Adaptive Budget Optimization for Multichannel Advertising Using Combinatorial Bandits

Extended Abstract

Briti Gangopadhyay
briti.gangopadhyay@sony.com
Sony
Japan

Zhao Wang
zhao.wang@sony.com
Sony
Japan

Alberto Silvio Chiappa
alberto.chiappa@epfl.ch
Sony, EPFL
Japan, Switzerland

Shingo Takamatsu
shingo.takamatsu@sony.com
Sony
Japan

ABSTRACT

Optimizing budget allocation is vital for digital advertising, yet practical algorithms remain scarce due to limited public datasets and realistic simulation environments. While multi-armed bandit (MAB) algorithms are well-studied, they struggle in non-stationary settings requiring rapid adaptation. This paper introduces three key contributions: (1) a simulation environment that emulates multichannel advertising campaigns using logged real-world data; (2) an enhanced combinatorial bandit strategy with efficient exploration, and change-point detection to adapt dynamically to market shifts; and (3) Empirical validation showing superior performance over baselines in reward and regret metrics across real-world campaigns.

KEYWORDS

Combinatorial Bandit, Non-stationarity, Digital Advertisement

ACM Reference Format:

Briti Gangopadhyay, Zhao Wang, Alberto Silvio Chiappa, and Shingo Takamatsu. 2025. Adaptive Budget Optimization for Multichannel Advertising Using Combinatorial Bandits: Extended Abstract. In *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Detroit, Michigan, USA, May 19 – 23, 2025*, IFAAMAS, 3 pages.

1 INTRODUCTION

Digital advertising, a rapidly growing field with a market size projected to exceed $830 billion by 2026 [2], involves managing diverse campaigns across formats (e.g., Search, Display) and platforms (e.g., Google, Meta) [11]. Effective budget allocation is crucial for maximizing Return on Ad Spend (ROAS) and ensuring ads reach high-quality traffic [20]. While multi-armed bandit (MAB) strategies [1, 6, 22] are popular for this task due to their simplicity, they often fail to adapt effectively in non-stationary environments where rapid adjustments are important [16]. A major challenge in developing adaptive algorithms is the lack of open-source datasets and simulation environments, as most business data is proprietary [5, 14, 27]. Testing algorithms on real traffic is costly and risky [21]. This paper addresses these gaps with the following contributions:

1. A public [9] simulation environment for multichannel, non-stationary ad campaigns to enable reproducible research.
2. An enhanced combinatorial bandit strategy with a modified mean function, targeted exploration, and change-point detection (CPD) for dynamic adaptation.
3. Theoretical guarantees of sub-linear regret ($O(\sqrt{T})$) and empirical results showing superior performance over SOTA baselines.

2 RELATED WORK

Budget allocation across multiple campaigns has been widely studied [6, 7], with significant contributions from industry players like Criteo, Netflix, and Lyft [4, 15, 18]. Typically, sub-campaigns are modeled as arms in a multi-armed bandit (MAB) problem, with combinatorial optimization used to allocate budgets based on expected rewards [28]. Parametric models, such as power laws [13] or sigmoids [11], combined with Thompson Sampling, often struggle in noisy environments, leading to deviations from true reward functions. Gaussian Process (GP) models [22, 23] offer flexibility with UCB or Thompson Sampling but lack domain knowledge integration and are limited to short timeframes, making them less effective in long-term settings. Addressing non-stationarity in MABs typically involves passive methods like sliding windows [26] or discounted rewards [10], which perform poorly in long campaigns with infrequent changes. Active methods, such as change point detection [3, 19], provide better adaptability, motivating our approach to incorporate dynamic adjustments.

3 METHODOLOGY

We follow the standard Automatic Budget Allocation (ABA) problem formulation [22], where an advertising campaign $\mathcal{A} = A_1, \ldots, A_n$, comprising N sub-campaigns, is managed over a finite time horizon T with a budget $\mathcal{B}$. Each day's budget $b_{j,t}$ for sub-campaign A_j must satisfy $\underline{b}t \leq bj, t \leq \bar{b}t$ while maximizing cumulative returns. The reward function $n_{j,t}$, mapping cost $x_{j,t}$ to feedback (e.g., clicks), changes dynamically due to market fluctuations, modeled as piecewise constant over *phases* $\mathcal{F}\phi$, separated by breakpoints $\mathcal{P}$. Within each phase, the reward function remains constant. We assume (1) reward changes exceed a detectable threshold τ, (2) breakpoints

Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Y. Vorobeychik, S. Das, A. Nowé (eds.), May 19 – 23, 2025, Detroit, Michigan, USA. © 2025 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

are separated by at least an unknown minimum duration T_p, and (3) reward functions are smooth, monotonically increasing, and exhibit diminishing returns [12, 14].

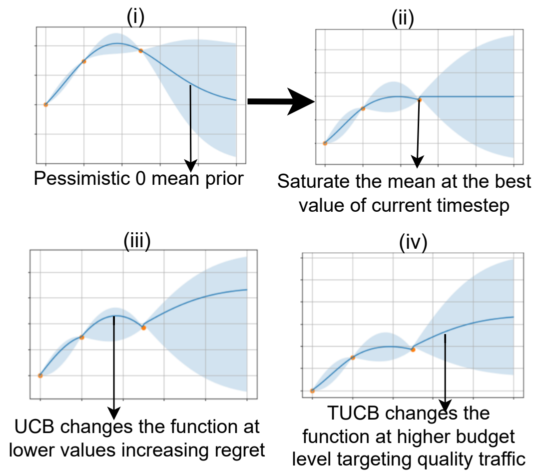

Figure 1: GP estimation with saturated mean and targeted UCB exploration.

Simulation Environment : To study the non-stationary multichannel budget allocation problem, we developed a simulation environment mimicking long-running ad campaigns from logged data of real campaigns. Daily cost consumption, $x_{j,t}$, is modeled as a truncated normal distribution, $x_{j,t} \sim \mathcal{N}(b_{j,t}, \sigma^2)$, constrained by $0 \leq x_{j,t} \leq 2b_{j,t}$. The cost-to-reward function, $n_j(x_{j,t}) = \alpha_c x_{j,t}^{\omega_c} + \epsilon$, incorporates noise ϵ, with parameters α_c and ω_c updated daily using curve fitting on the logged data. Abrupt reward function changes are modeled by maintaining future parameter estimates α_f and the model is updated when $|\alpha_c - \alpha_f| > 0.2$. This approach models dynamic, non-stationary campaign behavior in simulation.

Adaptive Budget Allocation (ABA) : The ABA algorithm involves four key steps: (1) estimating reward functions using Gaussian Processes (GP) [25] , (2) predicting rewards for each arm, (3) allocating budgets using a multi-choice knapsack [17], and (4) detecting change points. A standard GP model with a zero mean can restrict exploration to higher budget ranges. To address this, we use a saturating mean function defined as $\hat{n}_j = \hat{n}_{j\max}$ if $b_{j,i} > b_{j\max}$, and $\hat{n}_j = \hat{n}_j$ otherwise, where $\hat{n}j$ is the GP mean, $\hat{n}_{j\max}$ is the highest observed reward, and $b_{j\max}$ is the budget level corresponding to this reward. This ensures exploration focuses on effective regions.

$$\tilde{n}_j(\cdot) \leftarrow \hat{n}_j(\cdot) + \{\beta * (1-\theta_j) * \sigma_j\}|\mathbb{I}_{b_{j,i}>b_{j,max}} \tag{1}$$

We enhance exploration with a modified Upper Confidence Bound (UCB) (Eq. 1), where $\tilde{n}j(\cdot)$ incorporates campaign efficiency θ_j (e.g., normalized Cost per Click) and an indicator $\mathbb{I}_{b_{j,i}>b_{j\max}}$ to focus exploration on promising budget levels. This strategy reduces unnecessary exploration of low-reward regions.

For non-stationarity, two models are maintained: $\mathcal{M}j$, trained on data from the current phase, and $\tilde{\mathcal{M}}j$, using recent data. Change detection calculates the Mean Absolute Error (MAE) between their predictions as $pred_{diff} = \frac{1}{B}\sum_{i=1}^{B}|\mathcal{M}_j(b_i) - \tilde{\mathcal{M}}(b_i)|$, where B is the set of budgets. The reward model is refreshed when $pred_{diff}$ exceedes threshold τ, enabling dynamic adaptation to reward shifts.

4 EMPIRICAL STUDIES

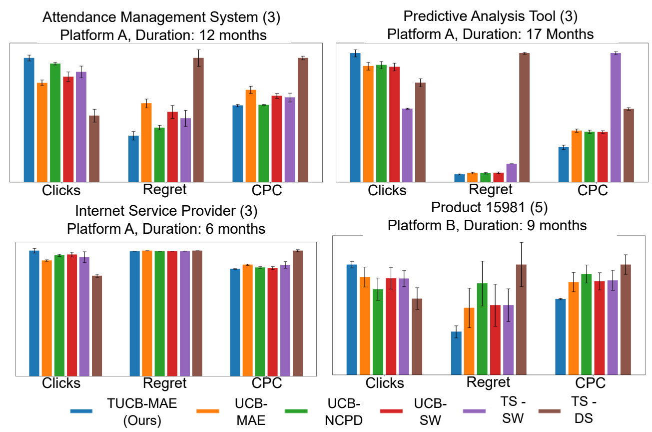

Figure 2: Comparison of normalized metrics (Clicks, Regret, CPC) across products w.r.t baselines

We conduct experiments using real logged campaign data from two platforms, denoted as Platform A and Platform B. Hyperparameters ($T_p = 20$, $window_{length} = 7$, budget granularity = 500). Campaigns are simulated in a long-horizon environment with noise $\epsilon \sim \mathcal{N}(0, 0.1)$. The proposed algorithm is compared to the following baselines: (1) UCB-MAE: Combines UCB exploration with MAE-based CPD [24]. (2) UCB-NCPD: UCB exploration without CPD, highlighting its importance. (3) UCB-SW: UCB with a 10-day sliding window [10]. (4) TS-SW: Thompson Sampling with a 10-day sliding window [8]. (5) UCB-DS: UCB with a 0.9 discounting factor for past data [10]. Results (Fig 2) are evaluated on: Clicks - Indicates user engagement and allocation effectiveness. Regret - Average cumulative regret compared to an oracle optimizer. Cost Per Click (CPC) - Lower CPC reflects higher ROAS and advertiser efficiency. Experiments span multi-channel campaigns (Display, Search-1, Search-2) running for over 5 months. Search-1 targets specific keywords; Search-2 uses broader terms. Results demonstrate the proposed algorithm's superior performance in clicks, regret, and CPC.

5 CONCLUSION

This paper explores the deployment of a combinatorial bandit algorithm for managing ad campaign budgets across multiple channels. We develop a simulation environment for long-horizon, logged data and propose enhancements like saturating mean, targeted UCB, and change point detection for better adaptation in non-stationary environments. Preliminary findings highlight the impact of non-stationarity and the potential for improved adaptability.

ACKNOWLEDGMENTS

The authors thank Sony Biz Networks Corporation for providing data for the ad campaign from services NURO Biz and AKASHI.

REFERENCES

[1] Vashist Avadhanula, Riccardo Colini Baldeschi, Stefano Leonardi, Karthik Abinav Sankararaman, and Okke Schrijvers. 2021. Stochastic bandits for multi-platform budget optimization in online advertising. In *Proceedings of the Web Conference 2021*. 2805–2817.

[2] Abraham Bagherjeiran, Nemanja Djuric, Kuang-Chih Lee, Linsey Pang, Vladan Radosavljevic, and Suju Rajan. 2024. AdKDD 2024. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (Barcelona, Spain) *(KDD '24)*. Association for Computing Machinery, New York, NY, USA, 6706–6707. https://doi.org/10.1145/3637528.3671476

[3] Yang Cao, Zheng Wen, Branislav Kveton, and Yao Xie. 2019. Nearly optimal adaptive procedure with change detection for piecewise-stationary bandit. In *The 22nd International Conference on Artificial Intelligence and Statistics*. PMLR, 418–427.

[4] Vincenzo D'Elia. 2019. On the causality of advertising. http://papers.adkdd.org/2019/invited-talks/slides-adkdd19-delia-causality.pdf. [Online; accessed September-2024].

[5] Qiyuan Deng, Kejia Hu, and Yun Fong Lim. 2023. Cross-Channel Marketing on E-commerce Marketplaces: Impact and Strategic Budget Allocation. *Available at SSRN 4332631* (2023).

[6] Yuan Deng, Negin Golrezaei, Patrick Jaillet, Jason Cheuk Nam Liang, and Vahab Mirrokni. 2023. Multi-channel autobidding with budget and ROI constraints. In *Proceedings of the 40th International Conference on Machine Learning* (Honolulu, Hawaii, USA) *(ICML'23)*. JMLR.org, Article 301, 28 pages.

[7] Paul W Farris, Dominique M Hanssens, James D Lenskold, and David J Reibstein. 2015. Marketing return on investment: Seeking clarity for concept and measurement. *Applied Marketing Analytics* 1, 3 (2015), 267–282.

[8] Marco Fiandri, Alberto Maria Metelli, and Francesco Trovò. 2024. Sliding-Window Thompson Sampling for Non-Stationary Settings. *arXiv preprint arXiv:2409.05181* (2024).

[9] Briti Gangopadhyay. 2025. Code cotributions for Adaptive Budget Optimization for Multichannel Advertising Using Combinatorial Bandits. https://github.com/sony/ABA. Accessed: 2025-01-23.

[10] Aurélien Garivier and Eric Moulines. 2011. On upper-confidence bound policies for switching bandit problems. In *International conference on algorithmic learning theory*. Springer, 174–188.

[11] Marco Gigli and Fabio Stella. 2024. Multi-armed bandits for performance marketing. *International Journal of Data Science and Analytics* (2024), 1–15.

[12] Marco Gigli and Fabio Stella. 2024. Multi-armed bandits for performance marketing. *International Journal of Data Science and Analytics* (2024), 1–15.

[13] Benjamin Han and Carl Arndt. 2021. Budget allocation as a multi-agent system of contextual & continuous bandits. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 2937–2945.

[14] Benjamin Han and Jared Gabor. 2020. Contextual bandits for advertising budget allocation. *Proceedings of the ADKDD* 17 (2020).

[15] Benjamin Han and Jared Gabor. 2020. Contextual bandits for advertising budget allocation. *Proceedings of the ADKDD* 17 (2020).

[16] EM Italia, A Nuara, Francesco Trovò, Marcello Restelli, N Gatti, E Dellavalle, et al. 2017. Internet advertising for non-stationary environments. In *Proceedings of the International Workshop on Agent-Mediated Electronic Commerce*. 1–15.

[17] Hans Kellerer, Ulrich Pferschy, David Pisinger, Hans Kellerer, Ulrich Pferschy, and David Pisinger. 2004. The multiple-choice knapsack problem. *Knapsack problems* (2004), 317–347.

[18] Randall Lewis and Jeffrey Wong. 2022. Incrementality bidding and attribution. *arXiv preprint arXiv:2208.12809* (2022).

[19] Fang Liu, Joohyun Lee, and Ness Shroff. 2018. A change-detection based framework for piecewise-stationary multi-armed bandit problem. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32.

[20] Phuong Ha Nguyen, Djordje Gligorijevic, Arnab Borah, Gajanan Adalinge, and Abraham Bagherjeiran. 2023. Practical Budget Pacing Algorithms and Simulation Test Bed for eBay Marketplace Sponsored Search.. In *AdKDD@ KDD*.

[21] Phuong Ha Nguyen, Djordje Gligorijevic, Arnab Borah, Gajanan Adalinge, and Abraham Bagherjeiran. 2023. Practical Budget Pacing Algorithms and Simulation Test Bed for eBay Marketplace Sponsored Search.. In *AdKDD@ KDD*.

[22] Alessandro Nuara, Francesco Trovò, Nicola Gatti, and Marcello Restelli. 2018. A Combinatorial-Bandit Algorithm for the Online Joint Bid/Budget Optimization of Pay-per-Click Advertising Campaigns. In *AAAI Conference on Artificial Intelligence*. https://api.semanticscholar.org/CorpusID:19147413

[23] Alessandro Nuara, Francesco Trovò, Nicola Gatti, and Marcello Restelli. 2022. Online joint bid/daily budget optimization of internet advertising campaigns. *Artificial Intelligence* 305 (2022), 103663.

[24] Gerlando Re, Fabio Chiusano, Francesco Trovò, Diego Carrera, Giacomo Boracchi, and Marcello Restelli. 2021. Exploiting History Data for Nonstationary Multi-armed Bandit. Springer-Verlag, Berlin, Heidelberg, 51–66. https://doi.org/10.1007/978-3-030-86486-6_4

[25] Eric Schulz, Maarten Speekenbrink, and Andreas Krause. 2018. A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. *Journal of mathematical psychology* 85 (2018), 1–16.

[26] Francesco Trovo, Stefano Paladino, Marcello Restelli, and Nicola Gatti. 2020. Sliding-window thompson sampling for non-stationary settings. *Journal of Artificial Intelligence Research* 68 (2020), 311–364.

[27] Xingfu wang, Pengcheng Li, and Ammar Hawbani. 2018. An Efficient Budget Allocation Algorithm for Multi-Channel Advertising. In *2018 24th International Conference on Pattern Recognition (ICPR)*. 886–891. https://doi.org/10.1109/ICPR.2018.8545777

[28] Haifeng Zhang and Yevgeniy Vorobeychik. 2017. Multi-channel marketing with budget complementarities. In *International Conference on Autonomous Agents and Multiagent Systems*.