Optimizing online advertising with muti-armed bandit algorithms

Chenyan Jiang

Hefei University of Technology, Hefei, Anhui, 230002, China

2022213473@mail.hfut.edu.cn

Abstract. The rapid digitalization of the global economy has significantly transformed the landscape of advertising, necessitating more sophisticated and adaptive strategies to reach and engage with consumers effectively. This paper explores the application of multi-armed bandit (MAB) algorithms as a powerful tool for optimizing online advertising processes. We examine how MAB algorithms can enhance various stages of the advertising cycle, from audience segmentation and creative development to bidding strategies and real-time optimization. Through an analysis of existing literature and practical applications, we demonstrate the potential of MAB algorithms to balance the trade-offs between exploration and exploitation, enabling advertisers to maximize click-through rates, conversion rates, and return on investment. Furthermore, we address specific challenges such as the cold-start problem and the optimization of search advertising, proposing innovative solutions that leverage the adaptive capabilities of MAB algorithms. Our findings suggest that integrating MAB algorithms into online advertising strategies can significantly improve targeting accuracy, user engagement, and overall advertising performance. We conclude by discussing the implications of these findings and suggesting directions for future research to further enhance the application of MAB algorithms in the evolving digital advertising landscape.

Keywords: Online advertising, multi-armed bandit algorithms, reinforcement learning, optimization.

1. Introduction

The rapid digitalization of the global economy has significantly transformed advertising, requiring more sophisticated, data-driven strategies to remain competitive. As traditional methods become less effective, the integration of reinforcement learning, particularly multi-armed bandit (MAB) algorithms, has gained traction in the advertising sector. These algorithms are well-suited to the dynamic nature of digital advertising, where balancing exploration and exploitation is crucial. Research highlights the potential of MAB algorithms like Thompson sampling and Upper Confidence Bound (UCB) to optimize ad performance across various stages, although challenges and research gaps persist [1].

In recent years, research has focused on applying MAB algorithms to various aspects of the online advertising process. Zhang explored how reinforcement learning bandit algorithms could be utilized to optimize advertising strategies, emphasizing the importance of adapting to changing market conditions and user preferences [2]. Baardman et al.investigated the application of MAB algorithms for learning optimal online advertising portfolios with periodic budgets, highlighting the benefits of robust

optimization and UCB techniques in mitigating information delays and budget constraints [3]. Similarly, Qiu and Ji introduced an embedded bandit algorithm based on agent evolution to address the cold-start problem in advertising recommendation systems [4]. This approach improved ad click-through and conversion rates, particularly under conditions of limited data. Xu et al. examined the impact of estimation bias and data noise on the effectiveness of MAB algorithms in search advertising optimization, proposing strategies to dynamically allocate budgets and adjust strategies based on realtime data [5]. Nicolas Gutowski et al. studied the problem of balancing global accuracy and diversity in recommendation systems and proposes a new algorithmic combination method called Gorthaur-EXP3 [6]. Marco Bonalumiinvestigated the bid strategy optimization problem in real-time bidding and proposes a new UCB algorithm to select the optimal bid in order to maximize the advertiser's expected profit [7]. Liu and Bo developed a machine learning-based intelligent ad recommendation and anomaly monitoring system, featuring a real-time recommendation algorithm using Gaussian processes and an ad traffic detection method with random forests, which improves digital marketing precision and maintains ad campaign integrity through pattern recognition and anomaly detection [8]. Schwartz et al. presented a new multi-armed bandit strategy," Thompson Sampling with a Hierarchical Generalized Linear Model" (TS-HGLM), which enhances online ad performance by 8% in customer acquisition without additional cost and shows that this model outperforms traditional methods in certain scenarios [9]. Yang et al. introduced the Dynamic Contextual Multi-Armed Bandit (DAC-MAB) model, which combines dynamic conversion rate prediction, contextual learning, and arm overlap modeling to address dynamic changes in online ad targeting, and demonstrates superior performance compared to traditional MAB variants in experiments [10]. Nuara and Alessandro examined internet advertising optimization from the advertiser's perspective, proposing various AI methods integrated with MAB algorithms to address key issues such as bidding and budget optimization, ad goals, and inter-activity dependencies, and demonstrates the effectiveness of these methods in enhancing ad campaign management and performance [11]. Mussi and Marco explored methods for dynamic pricing and ad budget optimization using online learning and multi-armed bandit algorithms, introducing new algorithms like DynaLT, DLB, and FRB, and demonstrating their effectiveness in enhancing sales and ad efficiency in real-world applications [12]. These studies underscore the potential of MAB algorithms to enhance targeting accuracy, user engagement, and overall advertising performance. However, despite these contributions, there are limitations in the existing research that need to be addressed.

One of the primary limitations of the current research is the focus on individual components of the advertising process, such as bidding or creative optimization, without fully integrating them into a cohesive framework that addresses the entire advertising ecosystem. Additionally, there is a lack of research on how MAB algorithms perform in multi-channel and multi-device environments, which are increasingly important in today's digital landscape. The impact of feedback delays and estimation biases on the effectiveness of MAB algorithms in real-world scenarios is also not fully understood. Moreover, the ethical implications of using MAB algorithms to target users based on their behavior and preferences have raised concerns about privacy and data security.

This study aims to address these gaps by exploring the application of MAB algorithms in optimizing online advertising processes comprehensively. The research focuses on enhancing various stages of the advertising cycle, from audience segmentation and creative development to bidding strategies and real-time optimization. Through an analysis of existing literature and practical applications, the study demonstrates the potential of MAB algorithms to balance the trade-offs between exploration and exploitation, enabling advertisers can maximize click-through rates, conversion rates, and return on investment. The research adopts a systematic approach, utilizing a combination of theoretical analysis and empirical validation to investigate the efficacy of MAB algorithms in online advertising. The methods include a review of current literature to identify existing research gaps, the development of a conceptual framework that integrates MAB algorithms into the entire advertising process, and the implementation of simulation experiments using synthetic and real-world ad data to validate the proposed solutions. The study also addresses specific challenges such as the cold-start problem and the

optimization of search advertising, proposing innovative solutions that leverage the adaptive capabilities of MAB algorithms.

2. The application of MAB in online advertising

The online advertising process is a multi-stage endeavor encompassing campaign goal setting, audience analysis, creative development, platform selection, bidding, and optimization. MAB algorithms are particularly well-suited for handling the complexities of this process. Initially, during the goal-setting and budgeting phase, MAB algorithms can be utilized to optimize budget allocation by exploring various strategies and exploiting the most effective ones. In audience analysis and segmentation, MAB algorithms can dynamically adjust targeting based on real-time data, improving the accuracy of audience predictions. During creative development, MAB techniques allow for the efficient testing of different ad creatives, helping advertisers quickly identify the most engaging and effective content. Platform selection involves evaluating various platforms, and MAB algorithms can aid in resource allocation by identifying platforms that yield the best results. In the bidding and auction phase, MAB algorithms optimize bidding strategies by analyzing historical data and adjusting bids to maximize visibility and user engagement. Furthermore, in the monitoring and optimization stage, MAB algorithms provide realtime feedback on key performance indicators (KPIs) such as click-through rates (CTR), conversion rates (CVR), and return on investment (ROI), enabling advertisers to fine-tune their strategies. Additionally, MAB algorithms enhance A/B testing by adaptively focusing on promising options, leading to more efficient experimentation. Overall, MAB algorithms offer a robust framework for addressing the challenges of the online advertising process, allowing advertisers to continuously adapt and optimize their strategies in response to evolving market conditions and user preferences. To further understand the application of Multi-Armed Bandit (MAB) algorithms in online advertising, we will now delve into three specific problems: online advertising portfolio optimization, the cold-start problem, and search advertising optimization. Each of these issues represents a distinct challenge within the online advertising process, and MAB algorithms offer tailored solutions to optimize and address these challenges effectively.

2.1. Online advertising portfolio optimization problem

The fundamental units of online advertising are" targets," which represent a combination of ads, customer segments, and online channels. These targets can be exemplified by a particular advertisement being shown to users searching for a specific keyword on a search engine like Google. Advertisers first identify which customer segments they want to reach. This involves selecting the right ad to show to the right users on the right channels. Publishers of online advertisements (e.g., Google, Facebook) operate real-time auctions to manage the high demand for ad slots compared to their limited supply. Advertisers place bids in these auctions to secure ad slots for their selected targets. Advertisers maintain and periodically update their portfolios of targets based on feedback from the ad platforms. Publishers use this feedback to adjust their bidding strategies and target selections for the next period. Advertisers generate revenue from customer clicks or conversions on their ads. At the same time, costs are incurred based on auction payments and customer interactions with the ads. These revenues and costs are reported with a delay by the ad platforms.

In the process, advertisers face several challenges. First, the limited data on each target makes it difficult to accurately estimate expected revenue and costs. To address this issue, the explorationexploitation framework helps balance the exploration of new targets with the exploitation of known profitable ones, optimizing ad performance with limited data. Second, feedback delays mean that advertisers cannot adjust their strategies in real-time. In this case, the Optimistic Robust Learning (ORL) algorithm combines robust optimization with Upper Confidence Bound (UCB) techniques to effectively use historical feedback for decision-making, mitigating the impact of information delays. Third, the uncertainty in revenue and cost estimates arises from random factors such as customer behavior and competitor bidding. UCB techniques within the ORL algorithm help balance exploration and

exploitation, reducing potential losses due to this uncertainty and improving revenue estimation accuracy. Lastly, budget constraints require advertisers to manage costs effectively. The robust optimization approach within the ORL algorithm creates an oracle strategy that approximates the optimal solution, ensuring total costs remain within budget while optimizing ad performance. Additionally, the design of bounded expected regret minimizes revenue losses from suboptimal target choices, and simulation validation using synthetic and real-world ad data demonstrates that the ORL algorithm significantly reduces regret and improves overall performance and revenue [3].

The ORL algorithm consists of the following main steps:

1. Historical Data Analysis:

- Analyze past ad performance data to build a model of the revenue distribution for each advertising portfolio.
- Calculate the expected revenue and confidence intervals for each portfolio.

2. UCB Strategy Application:

• Calculate the UCB value for each portfolio using the formula:

$$UCB_a(t) = \hat{\mu}_t(a) + \sqrt{\frac{2logt}{n_t(a)}}$$
(1)

where:

- $\mu^{a}(t)$ is the average revenue of portfolio *a* at time *t*;
- $n_a(t)$ is the number of times portfolio *a* has been selected by time t; t is the current time step.

3. Robust Optimization:

- Apply robust optimization techniques to consider the worst-case revenue scenario, taking into account all possible feedback delays and market fluctuations.
- Solve the following robust optimization problem:

$$\max_{\mathbf{x}} \min_{\xi \in \Xi} \mathbf{c}^{\mathsf{T}} \mathbf{x} - \boldsymbol{\xi}^{\mathsf{T}} \mathbf{x}$$
(2)

where:

- **x** is the selection vector for the advertising portfolio.
- c is the vector of revenue coefficients.
- ξ represents uncertainty factors; Ξ is the set of uncertainties.
- 4. **Decision Update:** Update the expected revenue and UCB values for each portfolio based on new feedback data.
 - Adjust the selection strategy of the advertising portfolio according to the optimization results.

The ORL algorithm effectively addresses the challenges of uncertainty and feedback delays in online advertising portfolio optimization by integrating UCB techniques with robust optimization. The experimental results demonstrate that the ORL algorithm significantly improves the efficiency of portfolio selection and maximizes the return on investment for advertisers.

2.2. Cold-start problem

Advertising recommendation systems are essential tools for e-commerce and internet platforms to enhance user experience and boost revenue by analyzing users' browsing histories, purchase records, and behavior patterns to display ads most likely to capture users' interest. This personalization increases click-through rates, conversion rates, and overall user satisfaction. However, a significant challenge they face is the cold start problem, which occurs when there is insufficient user or ad data, making effective recommendations difficult, especially for new users or advertisements. To address this, the MAB algorithm effectively balances exploration and exploitation, optimizing ad recommendations in uncertain environments. Exploration involves testing different ads to gain new insights into user preferences, while exploitation focuses on selecting the ads most likely to be accepted based on existing information. The paper " An Embedded Multi-Agent Multi-Armed Bandit Algorithm Based on Agent Evolution for the Cold Start Problem" introduces a MAB algorithm within an agent evolution framework to solve this challenge. This system develops multiple agents, each with different strategies to handle advertising recommendations, optimizing their approaches through competition and cooperation. Agents adjust their strategies by simulating natural selection and genetic mutations, adapting to user preferences and ad effectiveness. By integrating the MAB algorithm, each agent balances exploration and exploitation when recommending ads, selecting the optimal ad display strategy. This method offers several benefits: the agent evolution framework enables rapid adaptation to new users and ads without extensive data, effectively solving the cold start problem; embedding the MAB algorithm allows the system to explore new ads while leveraging existing user preference information, enhancing recommendation accuracy and user satisfaction; and the multi-agent system and evolutionary strategies provide flexibility and scalability, making it adaptable to various advertising scenarios. In practice, this system significantly improves ad click-through and conversion rates. For instance, in an e-commerce platform experiment, the system optimized ad performance under cold start conditions, enhancing the user shopping experience. This method demonstrates that leveraging advanced machine learning and intelligent algorithms can significantly improve the precision and effectiveness of digital marketing strategies [4].

The Embedded MAB Algorithm with Agent Evolution consists of the following main steps:

- 1. Initialize a population of agents, each with a set of strategies.
- 2. For each time step *t*:
 - (a) Each agent selects an ad to display (using the MAB algorithm) and receives a corresponding reward.
 - (b) Collect reward information from all agents and update the expected reward values.
 - (c) Evaluate and select the best-performing strategies based on cumulative rewards.
 - (d) Apply mutation and crossover operations to generate new agent strategies.
- 3. Return the strategy of the best-performing agent.

2.3. Search advertising optimization problem

Search advertising is an online marketing strategy where advertisers bid on keywords entered by users into search engines, allowing their ads to appear prominently in relevant search results. This targeted approach is designed to deliver ad content to users actively seeking related information, thereby increasing click-through and conversion rates. However, search advertising faces several complex challenges, including estimation bias, data noise, and limited user interaction data, all of which can hinder advertisers' ability to accurately assess ad effectiveness and optimize budget allocation. To address these challenges, MAB algorithms have been introduced into the optimization process of search advertising. MAB algorithms offer a flexible mechanism that balances exploration and exploitation by dynamically allocating budget across various ad options and adjusting strategies based on real-time data. Specifically, MAB algorithms enable advertisers to effectively identify and promote the best-performing

ads by maintaining a balance between exploration (testing different ads to gather more comprehensive data) and exploitation (prioritizing ads with known high effectiveness). This approach helps reduce the impact of estimation bias by continuously collecting and analyzing click and conversion data, allowing MAB algorithms to update their evaluation of ad performance and adjust ad allocations in real time. Such an adaptive method ensures that advertising budgets are efficiently utilized, directing more traffic towards ads with higher conversion rates, and thereby maximizing the advertiser's return on investment. Additionally, MAB algorithms can handle the interactions between different ads and account for the temporal variability of ad effectiveness, further enhancing the overall efficiency of search advertising. Through this innovative optimization strategy, advertisers can improve ad performance and achieve higher user acquisition rates and conversion rates without incurring additional costs[5].

1. Estimation Bias in MAB Algorithms: • In MAB algorithms, estimation bias primarily arises from the overestimation of rewards for arms (ads) that are selected fewer times. This bias can lead to incorrect conclusions about which ads are most effective, thereby reducing the overall performance of the advertising strategy.

2. Reward Estimation:

- Let a_t denote the action taken at time t (e.g., showing a specific ad). The true expected reward for action a is denoted by $\mu(a)$, and the estimated reward is denoted by $\hat{\mu}_t(a)$.
- The reward estimation for each arm (ad) is updated according to the following equation:

$$\hat{\mu}_t(a) = \frac{1}{n_t(a)} \sum_{i=1}^{n_t(a)} r_i(a)$$
(3)

where:

- $n_t(a)$ represents the number of times action a has been selected up to time t.
- $r_i(a)$ is the observed reward when action *a* is selected at time *i*.

3. Estimation Bias:

• The estimation bias $B_t(a)$ at time t for arm a is defined as the difference between the expected value of the estimated reward and the true expected reward:

$$Bt(a) = E[\mu t(a)] - \mu(a)$$
(4)

• This bias is influenced by the variance of the reward distribution and the number of times the arm has been pulled.

4. Methods to Reduce Estimation Bias:

(a) Bias-Corrected Reward Estimation:

• One approach to reduce estimation bias is to adjust the reward estimation by introducing a bias correction term. The corrected estimate $\hat{\mu}^{corrected}_t(a)$ is given by:

$$\hat{\mu}_t^{corrected}(a) = \hat{\mu}_t(a) - \frac{\sigma^2(a)}{n_t(a)}$$
(5)

where:

- $-\sigma^2(a)$ is the variance of the reward distribution for arm *a*.
- This correction term compensates for the overestimation that occurs when the arm is selected a limited number of times.

(b) Confidence Interval Adjustment:

• Another approach to addressing estimation bias is to adjust the confidence intervals used in the decision-making process. A more conservative method is to widen the confidence intervals to reflect the uncertainty in the reward estimation. The confidence interval $C_l(a)$ for arm *a* can be adjusted as follows:

$$C_t(a) = \hat{\mu}_t(a) \pm \sqrt{\frac{2\sigma^2(a)\log(t)}{n_t(a)}}$$
(6)

• This adjustment ensures that the algorithm takes into account the higher uncertainty associated with arms that have been explored less frequently.

3. Synthesis and Analysis

The existing literature on multi-armed bandit algorithms in online advertising demonstrates a consensus on their potential to optimize ad performance by effectively balancing exploration and exploitation. The research reveals that different MAB algorithms, such as Thompson sampling, UCB, and so on, have been successfully applied to various aspects of the online advertising process, including creative development, platform selection, and bidding strategies. These studies consistently show that MAB algorithms can dynamically adapt to changing market conditions and user preferences, leading to improved targeting and increased conversion rates.

3.1. Strengths and Weaknesses of Different MAB Algorithms

In advertising optimization, different multi-armed bandit (MAB) algorithms exhibit various strengths and weaknesses, each suited to different scenarios. The Upper Confidence Bound (UCB) algorithm balances exploration and exploitation by selecting the option with the highest upper confidence bound. It has a strong theoretical foundation and performs well in stable environments, but it can be slow to adapt in rapidly changing advertising settings. It is suitable for long-term optimization strategies where the advertising environment remains relatively stable. Thompson Sampling (TS) excels in dynamic environments due to its probabilistic sampling approach, which continually updates estimates of each ad's effectiveness. TS often outperforms UCB in environments with frequent changes but involves higher computational complexity, which may be a bottleneck in large-scale tasks. It is well-suited for real-time bidding systems and personalized ad recommendations. The Explore-then-Commit (ETC) strategy is straightforward, involving a phase of exploration followed by a commitment to the currently best-performing ad. It quickly converges in stable environments but performs poorly if the environment changes after the commitment phase. It is best used when a significant amount of exploration is needed initially, followed by rapid decision-making in a stable context. Online Reinforcement Learning (ORL) algorithms are highly adaptable, continually learning and adjusting to changes in the environment, and focus on long-term cumulative rewards. However, ORL algorithms are complex, requiring substantial computational resources and data, making them challenging to implement and tune. They are suitable for long-term advertising strategies that need to continually learn and adapt to changing user behaviors. The Embedded MAB Algorithm with Agent Evolution combines MAB algorithms with agent evolution, allowing continuous optimization through agent self-evolution. This approach maintains high adaptability in diverse and complex advertising environments but is complex to implement and tune. It is particularly effective in scenarios requiring long-term optimization and consideration of multiple objectives. In summary, UCB and ETC algorithms are effective in stable environments, making them suitable for long-term optimization. TS and ORL algorithms excel in dynamic settings, responding quickly to changes, while the Embedded MAB algorithm performs well in complex, multi-objective environments.

Proceedings of CONF-MLA 2024 Workshop: Semantic Communication Based Complexity Scalable Image Transmission System for Resource Constrained Devices DOI: 10.54254/2755-2721/83/2024GLG0063

3.2. Success Stories in Applying MAB Algorithms to Online Advertising

Several success stories have demonstrated the practical effectiveness of MAB algorithms in online advertising, which will be explored in detail in the following sections. DAC-MAB was successfully implemented on a leading Demand Side Platform (DSP) that processes billions of bidding requests daily, selecting thousands of ads from hundreds of campaigns. Experiments demonstrated that DAC-MAB effectively identified high-quality ad arms, significantly increasing ad revenue. It outperformed traditional multi-armed bandit algorithms like UCB, Thompson Sampling, and Contextual MAB, particularly in managing large data volumes and adapting to dynamic ad environments, showcasing its unique advantages [10]. In a related success story, MAB methods were used to improve customer acquisition rates through real-time ad placement. A large-scale experiment with a major retail bank showed that the implementation of Thompson Sampling led to an 8% increase in customer acquisition rates without additional costs, illustrating the power of MAB in optimizing online advertising strategies [9]. Another notable case involved the application of real-time bidding (RTB) and optimization algorithms. The online advertising market grew significantly in 2022, reaching \$209.7 billion, largely driven by RTB's use of user data and auction mechanisms. The introduction of first-price auctions and the UCB algorithm demonstrated significant improvements. In the Amazon AuctionGym simulation environment, UCB reduced cumulative regret and increased ad revenue by approximately 90% compared to baseline algorithms [7]. Finally, Gorthaur-EXP3 showcased its exceptional performance across various recommendation systems. It exhibited stability and effectiveness on the Control dataset, adaptability and accuracy in the RS-ASM dataset, and high efficiency with large-scale food recommendations. Gorthaur-EXP3 also outperformed other algorithms in diversity and accuracy on the Movie Lens dataset and improved recommendation diversity in the Jester dataset. Overall, Gorthaur-EXP3 demonstrated significant effectiveness and versatility across different recommendation scenarios, highlighting its broad application potential [6].

3.3. Challenges and Limitations of MAB Algorithms

Despite the successes of MAB algorithms, there are several limitations and challenges that need to be addressed. These include the computational complexity involved in real-time decision-making, particularly for algorithms like Thompson Sampling, which require frequent updating of posterior distributions. Moreover, the dynamic nature of online advertising environments poses a challenge for algorithms that may not adapt quickly enough to changing conditions, such as UCB. Additionally, the scalability of these algorithms when dealing with large datasets and high-dimensional feature spaces remains a critical area of concern.

3.4. Future Research Directions in MAB for Online Advertising

Future research in this domain should focus on developing more efficient and scalable algorithms that can handle the complexities of real-time online advertising. There is also a need for further exploration of hybrid models that combine the strengths of different MAB algorithms to optimize performance across various scenarios. Moreover, enhancing the adaptability of these algorithms to rapidly changing environments, possibly through the integration of machine learning techniques, could lead to significant improvements in ad selection and revenue generation. Finally, addressing the ethical implications of automated decision-making in online advertising, such as issues related to fairness and transparency, will be crucial as these technologies continue to evolve.

4. Conclusion

This study provides a comprehensive analysis of MAB algorithms and their significant role in optimizing online advertising. MAB algorithms like Thompson Sampling and UCB have proven effective in real-time decision-making environments by balancing exploration and exploitation. Thompson Sampling's probabilistic approach adapts well to dynamic environments, while UCB's deterministic nature minimizes cumulative regret, allowing for quicker convergence on optimal ad

placements. These algorithms manage large datasets and respond to changing conditions, showcasing their superiority over traditional methods.

A key success story is the implementation of the DAC-MAB framework on a leading Demand Side Platform (DSP), which processes billions of bidding requests daily. DAC-MAB successfully identified high-quality ad arms, significantly increasing ad revenue. Compared to traditional MAB algorithms like UCB and Thompson Sampling, DAC-MAB performed better in handling large data volumes and adapting to dynamic ad environments. This case highlights the practical benefits of MAB algorithms in a competitive industry.

However, this study also acknowledges certain limitations, particularly the computational complexity of MAB algorithms like Thompson Sampling, which require frequent updates. The adaptability of these algorithms in rapidly changing environments is another area for further research. These challenges indicate that while MAB algorithms are effective, there is room for improvement, especially in scalability and efficiency.

The findings of this study have important implications for future research in online advertising optimization. By analyzing the strengths and limitations of various MAB algorithms, this study lays the groundwork for developing more advanced models. Future research could focus on hybrid models that combine the strengths of different MAB algorithms or integrate machine learning techniques to enhance adaptability and performance in dynamic environments.

The main contribution of this study lies in its detailed examination of MAB algorithms and their practical applications in online advertising. It advances the understanding of these algorithms and provides practitioners with strategies to improve ad placement efficiency and revenue. For researchers, this study fills gaps in the existing literature by offering detailed comparisons and real-world case studies of MAB algorithms.

Looking forward, future research should address the identified limitations, particularly the need for scalable and efficient algorithms. Additionally, exploring the ethical implications of automated decision-making in online advertising, such as fairness and transparency, will be crucial as these technologies evolve.

References

- Kraus S,Jones P,Kailer N,Weinmann A,Chaparro-Banegas N and Roig-Tierno N 2021 Digital transformation:an overview of the current state of the art of research. Sage Open 11(3):21582440211047576
- [2] Zhang S 2024 Utilizing reinforcement learning bandit algorithms in advertising optimization. Highlights in Science, Engineering and Technology. 94:pp 195-200.
- [3] Baardman L,Fata E,Pani A and Perakis G 2019 Learning optimal online advertising portfolios with periodic budgets. Available at SSRN 3346642.
- [4] Qiu R,Ji W 2021 An embedded bandit algorithm based on agent evolution for cold-start problem. Int J Crowd Sci. 5(3):228–38.
- [5] Xu M,Qin T and Liu TY 2013 Estimation bias in multi-armed bandit algorithms for search advertising. Adv Neural Inf Process Syst. 26.
- [6] Gutowski N, Amghar T, Camp O and Chhel F 2021 Gorthaur-exp3: bandit-based selection from a portfolio of recommendation algorithms balancing the accuracy-diversity dilemma. Inf Sci. 546:378–96.
- [7] Bonalumi M 2022 An online learning algorithm for real-time bidding.
- [8] Liu B 2023 Based on intelligent advertising recommendation and abnormal advertising monitoring system in the field of machine learning. Int J Comput Sci Inf Technol. 1(1):17–23.
- [9] Schwartz EM,Bradlow ET and Fader PS 2017 Customer acquisition via display advertising using multi-armed bandit experiments. Mark Sci.36(4):500–22.
- [10] Yang H and Lu Q 2016 Dynamic contextual multi-arm bandits in display advertisement. In: 2016 IEEE 16th International Conference on Data Mining (ICDM). p. 1305–10.
- [11] Nuara A 2020 Machine learning algorithms for the optimization of internet advertising campaigns.

[12] Mussi M 2023 Online learning methods for pricing and advertising.