Lessons Learned from Auto-Bidding in the Large-Scale Auctions Competition at NeurIPS 2024

Zhilin Zhang ZHANGZHILIN.PT@ALIBABA-INC.COM
Shuai Dou DOUSHUAI.DS@ALIBABA-INC.COM
Yusen Huo HUOYUSEN.HUOYUSEN@ALIBABA-INC.COM
Alibaba Group

Zhijian Duan ZJDUAN@PKU.EDU.CN
Ningyuan Li LININGYUAN@PKU.EDU.CN
Peking University

Chuan Yu YUCHUAN.YC@ALIBABA-INC.COM
Alibaba Group

Xiaotie Deng XIAOTIE@PKU.EDU.CN
Peking University

Jian Xu* XIYU.XJ@ALIBABA-INC.COM
Bo Zheng BOZHENG@ALIBABA-INC.COM
Alibaba Group

Editors: Tao Qin, Kun Zhang and Jes Frellsen

Abstract

Decision-making in large-scale games is an essential research area in artificial intelligence, with significant real-world impact. An agent faces the critical task of making high-frequency strategic decisions in an uncertain and competitive environment. This environment is characterized by substantial randomness and rapidly changing strategies from numerous competitors. However, the shortage of large-scale, realistic game systems and datasets has hindered research progress in this area. To provide opportunities for in-depth research on this valuable problem, we have presented the Auto-Bidding in Large-Scale Auctions Competition, derived from online advertising, at NeurIPS 2024. This competition focuses on bidding decision-making in uncertain and competitive games. It specifically includes the AIGB (AI-Generated Bidding) track, which emphasizes the innovation of generative algorithms in the field of auto-bidding. This competition has attracted over 1,500 teams from around the world. After six months of intense competition, participants have proposed numerous innovative and forward-looking solutions, significantly promoting research and technological development in this field. In this competition, we have open-sourced AuctionNet, a benchmark for bidding decision-making in large-scale ad auctions derived from a real-world online advertising platform. We have also pre-generated a substantial dataset comprising 10 million ad opportunities, 48 diverse auto-bidding agents, and over 500 million auction records. Performance evaluations of baseline algorithms, such as linear programming, reinforcement learning, and generative models, are also presented as part of AuctionNet. AuctionNet has powered the competition at NeurIPS 2024. We believe that the proposed competition and AuctionNet provide opportunities for more researchers to gain insights and conduct research in this field, driving technical innovation for both academic research and real-world practical applications. Code: https://github.com/alimama-tech/AuctionNet.

* Corresponding author.

© 2026 Z. Zhang, S. Dou, Y. Huo, Z. Duan, N. Li, C. Yu, X. Deng, J. Xu & B. Zheng.

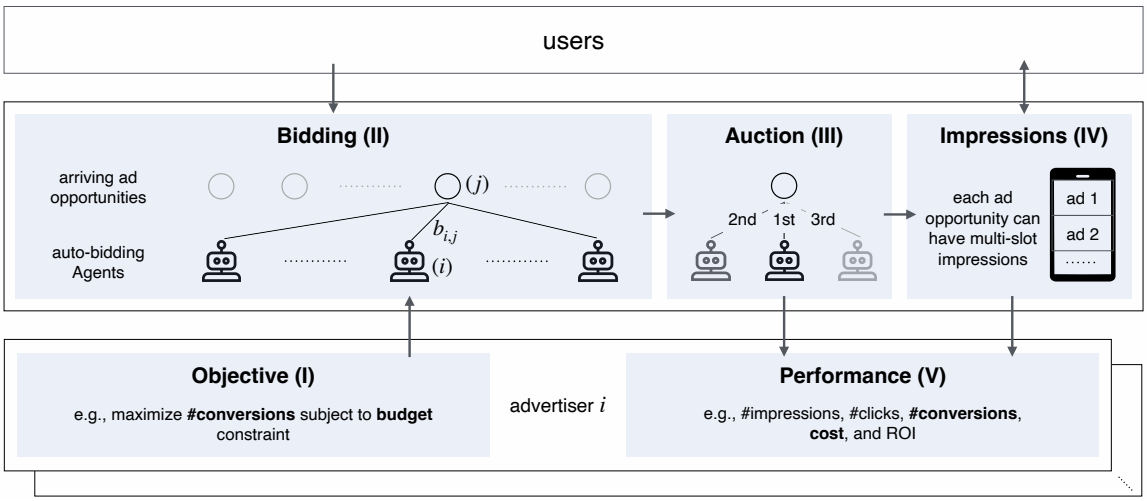


Figure 1: Overview of typical large-scale online advertising platform. Numbers 1 through 5 illustrate how an auto-bidding agent helps advertiser i optimize performance. For each advertiser's unique objective (I), auto-bidding agent make bid decision-making (II) for continuously arriving ad opportunities, and compete against each other in the ad auction (III). Then, each agent may win some impressions (IV), which may be exposed to users and potentially result in conversions. Finally, the agents' performance (V) will be reported to advertisers.

1. Introduction

Decision-making is a key research area in artificial intelligence with wide real-world applications, especially in large-scale games such as online advertising (Zhang et al., 2022), financial market (Hambly et al., 2023), e-commerce (Zhu et al., 2024), and energy trading (Wu et al., 2023). In large-scale games, an agent needs to make high-frequency strategic decisions to achieve its objective under resource constraints in an uncertain and competitive environment, characterized by significant randomness and rapidly changing strategies of massive competitors. However, the shortage of **large-scale**, **realistic** game systems and datasets for research purposes, caused by restricted data access within the industry, poses a challenge for academic researchers to study this problem and verify their strategies, hindering technological advancement in the field. Therefore, we believe that establishing a platform to provide opportunities for in-depth research on decision-making within uncertain, changing, and competitive games could capture the interest of researchers from the NeurIPS and broader artificial intelligence community, while also advancing the development of the related fields.

Therefore, we introduce the Auto-Bidding in Large-Scale Auctions challenge, a typical large-scale game-based competition derived from online advertising, which reached a market size of $626.8 billion in 2023 (Online Advertising Spending Worldwide, 2024). As shown in Figure 1, the auto-bidding agent sequentially bids for each impression across a large number of continuously arriving opportunities to maximize performance, adhering to certain

constraints on behalf of the advertiser. The key challenge is how to make effective bidding decisions in large-scale auctions, which requires effectively perceiving the **changes** in competitors' strategies and the **randomness** of opportunities' arrivals. In our competition, each participant is tasked with implementing an auto-bidding agent that considers the given optimization objective and constraints. They need to compete against numerous agents that employ confidential strategies provided by other participants or organizers. Besides dealing with the complexity of large-scale auctions, participants also need to address essential challenges deserving of research attention in real-world scenarios.

The competition focuses on intelligent decision-making technologies in uncertain and competitive games, which hold significant scientific research and industrial application value. The AIGB track is particularly challenging yet forward-looking. The exploration of Transformer-based and Diffusion Model-based methods is expected to promote the comprehensive implementation of generative technologies in the bidding domain. We believe that, with the continuous exploration and refinement by numerous practitioners, AIGB technologies will mature and address issues that traditional methods cannot solve. We offer a variety of baseline methods targeted at different tracks. In the General track, we provide reinforcement learning-based approaches, which are widely adopted in industry and have achieved significant results. In the AIGB track, we innovatively introduce methods based on Decision-Transformer and Decision-Diffusion, applying the latest AI technologies to the bidding problem and providing participants with a solid foundation for exploring innovative solutions.

This competition has received strong support from Alibaba, which owns the leading online advertising platform in China and is equipped with substantial resources. For the competition platform, we have developed a standardized ad auction system that allows numerous participants to evaluate their strategies' performance in large-scale games. This system effectively reproduces dynamic competition among agents, impression opportunity arrival patterns, and several essential industry characteristics. Additionally, we have established the competition website on Tianchi, which has successfully hosted over 500 competitions since 2014.

Since decision-making in large-scale games is a compelling research area in artificial intelligence with substantial influence on applications, our competition attracts a wide range of participants from related fields, including researchers and practitioners in machine learning, game theory, and data science. Researchers working on large language models (LLMs) and foundation agents will also be attracted, as these technologies play an increasingly active role in decision-making fields (Ajay et al., 2024; Nottingham et al., 2023). Our competition attracted 1,511 teams from 22 countries and regions, collectively submitting nearly ten thousand model iterations. After six months of intense competition, participants proposed numerous innovative and forward-looking solutions, significantly advancing research and technological development in this field. Additionally, we are formally open-sourcing the world's first standardized automatic bidding benchmark: AuctionNet. AuctionNet is designed for research on bid decision-making in large-scale advertising auctions and comprises three components: an advertising auction environment, a pre-generated dataset based on this environment, and performance evaluations of several baseline bidding decision algorithms. The dataset includes sequential trajectories of 48 different agents bidding against each other, with over 500 million records and a size of 80GB.

Our proposed competition offers opportunities for researchers to gain insights and conduct in-depth research in large-scale decision-making, thereby driving technical innovation in both academic research and real-world applications. After the competition, we have **open-sourced** the related frameworks and datasets, which further promote a broad impact in various fields.

2. Competition

2.1. Tasks and Application Scenarios

Participants are tasked with designing and implementing an auto-bidding agent that is capable of making efficient bidding decisions in a large-scale auction while fulfilling advertisers' specified objectives to optimize performance.

Specifically, the task in our competition is based on a significant and practical problem encountered in real-world scenarios called Target CPA (Cost Per Action) (About Target CPA bidding, 2024). The problem has been slightly optimized for academic inquiry. Achieving favorable results requires additional efforts to address challenges such as multiple ad slots and variance in conversion prediction, beyond those posed by complex large-scale auctions. This issue remains unsolved in both the academic and professional realms, making it a key research field.

Formally, given an advertiser j's budget B and the desired CPA C, the agent bids for N impression opportunities within an advertising period. The objective is to maximize the total conversion volume while ensuring that the realized CPA remains below C at the end of the period. All opportunities arrive sequentially, and the agent bids on each opportunity in turn. Here are the bidding and auction procedures for each opportunity i:

1. **Bidding**. **The agent bids** b_i based on $pvalue_i$[1], which represents the probability of a conversion action occurring when advertiser j's ad is exposed to the customer. Meanwhile, other competing agents simultaneously bid using their individual strategies, denoted by b_i^-, to compete for K ad slots in opportunity i.

2. **Auction**. The ad platform runs a GSP (Generalized Second Price) auction mechanism and returns the auction results to the agents. $x_i(b_i, b_i^-)$ indicates whether to win, $k_i(b_i, b_i^-)$ denotes the ad slot won, and $c_i(b_i, b_i^-)$ represents the cost. x_i, k_i, c_i[2] depends not only on b_i but also on b_i^- .

3. **Impression**. Whether the k_i-th slot is exposed to the customer is determined by a random variable defined as $E_i \sim \text{Bernoulli}\big(exposure_{k_i}\big)$, where $exposure_{k_i}$ is the probability of exposure for the slot k_i. The actual conversion occurrence is also a random variable defined as $V_i \sim \text{Bernoulli}\big(value_i\big)$, where $value_i \sim \mathcal{N}(pvalue_i, \sigma_i^2)$, and σ_i represents the variance, e.g., the uncertainty in prediction. If an ad in the slot is not exposed, the advertiser doesn't need to pay the cost and the customer won't make a conversion on the ad.

1. For simplicity, we usually omit the index j of the advertiser where it does not cause ambiguity.
2. For simplicity, $x_i(b_i, b_i^-)$ is generally denoted as x_i, so as k_i, c_i .

Therefore, this task can be formalized as follows:

$$\begin{aligned} \max_{b_1,\cdots,b_N} \quad & \sum_{i=1}^{N} x_i(b_i, b_i^-) \cdot E_i \cdot V_i \\ \text{s.t.} \quad & \sum_{i=1}^{N} x_i(b_i, b_i^-) \cdot E_i \cdot c_i(b_i, b_i^-) \le B, \\ & \frac{\sum_{i=1}^{N} x_i(b_i, b_i^-) \cdot E_i \cdot c_i(b_i, b_i^-)}{\sum_{i=1}^{N} x_i(b_i, b_i^-) \cdot E_i \cdot V_i} \le C. \end{aligned}$$

The realized CPA of the advertiser j is $CPA = \frac{\sum_i x_i(b_i, b_i^-) \cdot E_i \cdot c_i(b_i, b_i^-)}{\sum_i x_i(b_i, b_i^-) \cdot E_i \cdot V_i}$.

challenge. This task is not a traditional constrained optimization problem because when bidding for each opportunity, it is impossible to know all future opportunities in advance due to the randomness of arriving patterns, making it hard to obtain a closed-form analytical solution. Advertising is a competitive game where multiple agents bid simultaneously using their confidential bidding strategies. The game is also dynamic, with competitors continually adapting their strategies. Therefore, auto-bidding agents need to perceive the dynamic gaming environment, model the connection between bid and performance, and then carefully bid for each impression opportunity by taking into account the preceding bids and performance. Additionally, participants need to consider how to adhere to the CPA constraint, given complex features such as the uncertainty of prediction and sparse data. In this competition, we'll have 2 tracks. The first is the classic track, where any optimization method is allowed. The second is the AIGB track, encouraging the use of generative models for modeling.

2.2. Metrics

The auto-bidding agent aims to maximize conversion volume while satisfying the CPA constraint set by the advertiser. If the realized CPA exceeds the advertiser's desired CPA C, a penalty will be applied. The detailed evaluation metrics are defined as follows:

$$Score = \mathcal{P}(CPA; C) \cdot \sum_i x_i \cdot E_i \cdot V_i. \tag{1}$$

In real-world scenarios, if the advertiser's total costs exceed the budget during the advertising period, the ad platform typically stops the agent's bidding for the remaining opportunities. Therefore, the budget constraint can always be satisfied. The current evaluation metric focuses solely on the CPA constraint, where the penalty function $\mathcal{P}(CPA; C)$ for exceeding the CPA constraint is defined as:

$$\mathcal{P}(CPA; C) = \min\left\{\left(\frac{C}{CPA}\right)^{\beta}, 1\right\}$$

The penalty function is defined with a hyperparameter $\beta > 0$, typically set to 3. $\mathcal{P}(CPA; C)$ implies that the penalty is incurred only when $CPA > C$.

In each evaluation, the agent implemented by the participant is required to bid on behalf of one designated advertiser, given a specific budget, CPA, and other settings. To fully

evaluate the performance of the agent across various advertisers, we will run it multiple times in the auction system using different advertiser profiles and periods, and then average the results as the final score. During evaluation, profiles and strategies of competitors' auto-bidding agents are kept confidential.

2.3. Competition Stages

The competition consists of two rounds:

Preliminary Round Participants need to implement their agents to compete against the pre-trained agents provided by the organizer. Each participant's evaluation is conducted separately and independently. All participants are consistently ranked according to their evaluation scores, with the top m advancing to the finals. The parameter m will be set according to participant numbers to guarantee sufficient finalists.

Final Round In the final round, advancing participants need to further optimize their agents. Final rankings will be determined using the same evaluation metrics and methods as in the preliminary round. The winning teams will be determined by the results of the final round.

3. AuctionNet

3.1. Ad Auction Environment

To ensure the quality of the competition and to comprehensively demonstrate real-world large-scale games, we have developed a standardized auction system specifically for the competition, as shown in Figure 2. This system effectively reproduces dynamic competition among agents, randomness of impression opportunity arrivals, and several essential industry features deserving of research attention. To better simulate a large-scale auction, a huge amount of impression opportunities will be fed into the system and configured with hundreds of competing agents. These agents are equipped with different confidential bidding strategies developed by the organizers or submitted by participants.

To simplify the auto-bidding process for participants, we divide impression opportunities in an advertising period into T decision time steps. Given the objective, the auto-bidding agent sequentially bids at each step, using the results from step t and prior historical information to refine its strategy for step $t+1$. This approach enables agents to continuously optimize their bidding strategies in order to adapt to the changing environment. Within each step, all impression opportunities are executed independently and in parallel. At the end of the period, the system provides the final performance for the agent.

Regarding limitations, while the generated data in the AuctionNet environment and the real-world data are similar in general, there are biases in some details, and the performance of the generative model can be improved.

3.2. Pre-Generated Dataset

The dataset is derived from game data generated within the environment, where numerous auto-bidding agents compete against each other. We have pre-generated large-scale game data to assist researchers in gaining deeper insights into the auction ecosystem. This data can be used to model the environment and to train the auto-bidding agents effectively.

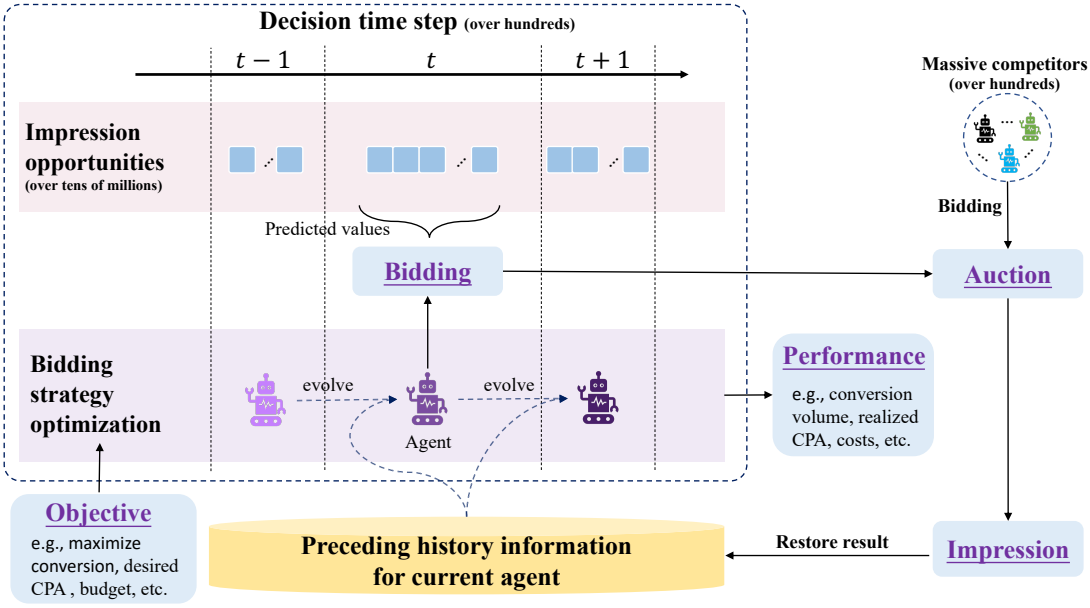


Figure 2: Overview of the standardized ad auction system. The auto-bidding agent sequentially bids for impression opportunities at each step, evolving based on the preceding information.

The dataset contains 10 million ad opportunities, including 21 advertising episodes. Each episode contains more than 500,000 ad opportunities, divided into 48 steps. Each opportunity includes the top 48 agents[3] with the highest bids. The dataset comprises over 500 million records, totaling 80 GB in size. Each record includes information such as the predicted value, bid, auction, and impression results, among other details.

3.3. Baselines

According to research (He et al., 2021), the optimal bid for opportunity i in the single-slot GSP mechanism is given by the formula:

$$b_i = \alpha \cdot C \cdot pvalue_i$$

Here, C denotes the CPA constraint, while α represents the parameter that the auto-bidding agent needs to optimize. In order to adapt to changes in competitors' agents and randomness of opportunity arrival, α must be dynamically adjusted at each step based on previous performance within a period. This constitutes a typical sequential decision-making problem to determine α, which can be addressed using decision-making algorithms.

We offer a variety of baseline methods targeted at different tracks. In the General track, we provide reinforcement learning-based approaches, which are widely adopted in the

3. Real-world data show that 48 agents can ensure competitive pressure for auto-bidding agent training.

industry and have achieved significant results. In the AIGB track, we innovatively introduce methods based on Decision-Transformer and Decision-Diffusion, applying the latest AI technologies to the bidding problem and providing participants with a solid foundation for exploring innovative solutions.

Reinforcement Learning-Based Baseline The automatic bidding problem is essentially a sequential decision-making problem; therefore, reinforcement learning (RL) has become the mainstream approach for addressing this issue (He et al., 2021; Li et al., 2024). Traditional RL methods rely on interactions with the environment, but in the context of advertising bidding, such interactions can directly affect the advertiser's budget consumption, and direct online exploration may lead to economic losses. Offline reinforcement learning has gained widespread application in bidding scenarios by leveraging historical data for training, thus avoiding direct interaction with the online environment. We provide two offline reinforcement learning methods: Behavior Cloning (BC) and Implicit Q-Learning (IQL). BC is the simplest imitation learning method, whose core idea is to directly mimic the behavioral policy in historical data to minimize the discrepancy between predicted bids and actual bids. IQL is a conservative offline RL algorithm characterized by learning the Q function in an implicit manner, which prevents the overestimation issues often associated with explicit maximization and thereby enhances the stability and reliability of the algorithm.

Decision-Diffusion Baseline Decision-Diffusion is a simplified implementation of the AIGB (Guo et al., 2024) algorithm proposed by Alibaba, adopting a "Planning and Control" architecture. In the planning phase, this method treats the bidding, optimization objectives, and constraints as a joint probability distribution and employs a diffusion model to learn a conditional distribution of future states based on the optimization objectives and constraint items. In the control phase, the historical states and future states generated during the planning phase are input, producing the corresponding bids as output. AIGB views the automatic bidding problem as a generative sequential decision-making problem, directly capturing the correlation between the optimization objective and the entire delivery trajectory, thus overcoming the bottleneck of traditional RL in handling long sequences with sparse rewards.

Decision-Transformer Baseline The Decision-Transformer also views the automatic bidding problem as a generative sequential decision-making problem and employs a Transformer as its core architecture. Unlike Decision-Diffusion, it features an end-to-end algorithmic structure. Specifically, this method encodes the sequences of states, actions, and returns from historical trajectories, incorporating positional encoding information to form the input sequence. Subsequently, it captures long-term dependencies within the sequence through a standard Transformer structure, which includes multiple layers of self-attention mechanisms and feedforward networks. Finally, an action prediction head is added to the output layer of the Transformer, generating the action for the next time step.

4. Result

Our competition attracted a total of 1,511 teams from 22 countries and regions, with nearly ten thousand iterations of model submissions. Participants performed in-depth optimizations on baseline models for the bidding problem. The following sections provide detailed descriptions of various baseline optimization strategies.

4.1. Optimization of Reinforcement Learning

Model Optimization: Participants implemented targeted optimizations on reinforcement learning models for the bidding problem. For instance, they introduced features such as the distribution of the number of winning bids in different positions and the vectorization of advertiser category IDs to enhance the model's perception of historical information. In terms of reward design, continuous rewards incorporating a super-cost penalty term were employed to replace discrete rewards, alleviating issues of sparse rewards and enabling the model to better perceive penalties for exceeding limits. To address the instability in training reinforcement learning models, participants utilized grid search for offline training of multiple models, achieving model ensemble through policy combination. Additionally, because the traffic distributions of different advertisers vary, training personalized models for different advertisers has been proven to be an effective strategy.

Data Optimization: The performance of offline reinforcement learning is highly dependent on data quality. We provided bidding logs from all advertisers across various traffic types, and some participants calculated the optimal bids for each advertiser using linear programming, thereby augmenting their training data with high-quality information. Others constructed offline simulation environments using the bidding logs, effectively simulating exposure probabilities for different positions and transforming offline reinforcement learning tasks into online ones.

Strategy Optimization: The uncertainty of traffic presents ample opportunities for participants in strategy design. Some participants devised risk-averse bidding strategies to moderately suppress high-uncertainty traffic. Conversely, other participants chose to increase bids for high-uncertainty traffic when budgets were sufficient and time was limited. Moreover, reasonable traffic control strategies exhibited significant improvements in budget consumption rates, for example, adopting more aggressive bidding when an advertiser has a large remaining budget but few time steps left.

4.2. Optimization of Decision-Diffusion

The Decision-Diffusion model leverages the powerful expressiveness of diffusion generative models to better fit complex data distributions. However, trajectory prediction-based approaches suffer from high computational complexity and require comprehensive modeling of environmental uncertainties, resulting in slow decision-making speeds and diminished performance. The authors of the CleanDiffuser library proposed the DARE (Diffusion Bidding Policy Optimization with Action Relabeling via Ensemble Q-Learning) strategy, which integrates action relabeling with an ensemble Q-function. This method reduces the complexity of fitting the Q-function by establishing a discretized ensemble Q-function table to approximate the expected values of different types of data. By backtracking data states using the trained Q-function, the theoretically optimal actions can be obtained and relabeled. Finally, the optimal actions are fitted using an end-to-end multimodal distribution via the diffusion model. Compared to Decision-Diffusion, DARE directly learns low-dimensional action distributions, avoiding the fitting of excessive and complex trajectory sequence distributions, which effectively reduces redundant information predictions and decreases

the number of sampling steps in the diffusion model, thereby achieving high-frequency decision-making capability.

4.3. Optimization of Decision-Transformer

The Decision Transformer (DT) model, due to its streamlined architecture, has become the optimization choice for the majority of participants in the AIGB track. The optimizations can be categorized into three directions:

Return to Go ($\widehat{R}$) Optimization: At the prediction stage, $\widehat{R}$ serves as the generation condition and its setting directly influences the model's final performance. The simplest optimization method is to differentiate between different advertisers based on their budget and category characteristics, assigning personalized RTG values. Some participants referenced a solution proposed by Google, adjusting the input order from $\widehat{R}, s, a$ to $s, \widehat{R}, a$ at the same moment in time, thereby predicting $\widehat{R}$ first and then using it for action prediction. Additionally, another optimization approach combines the Q/V function approximation from reinforcement learning with DT, introducing additional $Q(s, a)$ or $V(s)$ functions to estimate $\widehat{R}$.

Model Optimization: The original DT paper only predicted actions and calculated losses in the output phase. However, in reality, during the model output phase, information such as rewards, states, and actions at that moment can also be predicted simultaneously. This multi-target prediction strategy can effectively enhance the model's expressiveness. Participants demonstrated through experiments that appropriately adding predictions of current rewards is one of the effective ways to improve model performance. Furthermore, constraint information such as real-time CPA limit rates and advertiser industry category information is crucial for the subsequent predictions of the model. Incorporating these constraints alongside $\widehat{R}, s, a$ as inputs, encoded with a separate embedding layer, has been shown to significantly enhance model performance.

Data Optimization: The effectiveness of the DT model is highly dependent on the quality of the training data. Participants and relevant academic research have shown that a mixed training strategy involving expert samples and non-expert samples yields better results. While expert samples exhibit superior performance, they often provide a singular trajectory; non-expert samples, by contrast, offer diversity that more comprehensively covers the decision space. To address the challenge of implementing expert sample-guided reasoning during the prediction phase of mixed sample training, competition participants successfully introduced sample category embeddings in the embedding layer, uniformly utilizing expert sample embeddings during the prediction phase to realize an expert sample-guided reasoning mechanism. Moreover, data augmentation methods from the general track can also be applied to optimize the DT model.

5. Related Works

The competition we propose is a brand-new event designed as a novel benchmark for large-scale decision-making, rooted in real-world practical applications. It offers a large-scale, realistic auction system and dataset, and features a distinct task with specific evaluation

metrics. This initiative bridges an existing gap in the community for effectively benchmarking research questions in this field. Ultimately, it aims to facilitate in-depth research in the field and spur the development of numerous valuable new research topics.

To our best efforts, we identified two past competitions (Jordan and Wellman, 2009; Schain and Mansour, 2013) related to the optimization of online advertising, held more than a decade ago. These competitions were narrowly focused on distinct advertising application issues and conducted in small-scale simulated environments with fewer than 10 participants. In contrast, we distill fundamental research questions from the area of online advertising, particularly those concerning decision-making in large-scale auctions. To this end, we have developed a scalable competition platform designed to accommodate thousands of participants.

6. Conclusion

The Auto-Bidding in Large-Scale Auctions challenge has significantly advanced research in decision-making within artificial intelligence by addressing critical complexities in high-frequency bidding scenarios. With over 1,500 participating teams, the competition has fostered innovation and collaboration, generating numerous forward-looking solutions. The introduction of AuctionNet as a benchmark for bid decision-making, along with the extensive dataset provided, equips researchers with valuable resources to explore generative algorithms and advanced methodologies. This initiative not only drives technological advancements in academic research but also enhances practical applications in industries such as online advertising and beyond.

References

About Target CPA bidding. https://support.google.com/google-ads/answer/6268632. 2024.

Online Advertising Spending Worldwide. https://www.statista.com/statistics/237974/online-advertising-spending-worldwide. 2024.

Anurag Ajay, Seungwook Han, Yilun Du, Shuang Li, Abhi Gupta, Tommi Jaakkola, Josh Tenenbaum, Leslie Kaelbling, Akash Srivastava, and Pulkit Agrawal. Compositional foundation models for hierarchical planning. *Advances in Neural Information Processing Systems*, 36, 2024.

Jiayan Guo, Yusen Huo, Zhilin Zhang, Tianyu Wang, Chuan Yu, Jian Xu, Yan Zhang, and Bo Zheng. Aigb: Generative auto-bidding via diffusion modeling. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2993–3001, 2024.

Ben Hambly, Renyuan Xu, and Huining Yang. Recent advances in reinforcement learning in finance. *Mathematical Finance*, 33(3):437–503, 2023.

Yue He, Xiujun Chen, Di Wu, Junwei Pan, Qing Tan, Chuan Yu, Jian Xu, and Xiaoqiang Zhu. A unified solution to constrained bidding in online display advertising. In *Proceedings*

of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pages 2993–3001, 2021.

Patrick R. Jordan and Michael P. Wellman. Designing an ad auctions game for the trading agent competition. In *International Workshop on Agent-Mediated Electronic Commerce and Trading Agents Design and Analysis*, 2009. URL https://api.semanticscholar.org/CorpusID:16408158.

Haoming Li, Yusen Huo, Shuai Dou, Zhenzhe Zheng, Zhilin Zhang, Chuan Yu, Jian Xu, and Fan Wu. Trajectory-wise iterative reinforcement learning framework for auto-bidding. *Proceedings of the ACM Web Conference 2024*, 2024. URL https://api.semanticscholar.org/CorpusID:267898008.

Kolby Nottingham, Yasaman Razeghi, Kyungmin Kim, JB Lanier, Pierre Baldi, Roy Fox, and Sameer Singh. Selective perception: Learning concise state descriptions for language model actors. In *NeurIPS 2023 Foundation Models for Decision Making Workshop*, 2023.

Mariano Schain and Yishay Mansour. Ad exchange – proposal for a new trading agent competition game. In *Agent-Mediated Electronic Commerce. Designing Trading Strategies and Mechanisms for Electronic Markets*, pages 133–145, 2013. ISBN 978-3-642-40864-9.

Yuxin Wu, Tianyang Zhao, Haoyuan Yan, Min Liu, and Nian Liu. Hierarchical hybrid multi-agent deep reinforcement learning for peer-to-peer energy trading among multiple heterogeneous microgrids. *IEEE Transactions on Smart Grid*, 2023.

Wei Zhang, Yanjun Han, Zhengyuan Zhou, Aaron Flores, and Tsachy Weissman. Leveraging the hints: Adaptive bidding in repeated first-price auctions. *Advances in Neural Information Processing Systems*, 35:21329–21341, 2022.

Yiheng Zhu, Yang Zhan, Xuankun Huang, Yuwei Chen, Jiangwen Wei, Wei Feng, Yinzhi Zhou, Haoyuan Hu, Jieping Ye, et al. Ofcourse: A multi-agent reinforcement learning environment for order fulfillment. *Advances in Neural Information Processing Systems*, 36, 2024.