Optimizing Advertising Efficacy: Implementing Cost-Effective Multi-Armed Bandit Algorithms

Biao Xiang *

Johns Hopkins Carey Business School, Baltimore, Maryland, 21202, United States

* Corresponding Author Email: bxiang2@jhu.edu

Abstract. This paper proposes a novel approach of multi-armed bandit (MAB) algorithms for online advertising across multiple platforms with no budget limitation. The study adapts the application of two algorithms, Cost-Subsidized Upper Confidence Bound Bandit (CS-UCB) and Cost-aware Cascading Upper Confidence Bound Bandit (CC-UCB), with the objective to maximize the click-through rate (CTR) while considering cost efficiency. Departing from traditional budget-constrained models such as Bandits with Knapsacks (BwK), this study assumes a competitive advertising environment where market share and ad exposure are prioritized over strict budget adherence and cost control. It provides advertisers flexibility in cost control, maximizing CTR within the bounds of acceptable cost concessions, balancing the dual objectives of minimize quality regret and cost regret. The introduction of this novel appraoch extends the application of Multi-Armed Bandit (MAB) algorithms to advertising strategies in highly competitive markets without budget constraints. This study analysises the performance of both CC-UCB and CS-UCB algorithms with empricial research by using real-world data, showing online advertisements bidding strategies from multiple ad platforms in real-time bidding (RTB) system.

Keywords: Online Advertising; Multi-armed Bandit; Cost-Efficiency; No-Budget Advertising.

1. Introduction

The realm of online advertising has emerged as a cornerstone in the digital economy, revolutionizing the way businesses engage with their target audience. The industry of online advertising has witnessed a meteoric rise, reaching an estimated value of $209.7 billion in the United States by the year 2022 [1]. This expansion is driven by the rise of digital platforms which have become integral to the advertising ecosystem, catalyzing competitive auctions for ad impressions that attract a diverse array of advertisers [2]. Ad exchanges utilize auctions to allocate user impressions through both guaranteed and nonguaranteed selling channels. In the guaranteed selling approach, advertisers acquire a specific number of ad impressions at a pre-set price, which guarantees stable ad placement. Conversely, non-guaranteed selling adopts a dynamic, real-time bidding (RTB) system, where ad impressions are auctioned in real-time, allowing for variable pricing and increased placement flexibility. The focus of this study is on optimizing bidding strategies within the RTB system with conditions of unconstrained budget limitations.

Advertisers participating in RTB auctions confront significant uncertainty about the value of the impressions they are bidding for, exacerbated by the varying auction formats and target user base across different ad exchange platforms. With a limited understanding of the true valuation in a variety of platforms, advertisers have to bid on different platforms to learn the distribution of returns in each one. The core challenge lies in effectively allocating funds across these platforms to learn the true return in terms of ad clicks while maximizing ad click-through rates within limited budgets.

The budget-constrained model such as Bandits with Knapsacks (BwK) model has been commonly studied in literatures to optimize advertisement bidding strategies within predefined budget constraints in second-price auctions (SPA) [3-5]. To set up this problem as a BwK problem, suppose the advertiser bids with J platforms with a limited budget B and value distribution in each auction $V = (V_1, \dots, V_j)$. By playing an arm x_{b_j} for each time $t \in T$, the advertiser chooses a bid $b_{j,t}$ which corresponds a CTR $x_{j,t} \epsilon$ [1,0], an expected value $v_{i,t}$ and payment $p_{j,t}$. Reward distribution

is the difference between expected gain and expected price paid, i.e. $v_j\bar{x}_j(b) - \bar{p}_j(b)$ [6]. Advertiser tries to maximize the cumulative reward before running out of budget.

The budget-constrained models are based on the known budget limitation, which focuses on bidding strategies of maximizing CTR with minimum cost. However, in highly competitive markets, advertisers who are driven to capture and retain market share, might engage in advertising campaigns without predefined budget constraints to enhance product exposure. To survive the competition, advertisers prioritize market share and ad exposure supersedes strict adherence to budget limits and cost control. Consequently, this strategy, often adopted to surpass competitors and establish market presence, challenges the adequacy of the traditional BwK model. Additionally, while the classic MAB problem focuses primarily on CTR, it neglects the cost implications of ad placements. Advertisers, even without set budget caps, still endeavor to allocate their resources effectively, aiming to maximize CTR within a cost-efficient framework. This necessitates a novel approach to the MAB problem, one that integrates cost considerations with CTR maximization in an unconstrained budget environment, enabling advertisers to pursue an optimal balance between ad exposure and cost efficiency.

In response to the non-budget-constrained advertising in competitive market, this research applies the Cost-Subsidized Upper Confidence Bound Bandit and Cost-aware Cascading Upper Confidence Bound Bandit algorithms to maximize CTR with unknown budget while considering cost-efficiency. Instead of using cost as first priority in budget-constrained models, this research focus on maximizing CTR, and provides flexibility for advertisers in the extent of cost control in the consideration of advertising decisions, and applies empirical research with real-world data. This advancement allows for a more nuanced approach to balancing ad exposure with cost-effectiveness, tailored to the complexities of contemporary online advertising strategies.

2. Theoretical Background and Developments

2.1. Bandits with Knapsacks

The Bandits with Knapsacks (BwK) model, initially introduced by Badanidiyuru et al. represents an extension of the MAB framework, integrating it with linear optimization from the Knapsack problem [7, 8]. The general format of BwK model considers MAB problem under a resource constraint. It operates on a known set of resources d, and K arms correspond with unknown reward distribution. By pulling arm a at time t, it yields a reward r_t and a cost c_t^i for resource i. The player's objective is to maximize the cumulative reward before running out of any of resource.

The BwK framework encompasses a broad spectrum of constrained bandit problems, particularly relevant in the context of second-price auctions in online advertising's RTB systems [9]. The topic has been extensively explored in various studies focusing on stochastic bandits under knapsack constraints [10-13]. However, in a highly competitive market, to maintain the product exposure and market share, advertisers may not have a hard budget limitation, where budget-constrained model is inadequate.

2.2. Multi-Armed Bandit with Cost Subsidy

In traditional stochastic MAB problem, a player is presented K arms with unknown but fixed reward distribution. The player repeatedly pulls an arm $k \in [K]$ in time T, and receive a reward r_t in each time t. The player aims to maximize cumulative reward, or equivalently, minimize cumulative regret. However, there's a dilemma: to get a more accurate estimation of reward distribution of each arm by pulling every arm (exploration) or to continually pull the arm already know to yield highest reward (exploitation) [14]. Upper Confidence Bound is one of the commonly used algorithms to handle the exploration-exploitation tradeoff in MAB problem.

Sinha et al. proposed a novel variant of MAB problem, MAB with cost subsidy, which considers cost management [15]. In this model, after pulling an arm, player receives a reward r_t from a fixed but unknown distribution $\mathcal{F}_i$ with mean μ_i, and incurs a known cost c_i. In each step t, the player will

pull the cheapest arm from a set of arms, each exceeding a smallest tolerated reward $(1-\alpha)\mu_{m_*}$, where m_* denote the arm with highest expected mean and α is a fixed known value refer to as subsidy factor.

2.3. Cost-aware Cascding Bandits

Kventon et al. proposed Cascading Bandit as a learning variant of cascading model. Cascading Bandit considers the farmwork where at each step, generating an ordered list L out of K arms, proceeding to pull arms form the beginning to the end of L, until get first positive reward (user click) [16]. Gan et al. introduced Cost-aware Cascading Bandit (CCB), considering the cost of pulling individual arms. In CCB model, the ranking of arms in L is dependent on reward and cost of arms, which affects the expected net reward [17].

3. Systematic Analysis and Empirial Research

3.1. Model Formulation and Setup

The empirical research uses the real-world dataset iPinYou to evaluate the performance of CS-UCB and CC-UCB in RTB with no budget limitation [18]. The periods of research will be set as $T = 10{,}000$. to formulate models, suppose that there are K platforms, in each step $t \in [T]$ the adviser makes a bid in a platform $i \in [K]$. With each bidding at step t, advertise receives a reward (click) $r_{i,t} \in \{0,1\}$ from an unknown distribution $\mathcal{F}_{r,i}$ with mean $\mu_{r,i} \in [0,1]$, and generates a cost $c_{i,t}$ from an unknown distribution $\mathcal{F}_{c,i}$ with mean $\mu_{c,i}$. The UCB padding term at step t is $\beta_i(t) \coloneqq \sqrt{\frac{4\log T}{T_i(t)}}$.

3.2. Efficacy of Cost Subsidy UCB Algorithm

Let m^* denotes the platform with highest CTR, $i.e., m^* = \text{argmax}_{i\in[K]}\mu_{r,i}$With a priori known subsidy factor α, $(1-\alpha)\mu_{r,m^*}$ is defined as smallest tolerated reward, and C_* is the set of platforms whose expected CTR is higher than the smallest tolerated reward, $i.e., C_* = \{i \in [K] \mid \mu_{r,i} \geq (1-\alpha)\mu_{r,m^*}\}$. Advertiser aims to bid with platform i_* which has the minimum average cost in C_*, $i.e. i_* = \text{argmin}_{i\in C_*}\mu_{c,i}$

Since the model considers both cost and reward, Gan et al. proposed two notions of regret – quality and cost regret to measure the performance of any policy π, and any policy must incur a regret of $\Omega(K^{1/3}T^{2/3})$ on at least one of the regret metrics: $Quality\_Reg_\pi(T,\alpha,\mu,c) = \mathrm{E}\left[\sum_{t=1}^{T}\max\{(1-\alpha)\mu_{m_*} - \mu_{\pi_t}, 0\}\right]$, $Cost\_Reg_\pi(T,\alpha,\mu,c) = \mathrm{E}\left[\sum_{t=1}^{T}\max\{c_{\pi_t} - c_{i_*}, 0\}\right]$, where $\boldsymbol{c} = (c_1, \dots c_K), \mu = (\mu_1, \dots \mu_K)$.

In CS-UCB algorithm, UCB indices is $\mu_i^{score}(t) \coloneqq \min\{\hat{\mu}(t) + \beta_i(t), 1\}$. For each step t, the learning agent pull the arm with the lowest average cost in the feasible set $Feas(t) = \{i: \mu_i^{score}(t) - (1-\alpha)\mu_{m_t}^{score}(t) \geq 0\}$, $i.e. I_t = \text{argmin}_{i\in Feas(t)}\mu_{c,i}$ observes reward $r_{i,t}$ and cost $c_{i,t}$, then iterate with updated information.

Fig 1 and Fig 2 demonstrate the cost regret and quality regret of CS-UCB with different subsidy factor α. The width of error bands is 3 standard deviations based on 10 runs.

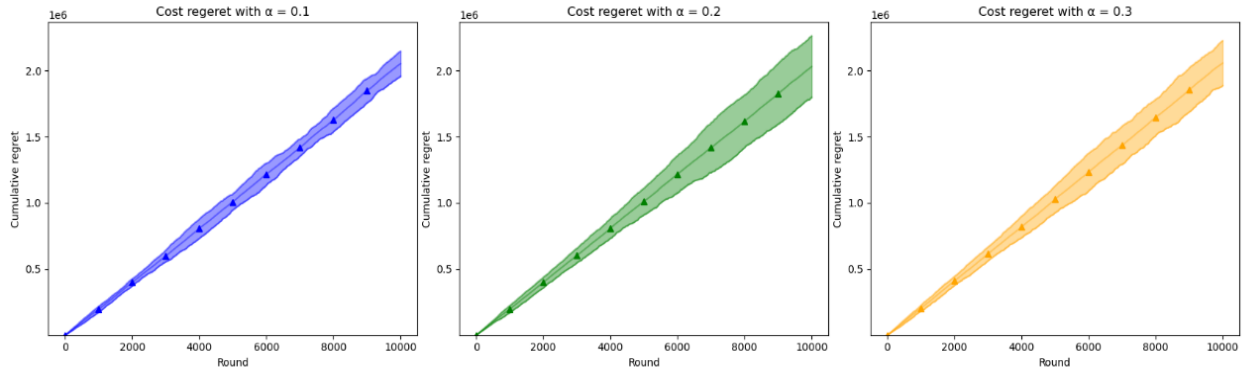

Fig 1. Cost regret of CS-UCB (Photo/Picture credit: Original).

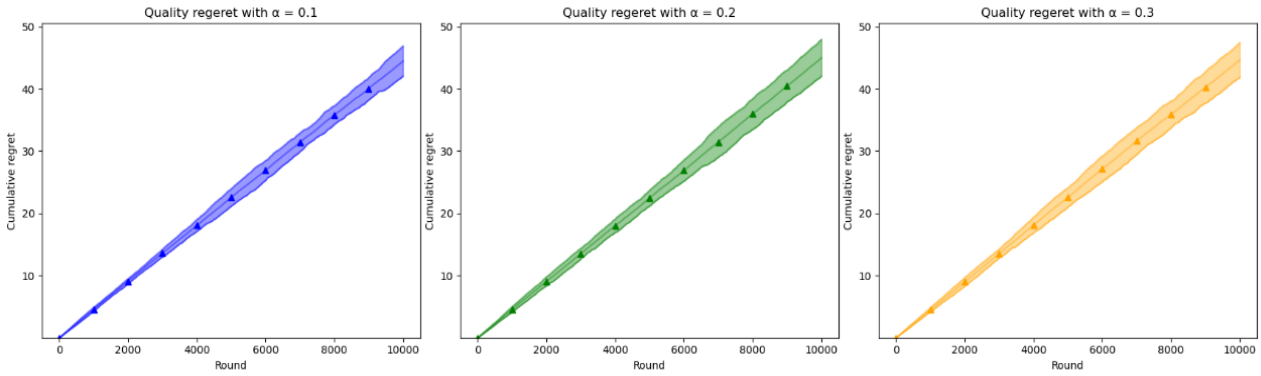

Fig 2. Quality regret of CS-UCB (Photo/Picture credit: Original).

3.3. Performance Evaluation of Cost-Aware Cascading UCB

Let $I_m = \{I_m(1), I_m(2), \ldots, I_m(K))\}$ denotes an ordered list of arms for all arms in$[K]$, where $I_m(n)$ is the nth arm to be pulled. In all T rounds, it has m ordered lists I_m with$m \times n = T$. Adviser will bid in platform from list I_t sequentially until receive a $r_{i,t} = 1$. Denote $\tilde{I}_m \in I_m$ as the list of arms that have been actually pulled inI_m.

In CC-UCB algorithm, denotes $U_{i,m} = \hat{\mu}_{i,m} + \beta_i(m)$ and $L_{i,m} = max(\hat{c}_{i,m} - \beta_i(m), 0)$ as the UCB indices of reward and cost, I_m is ranked in descending order of$\frac{U_{i,m}}{L_{i,m}}$. The first arm with highest $\frac{U_{i,m}}{L_{i,m}}$ in I_m isi_* $i.e. i_* = \text{argmax}_{i \in I_m} \frac{U_{i,m}}{L_{i,m}}$. Let $I_* = \{i \in [K] | \frac{U_{i,m}}{L_{i,m}} \geq (1-\alpha)\frac{U_{i^*,m}}{L_{i^*,m}}\}$ denotes the list of smallest acceptable arms. The learning agent will sequentially pull all arms in I_* until it gets a$r_{i,t} = 1$, observe cost and reward of all pulled arms, then iterate with updated information.

Fig 3 and Fig 4 demonstrate the cost regret and quality regret of CC-UCB with different subsidy factorα. The width of error bands is 3 standard deviations based on 10 runs.

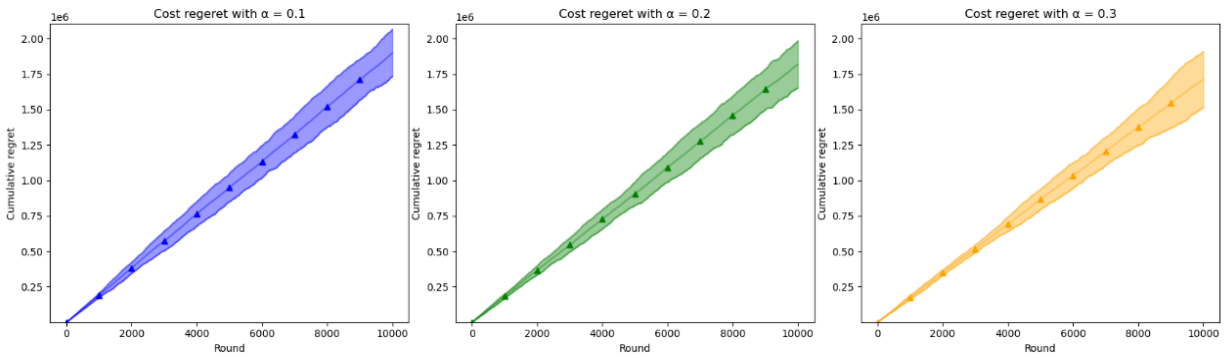

Fig 3. Cost regret of CC-UCB (Photo/Picture credit: Original).

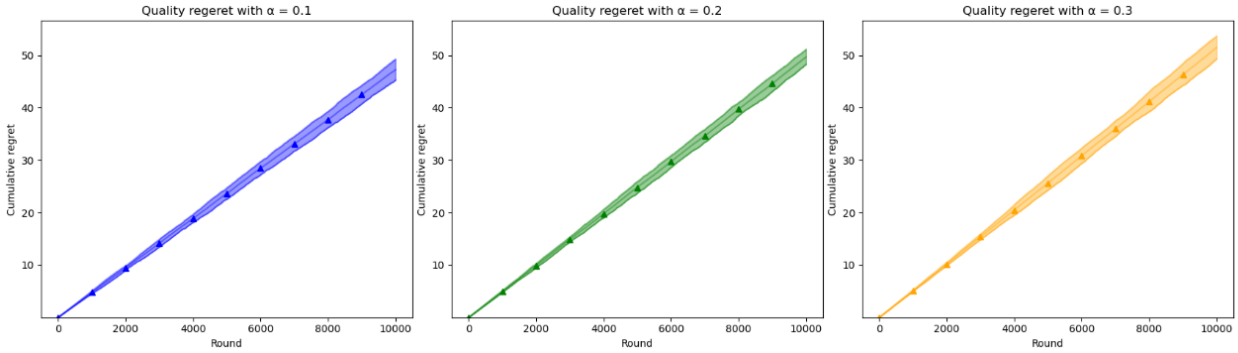

Fig 4. Quality regret of CC-UCB (Photo/Picture credit: Original).

3.4. Comparative Analysis of Both Algorithms

Comparing the outcomes of both CS-UCB and CC-UCB, as it is shown in Fig 1-4, with the increase of subsidy factorα, both algorithms exhibit a trend of increasing accumulated quality regret and a decreasing accumulated cost regret. The value of α determines the range of feasible set $Feas(t)$ in CS-UCB and lists of smallest acceptable arms I_* in CC-UCB, from which we choose the final platform to bid with the consideration of cost control in different strategies. Departing from directly bid with platform with the highest CTR, advertiser choose to bid with a cost-efficient one which may have a lower CTR. As α incerases, indicating a heightened consideration of cost control by the advertiser, both algorithms have a higher range of choices. Consequently, advertiser has a higher chance to bid in the platform with lower CTR, which result in a higher quality cumulative regret and lower cost cumulative regret.

Different from CS-UCB, in the Fig 2 and Fig 4, it demostrates that CC-UCB has more convergent accmulative quality regret. This may be attributed to the fact that CC-UCB does not directly choose the cheapest platform in the smallest acceptable list, but bid with all arms in the list with an order relies on$\frac{U_{i,m}}{L_{i,m}}$, which is dependent on both cost and reward of arm.

4. Challenges and Future Prospects

This study, stand in the perspective of individual advertiser, optimizes multi-platform online advertising strategies to achieve optimal CTR. However, it does not take into account of the complex bidding procedure in the RTB system. A direction of future work is considering the competitors' bidding price in various auction format across multi-platform, especially in a highly competitive environment.

5. Conclusion

This research introduces a groundbreaking application of the Multi-Armed Bandit algorithm for decision-making in online advertising bidding across various ad exchange platforms. It operates under the premise of a highly competitive market where the imperative for product exposure surpasses the constraints of stringent budgeting and cost control. In such a non-budget-limited context, the study adapts two algorithms, the Cost-Subsidized Upper Confidence Bound and the Cost-aware Cascading Upper Confidence Bound, with an innovative approach to maximize Click-Through Rate while maintaining cost efficiency. The adaptation involves the integration of a subsidy factor α into the algorithms, a strategic element that plays a crucial role in aligning the pursuit of high CTR with the pragmatic aspect of cost effectiveness. This methodology allows for a nuanced approach to online ad bidding, where decisions are not solely driven by budget constraints but are instead focused on maximizing visibility and engagement. The application of CS-UCB and CC-UCB algorithms in this setting showcases a significant shift in online advertising strategies, catering to the dynamics of a

competitive market while balancing the dual objectives of maximizing exposure and maintaining cost efficiency.

References

[1] Statista. (2022). Online advertising revenue in the United States from 2000 to 2022. Retrieved from https://www.statista.com/statistics/183816/us-online-advertising-revenue-since-2000/.

[2] Choi, H., Mela, C. F., Balseiro, S. R., et al. (2020). Online display advertising markets: A literature review and future directions. Information Systems Research, 31(2), 556-575.

[3] Conitzer, V., Kroer, C., Sodomka, E., et al. (2022). Multiplicative pacing equilibria in auction markets. Operations Research, 70(2), 963-989.

[4] Waisman, C., Nair, H. S., Carrion, C. (2019). Online causal inference for advertising in real-time bidding auctions. arXiv preprint arXiv:1908.08600.

[5] Haoyu, Z., Wei, C. (2020). Online second price auction with semi-bandit feedback under the non-stationary setting. Proceedings of the AAAI Conference on Artificial Intelligence, 34(04), 6893-6900.

[6] Chen, Z., Wang, C., Wang, Q., et al. (2022). Dynamic budget throttling in repeated second-price auctions. arXiv preprint arXiv:2207.04690.

[7] Badanidiyuru, A., Langford, J., Slivkins, A. (2014). Resourceful contextual bandits. Conference on Learning Theory. PMLR, 1109-1134.

[8] Badanidiyuru, A., Kleinberg, R., Slivkins, A. (2018). Bandits with knapsacks. Journal of the ACM (JACM), 65(3), 1-55.

[9] Sankararaman, K. A., Slivkins, A. (2021). Bandits with knapsacks beyond the worst case. Advances in Neural Information Processing Systems, 34, 23191-23204.

[10] Avadhanula, V., Colini Baldeschi, R., Leonardi, S., et al. (2021). Stochastic bandits for multi-platform budget optimization in online advertising. Proceedings of the Web Conference 2021, 2805-2817.

[11] Susan, F., Golrezaei, N., Schrijvers, O. (2023). Multi-Platform Budget Management in Ad Markets with Non-IC Auctions. arXiv preprint arXiv:2306.07352.

[12] Celli, A., Colini-Baldeschi, R., Kroer, C., et al. (2022). The parity ray regularizer for pacing in auction markets. Proceedings of the ACM Web Conference 2022, 162-172.

[13] Gaitonde, J., Li, Y., Light, B., et al. (2022). Budget pacing in repeated auctions: Regret and efficiency without convergence. arXiv preprint arXiv:2205.08674.

[14] Robbins, H. (1952). Some aspects of the sequential design of experiments.

[15] Zhu, X., Zhao, Z., Wei, X., & others. (2021). Action recognition method based on wavelet transform and neural network in wireless network. In 2021 5th International Conference on Digital Signal Processing (pp. 60-65).

[16] Kveton, B., Szepesvari, C., Wen, Z., et al. (2015). Cascading bandits: Learning to rank in the cascade model. International conference on machine learning. PMLR, 767-776.

[17] Gan, C., Zhou, R., Yang, J., et al. (2020). Cost-aware cascading bandits. IEEE Transactions on Signal Processing, 68, 3692-3706.

[18] Liao, H., Peng, L., Liu, Z., et al. (2014). iPinYou global rtb bidding algorithm competition dataset. Proceedings of the Eighth International Workshop on Data Mining for Online Advertising, 1-6.