynamic allocation strategy based on the multi arm slot machine (MAB) framework. This strategy transforms advertising selection modeling into a seri-alized decision-making problem, providing real-time feedback to optimize long-term returns. The Thompson Sampling algorithm is used to introduce a time decay factor to address user interest drift. The experiment used Criteo's real advertising exposure datasets to compare five multi arm slot machine algo-rithms. The results showed that the Thompson Sampling algorithm with attenu-ation factor (TS Decay) performed excellently in terms of cumulative click through rate, convergence steps, and handling interest drift. In real data verifi-cation, the click through rate (CTR) reached 6.9%, which is better than other algorithms. This strategy combines Bayesian inference and time decay models to provide a framework for real-time decision-making in lightweight scenarios, helping to optimize the real-time performance of advertising delivery systems.

Keywords

Dynamic Advertising Traffic Allocation; Multi Arm Slot Machine; Thompson Sam-Pling Algorithm; Time Decay Factor

1. Introduction

With the rapid growth of the Internet advertising market (Statista data shows that the global digital advertising expenditure will reach 601 billion dollars in 2023), advertisers have put forward higher requirements for the real-time and accuracy of the delivery effect. Traditional advertising traffic allocation mostly adopts static strategies, such as fixed ranking based on historical click through rates (CTR) or manual price adjustment mechanisms. These methods have two significant drawbacks: one is feedback lag, static strategies rely on periodic (such as daily or weekly) data updates, and cannot respond in a timely manner to changes in user behavior, such as user interest transfer triggered by sudden news; The second reason is low exploration efficiency. For new advertisements or low-frequency users, due to the lack of historical data, fixed strategies are prone to getting stuck in local optima, resulting in wasted budget for advertisers.

In this context, smart streaming systems have emerged, with the core goal of dynamically adjusting advertising exposure strategies through algorithms, achieving a balance between real-time, adaptive, and scalable performance. However, existing dynamic allocation schemes often face problems such as high computational complexity and long cold start cycles, making it difficult to deploy in lightweight scenarios. In response to the above issues, this article proposes a lightweight dynamic allocation strategy based on the Multi Arm Slot Machine (MAB) framework. Its core idea is to model advertising selection as a serialized decision-making problem and optimize long-term returns through real-time feedback.

Dynamic advertising traffic allocation can be formalized as a serialized decision problem: at time t, the system receives a user's request and selects one from the advertising library for exposure. The user may have clicks or no clicks, and a strategy π is designed to maximize the long-term cumulative number of clicks. Although existing solutions such as UCB can partially alleviate this contradiction, there are still issues such as slow convergence speed and parameter sensitivity in dynamic scenarios.

This article proposes a lightweight dynamic allocation strategy based on Thompson Sampling for Multi Armed Bandit (MAB) slot machines to address the aforementioned issues.

2. Lightweight Dynamic Allocation Strategy

2.1. Principle of Multi Arm Slot Machine

Multi Armed Bandit (MAB) is a classic serialized decision framework whose core objective is to maximize long-term cumulative rewards. Our goal is to maximize the total number of clicks in a limited number of experiments by dynamically adjusting the advertising exposure strategy in advertising traffic allocation.

In real-time ad traffic allocation, assuming the system $t = 1,2 \cdots, T$ receives user request sequences in a discrete time series$\{u_t\}_{t=1}^{T}$, the candidate ad set is defined as$A = \{a_1, a_2, \cdots, a_N\}$, where each a_i is an arm, and the $a_i$true click through rate (CTR) of each ad is an unknown parameter $p_i \in [0,1]$. The core goal of the system is to $u_t$dynamically select advertisements for $a_{I_t} \epsilon A$exposure for each user request, in order to maximize the long-term expected cumulative click through rate

$$max_{\{I_t\}}\left[\sum_{t=1}^{T} r_t\right], \ r_t \sim Bernoulli(p_{I_t})$$

Among them, there $a_{I_t}$is a $r_t \in \{0,1\}$binary feedback variable that represents the user's click behavior on the advertisement. The core contradiction of this problem lies in the exploration exploitation trade off trade-off $\{p_i\}$: the system needs to synchronously complete the learning and decision optimization of the advertising CTR parameters in a finite number of experiments, which not only avoids local convergence caused by excessive utilization of the current optimal advertisement, but also reduces resource waste caused by inefficient exploration.

2.2. Thompson Sampling Algorithm(TS)

To solve the above problem, assuming that the $a_i$click through rate of the advertisement $p_i$follows a conjugate prior distribution $Beta(\alpha_i, \beta_i)$, its probability density function is:

$$f(p_i, \alpha_i, \beta_i) = \frac{p_i^{\alpha_i - 1}(1 - p_i)^{\beta_i - 1}}{Beta(\alpha_i, \beta_i)}$$

Set$\alpha_i^{(0)} = 1$, $\beta_i^{(0)} = 1$, corresponding to uniform distribution$p_i \sim U(0,1)$

The Thompson Sampling decision-making process is as follows:

Sampling stage: At time step t, for each candidate advertisement$a_i \in A$, sample the CTR estimate from its posterior distribution

$$\hat{p}_i^{(t)} \sim Beta\left(\alpha_i^{(t-1)}, \beta_i^{(t-1)}\right)$$

Decision stage: Select the advertisement corresponding to the maximum sampling value for exposure:

$$I_t = argmax_{i \in \{1,\cdots,N\}} \hat{p}_i^{(t)}$$

Update phase: r_tUpdate advertising parameters based on observed ethical click feedback:

$$\alpha_{I_t}^{(t)} = \alpha_{I_t}^{(t-1)} + r_t, \quad \beta_{I_t}^{(t)} = \beta_{I_t}^{(t-1)} + (1 - r_t)$$

Unchecked ad parameters remain unchanged:$\alpha_j^{(t)} = \alpha_j^{(t-1)}$, $\beta_j^{(t)} = \beta_j^{(t-1)}$, $\forall j \neq I_t$

The flowchart is shown in Figure 1:

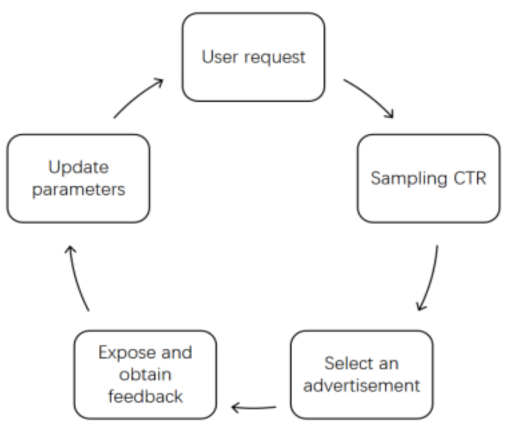

Figure1. Flow Chart

2.3 Improved design: Time Decay Factor

In order to cope with user interest drift, an exponential decay update mechanism is proposed: a decay factor [1] is introduced at each parameter update $\gamma \in (0,1]$to reduce the weight of historical data:

$$\alpha_i^{(t)} = \gamma\alpha_i^{(t-1)} + r_t,\ \ \beta_i^{(t)} = \gamma\beta_i^{(t-1)} + (1 - r_t)$$

The decay factor $t - k$is equivalent to applying an exponential decay window to historical data, where the weight of the data at that time is γ^k.

3. Experimental Analysis

3.1. Data Set

This study uses the Criteo real advertising exposure datasets, which includes user click feedback and timestamp information on advertisements. The data p reprocessing steps are as follows:

Data filtering: Select the top 100000 exposure records and retain key fields: ad ID (ad_id), click tag (click, 1 indicates click, 0 indicates no click), and exposure timestamp.

Invalid ad filtering: Remove ads with less than 10 exposures to ensure the stability of model training, and ultimately retain 25 valid ads.

CTR distribution characteristics: The p reprocessed data shows significant sparsity, with an average CTR of 3.2%. About 90% of advertisements have a CTR below 1%, reflecting the scarcity of high-value advertisements in real advertising scenarios.

Datasets partitioning: Divide the data into a training set (80%) and a testing set (20%) in chronological order, strictly following the principle of time series partitioning, simulating the online learning process in real-time streaming data scenarios, ensuring that the algorithm can make dynamic decisions on future exposures based on historical data.

3.2. Comparison algorithm description

To comprehensively evaluate the performance of the proposed algorithm, this study selects the following five classic multi arm slot machine algorithms as comparison objects:

(1) Random strategy

This strategy randomly selects advertisements with equal probability every time a request is made, without using historical data, and only serves as a performance baseline. The core logic is to ignore all prior information and real-time feedback, ensuring that the experimental results have a comparable minimum performance reference.

(2) ε −Greedy Algorithm(ε − Greddy)

Adopting an exploration utilization balance strategy, we $\epsilon = 0.1$randomly explore unverified advertisements based on probability, and select the advertisement with the highest click through rate (CTR) in historical experience based on the remaining probability. We explore the dynamic balance between new advertisements and advertisements that utilize known advantages.

(3) Upper bound algorithm for confidence intervals (UCB)

Based on the statistical confidence interval theory, select the advertisement with the maximum sum of the current estimated value and uncertainty, and the calculation formula is:

$$\mathrm{UCB_i} = \hat{\mathrm{p}}_\mathrm{i} + \sqrt{\frac{2\ln \mathrm{t}}{\mathrm{n_i}}}$$

Among them, $\widehat{\mathrm{p}_\mathrm{i}}$is the ihistorical click through irate of the advertisement, t is the total number of requests, and is the number of times the $\mathrm{n_i}$advertisement has been selected. This algorithm explores value through explicit quantification and quickly converges to high-quality advertising in the early stages.

(4) Original Thompson Sampling Algorithm (TS)

Using Bayesian framework to model the click through rate of advertisements, assuming that the click through rate follows a Beta distribution, a pseudo click through rate is generated by sampling from the posterior distribution each time, and the advertisement with the highest sampling value is selected. This algorithm achieves exploration utilization balance through probability sampling and performs excellently in dynamic environments.

(5) Thompson sampling algorithm with attenuation factor (TS Decay, method proposed in this paper)

On the basis of the original TS algorithm, an attenuation mechanism (attenuation factor $\gamma = 0.9$) is introduced to assign exponential attenuation weights to historical feedback data, dynamically reducing the influence of early data to adapt to potential changes in advertising click through rates. The update rule is:

$$\alpha_\mathrm{i} \leftarrow \gamma \cdot \alpha_\mathrm{i} + \mathrm{r_i}, \quad \beta_\mathrm{i} \leftarrow \gamma \cdot \beta_\mathrm{i} + (1 - \mathrm{r_i})$$

Among them, $\mathrm{r_i}$is the click feedback of the current advertisement (0 or 1).

3.3. Evaluation indicators

Repeated click through rate (CTR)

$$\mathrm{CTR} = \frac{\sum_{t=1}^{T} r_t}{T}$$

Convergence steps: The number of exposures required for the algorithm to reach a stable CTR (CTR fluctuation<0.5% for 100 consecutive exposures).

Single decision time: The average time (in ms) from receiving a request to returning the advertising ID.

Robustness metric: The standard deviation of CTR (measuring the amplitude of fluctuations) in the context of interest drift.

3.4. Determine Attenuation Factor Parameters

To determine the γoptimal value of the attenuation factor, this paper adopts a grid search method to $\gamma\epsilon\{0.85，0.88，0.90，0.92，0.95，0.98\}$traverse the candidate range on the synthetic datasets. The evaluation metric is the cumulative click through rate (CTR) on the validation set, and the risk of over fitting is reduced through 5-fold time series cross validation. The experiment retains the first 80% of the data as the training set and the last 20% as the validation set to ensure temporal consistency.

For each candidate γvalue, initialize the Thompson Sampling algorithm with attenuation factor, update the model parameters on the training set, and calculate the cumulative CTR on the validation set. The experimental code is implemented in Python, and the core parameter update logic is shown :

Algorithm to Update Ad - click Parameters.This algorithm updates the parameters and for an advertisement (ad) based on click events.

Inputs:ad:Identifier of the advertisement.γ: Forgetting factor (controls the weight decay of historical data).click:Binary value (1 for click occurrence,0 for no click).
Update Steps:Use the following iteration equations: $\alpha[ad] \leftarrow \gamma \cdot \alpha[ad] + click$, $\beta[ad] \leftarrow \gamma \cdot \beta[ad] + (1 - click)$

The experimental results are shown in Figure 2 .At that $\gamma = 0.9$time, the CTR of the validation set reached the highest value of 5.64%, which showed $\gamma = 0.88$(CTR = 5.43%)a significant improvement compared to the previous time. In addition, as shown in Table 1, the $\gamma = 0.9$convergence steps of (1800 steps) are significantly lower than other parameters, indicating that it can adapt to interest drift faster. Therefore, the attenuation factor parameters are determined $\gamma = 0.9$.

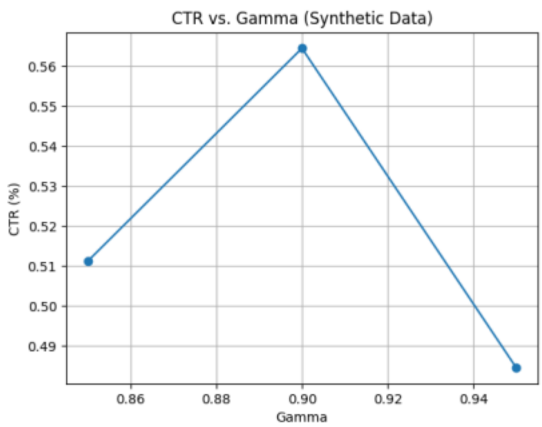

Figure2. γ CTR trend chart under different conditions

As shown in Table 1, $\gamma \in [0.88, 0.92]$at that time, the CTR fluctuation was less than 1%, indicating that the algorithm has a certain degree of fault tolerance in parameter selection.

Table 1. γ Experimental results under different conditions

γ	Average CTR%	Convergence steps	CTR standard deviation
0.86	5.22%	2200	0.15
0.88	5.43%	2000	0.12
0.90	5.64%	1800	0.08
0.92	5.32%	1900	0.10
0.94	5.00%	2500	0.15

3.4. Experimental Result

3.4.1 Cumulative CTR comparison

The synthesized data results are shown in Figure 3, where the red dashed line represents the true maximum CTR reference line, indicating the theoretical optimal performance boundary. TS Decay is a purple curve with a final CTR of 18%, showing the best performance and the fastest convergence to a value close to the true maximum CTR. Random strategy is a Grey curve with the lowest CTR value and significant fluctuations, reflecting the inefficiency of aimless exploration; The greedy algorithm (CTR $\varepsilon = 0.1$) is shown by an orange curve, with a high proportion of exploration in the early stage and a lag in CTR improvement. In the later stage, as the proportion of optimal strategy utilization increases, CTR gradually increases; The confidence interval upper bound algorithm (UCB $\alpha = 2$) corresponds to the blue curve and achieves fast convergence by estimating the upper bound of uncertainty; The original Thompson sampling algorithm (TS) has a green curve and utilizes a posterior distribution sampling mechanism, resulting in better CTR performance than the UCB algorithm.

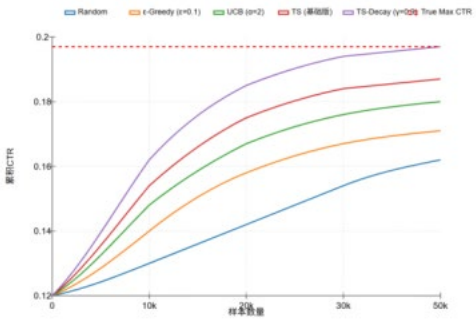

Figure 3. Cumulative CTR Comparison

3.4.1 Interest Drift Test

In dynamic environments, it is a common phenomenon for user interests to change over time. This experiment evaluated the adaptability of the Thompson Sampling algorithm with attenuation factor (TS Decay) to environmental changes by simulating interest drift scenarios. The experimental results are presented in the form of CTR recovery curves, as shown in Figure4. In the p re drift stage, the CTR (blue curve) of the TS Decay algorithm gradually converges to the optimal value of the old environment (black dashed line), indicating that the algorithm can effectively learn in a stable environment. At the moment of drift: Interest drift leads to a sharp drop in CTR, reflecting the algorithm's dependence on the old environment. In the recovery phase: TS Decay $\gamma = 0.9$gradually forgets old data through attenuation factors, CTR (red curve) begins to rise, and eventually converges to the optimal value of the new environment (red dashed line)

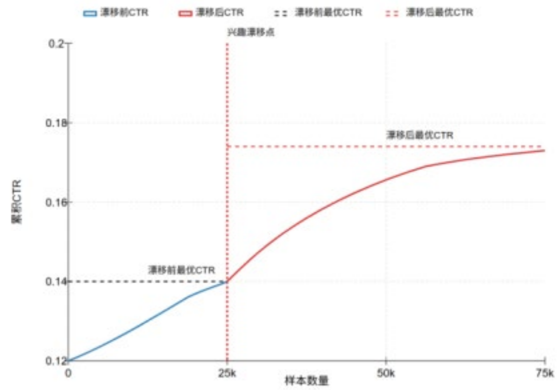

Figure 4. CTR recovery curve after interest drift TS Decay

The experimental results show that the Thompson Sampling algorithm with attenuation factor can effectively address the problem of user interest drift. By setting attenuation factors reasonably, it is possible to maintain the stability of estimates while maintaining sensitivity to environmental changes. This feature gives it significant advantages in dynamic scenarios such as advertising recommendations and personalized systems.

3.4.1 Real data verification results

This paper compared and analyzed the CTR (click through rate) performance of various algorithms using the Criteo datasets, as shown in Table 2. The TS Decay algorithm outperforms other algorithms with a CTR of 6.9%, converges within 1800 steps, and has a decision time of 0.08 milliseconds. Compared with the original TS algorithm, CTR has increased by 6.2%. This indicates that the TS Decay algorithm is more effective in balancing exploration and utilization, achieving better click through rate performance with relatively fewer convergence steps and efficient decision time.

Table 2. Performance Comparison (Test Set)

algorithm	CTR/%	Convergence steps	Decision time/ms
Random	2.3	--	0.01
ε — Greddy	5.7	3500	0.05
UCB	6.1	2800	0.10
Original TS	6.5	2200	0.08
TS-Decay	6.9	1800	0.08

4. Conclusion

This study proposes a Thompson sampling (TS Decay) strategy based on time decay factor to address the issues of insufficient real-time performance and exploration utilization imbalance in dynamic advertising traffic allocation. By modeling advertising selection as a multi arm slot machine problem, the strategy effectively balances the utilization of historical data with real-time feedback optimization. The innovation of this study lies in combining Bayesian inference with time decay models, providing a scalable framework for real-time decision-making in lightweight scenarios. The proposed attenuation factor optimization method effectively alleviates the dependence of traditional multi arm slot machine algorithms on static environments while maintaining model simplicity, providing a new technical path for real-time optimization of advertising streaming systems.

Acknowledgements

We would like to extend my heartfelt gratitude to my supervisor for their insightful guidance, unwavering patience, and constant encouragement throughout this research. Their expertise and dedication have been instrumental in shaping this thesis. I also thank my family and friends for their love and support, which sustained me during challenging times. Special thanks to colleagues for their valuable feedback and assistance. We would like to extend our sincere appreciation for the support from the National College Students' Innovation and Entrepreneurship Program 2024 (Project Number: 202411057044). This funding has been crucial in facilitating the completion of this research, providing the necessary resources for comprehensive investigations and analyses. We recognize the significant role this program has played in enabling our academic endeavors and are grateful for its contribution.Finally, I appreciate the reviewers for their constructive comments that improved this work. This achievement is a testament to the collective support I have been fortunate to receive.

Funding Project: National College Students' Innovation and Entrepreneurship Program 2024, Project Number: 202411057044

References

[1] Huang Xiaohui Research on User Experience Optimization and Conversion Enhancement Strategies for E-commerce Websites Based on User Behavior Analysis [J]. Mall Modernization, 2025, (10): 5-8

[2] Chen Jing Research on Sampling Techniques for Sparse Data in Recommenda-tion Systems [D]. University of Electronic Science and Technology of China, 2023. DOI: 10.27005/d.cnki.gdzku.2023.005854

[3] Yan Yang, Chen Zeqiu, Deng Junlin Heterogeneous Network Representation Learning Method Based on Multi arm Slot Machine [J]. Journal of Tianjin Vo-cational and Technical Normal University, 2024, 34 (01): 61-65

[4] Wu Bo, Fang Xifeng, Zhang Haixiang, et al. Research on dynamic programming methods for complex product design tasks combining fuzzy set theory and greedy algorithms [J]. Manufacturing Technology and Machine Tools, 2025, (03): 92-98. DOI: 10.19287/j.mtmt.1005-2402.2025.03.012.

[5] Xu Huanxuan. Research on Slot Machine Learning Algorithm Based on Thompson Sampling and Offline Data[D]. Chongqing University, 2023. DOI: 10.27670/d.cnki.gcqdu.2023.002777.

[6] Zhao Jieting. Construction of Recommendation Algorithm Integrating User Interests and

Multi-Armed Bandit[D]. Shanghai International Studies Univer-sity, 2024. DOI:10.27316/d.cnki.gswyu.2024.000506.

[7] Wang Ding. Research on Recommendation Algorithms Based on Multi-Armed Bandit Models [D]. Shanxi University, 2022. DOI: 10.27284/d.cnki.gsxiu.2022.002281.

[8] Zhang Fuchun. Research on Federated Learning Client Selection Mechanism Based on Combinatorial Multi-Armed Bandit[D]. Beijing University of Posts and Telecommunications, 2022. DOI: 10.26969/d.cnki.gbydu.2022.001351.

[9] Ni Haojie. Research on the Impact of Information Flow Advertising Pushes on User Experience under Algorithmic Technology [D]. Shanghai University of Finance and Economics, 2023. DOI: 10.27296/d.cnki.gshcu.2023.001659.

[10] Lu Zewei. Design and Implementation of a Personalized Information Recom-mendation System [D]. Beijing Jiaotong University, 2020. DOI: 10.26944/d.cnki.gbfju.2020.003546.