Knowledge-informed Bidding with Dual-process Control for Online Advertising

Huixiang Luo[1], Longyu Gao[1], Qianqian Chen[1], Tianning Li[1⋆], Yaqi Liu[2], and Pingchun Huang[2]

[1] Alibaba Health, Hangzhou, China
{luohuixiang.lhx, longyu.gly, fangqi.cqq, tianning.litn}@alibaba-inc.com
[2] Alibaba Health, Beijing, China
{lyq418445, pingchun.hpch}@alibaba-inc.com

Abstract. Bid optimization in online advertising relies on black-box machine learning models that learn bidding decisions from historical data. However, these approaches sometimes fail to replicate human experts' adaptive, experience-driven, and globally coherent decisions. Specifically, they generalize poorly in data-sparse cases because of missing structured knowledge, make short-sighted decisions that ignore long-term interdependencies, and struggle to adapt in out-of-distribution scenarios where human experts succeed. To address this, we propose KBD (Knowledge-informed Bidding with Dual-process control), a novel method for bid optimization. KBD embeds human expertise as inductive biases through the Informed Machine Learning paradigm, uses Decision Transformer (DT) to globally optimize multi-step bidding sequences, and implements dual-process control by combining a fast rule-based PID (System 1) with DT (System 2). Extensive experiments highlight KBD's advantage over existing methods and underscore the benefit of grounding bid optimization in human expertise and dual-process control.

Keywords: Informed Machine Learning · Dual-process Theory · Decision Transformer · Auto-bidding · Online Advertising

1 Introduction

Auto-bidding now dominates digital impression purchasing, projected to exceed $200 billion in 2026 [23]. Advertisers specify high-level performance goals (tROI, tCPA, tCPC) for campaigns, while platforms handle per-impression bidding. Since these goals directly affect revenue (e.g., GMV), calibrating them effectively is crucial. Prior works [3, 20, 21] replace human experts with black-box ML models learned from historical data. However, these models underperform in data-scarce settings, optimize myopically, and generalize poorly to distribution shifts.

To address these limitations, we propose KBD (Knowledge-informed Bidding with Dual-process Control), a two-stage bid optimization method. At the macro (daily) stage, we adopt Informed Machine Learning (IML) paradigm [32] to embed

⋆ Corresponding author.

human expertise into model across hypothesis, algorithm and data levels. The hypothesis level design is central to macro stage bid optimization. We construct a price-volume model with hybrid cognitive architectures [25] to produce base bid values that are robust to data sparsity. At the micro (hourly) stage, we introduce Decision Transformer (DT) [4] to optimize multi-step future rewards, mitigating the short-sightedness of step-wise optimization. To handle abrupt distribution shifts, we draw on dual-process control [24], which posits that human cognition comprises two parallel systems: System 1, a fast, intuitive heuristic process, and System 2, a slow, deliberate reasoning process. We instantiate System 1 as a rule-based PID controller [2] that guides the complicated DT model (System 2) during training, and fuse their outputs to improve robustness in extreme scenarios.

In summary, our contributions are threefold:
(1) We propose KBD, a two stage bid optimization method that couples expert-driven daily calibration with sequential hourly control to optimize long-term bidding rewards.
(2) We improve model performance and robustness with a dual-process controller, where PID (System 1) guides and is fused with DT (System 2) to handle data distribution shifts.
(3) Experiments on two datasets show the effectiveness and broad applicability of KBD across diverse environments.

2 Related Works

2.1 Auto-bidding Strategies in Ad Platforms

Ad platforms offer various auto-bidding strategies to meet advertisers' needs for ad procurement. These strategies eliminate manual per-impression bidding, substantially reducing the complexity of ad procurement and enabling advertisers to get started with minimal effort. However, as these strategies maximize platform revenue, low-value impressions are mixed with high-value ones, prompting advertisers to calibrate performance goals to optimize bids and improve returns. Mainstream auto-bidding strategies fall into three groups: (1) rule-based controllers, e.g., RTBControl [33], MPID [36]; (2) reinforcement learning methods, e.g., USCB [15], MAAB [34]; and (3) generative-model-based methods, e.g., DiffBid [12], GAVE [10]. To cope with the increasingly complex strategies, advertisers begin to use ML models to learn bidding patterns and combine them with human expertise for efficient ad procurement.

2.2 Bid Optimization by Performance Goal Control

Following the 'Predict, then Optimize' (PO) paradigm [9], advertisers commonly adopt a two-stage approach for bid optimization. The first stage centers on a price-volume model that characterizes how bid price varies with volume (e.g. cost, GMV). The model learns from historical data with monotonicity constraints as expert priors to improve model reliability. The second stage formulates an

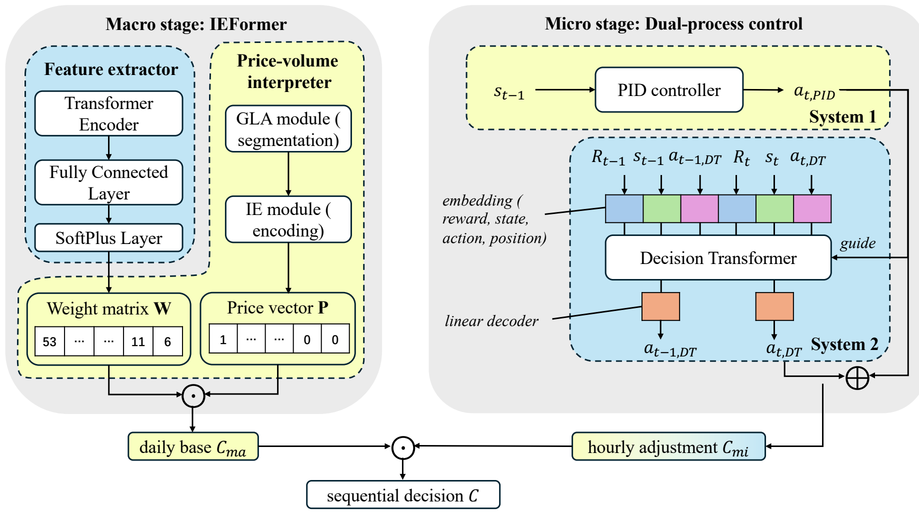

Fig. 1. The architecture of the KBD method. Blue-shaded components denote black-box modules implemented via neural networks, yellow-shaded components indicate rule-based interpretable modules, and mixed-color regions represent their fusion.

optimization problem to derive the optimal bidding decision. PUROS [21] employs monotonic XGBoost for price-volume modeling with single-step optimization, while GCB-safe [3] uses Gaussian processes with smoothness priors. Overall, these first-stage models are black-box ML methods learned mostly from historical data, with limited incorporation of expert knowledge. Moreover, their second-stage optimization is typically myopic, maximizing single-step reward while ignoring how current bids affect future decisions and rewards, which hinders global revenue maximization. Under data distribution shifts caused by sales promotions or new product launches, due to the mismatch between online and historical data distributions, price-volume models exhibit degraded performance. The resulting prediction errors propagate into the second stage, amplifying the unreliability of bidding decisions and potentially causing losses for advertisers.

2.3 Dual-process Theory in Industrial Applications

Industrial implementations of dual-process theory instantiate System 1 as a robust, low-latency heuristic and System 2 as a deliberative planner, fusing them to balance efficiency and performance [22, 13, 14].

3 Problem Formulation

We consider a sequence of M ad impressions arriving within a day, indexed by $i = 1, \ldots, M$. Each impression i is associated with a predicted click-through rate

(pCTR_i), a predicted conversion rate (pCVR_i), a predicted payment amount (ppay_i), and a winning price (wp_i). The advertiser employs the platform-provided tCPA auto-bidding strategy to maximize GMV. Under the generalized second-price (GSP) auction mechanism [8], the strategy submits bid_i and wins the impression if $\mathrm{bid}_i > \mathrm{wp}_i$. Let x_i be the win indicator and $v_i = \mathrm{pCTR}_i \cdot \mathrm{pCVR}_i \cdot \mathrm{ppay}_i$ the expected conversion value of impression i. Given budget B and tCPA target C, the bid optimization problem is formulated as:

$$\max_{\{x_i\}} \quad \sum_{i=1}^{M} x_i \cdot v_i \tag{1}$$

$$\text{s.t.} \quad \sum_{i=1}^{M} x_i \cdot \mathrm{wp}_i \leq B \tag{2}$$

$$\textstyle\sum_{i=1}^{M} x_i \cdot \mathrm{wp}_i \leq C \cdot \sum_{i=1}^{M} x_i \cdot \mathrm{pCTR}_i \cdot \mathrm{pCVR}_i \tag{3}$$

$$x_i \in \{0,1\}, \quad \forall i. \tag{4}$$

Following the MPID framework [36], the optimal bid can be expressed as:

$$\mathrm{bid}_i = \frac{1}{p+q} \cdot v_i + \frac{q}{p+q} \cdot \mathrm{pCTR}_i \cdot \mathrm{pCVR}_i \cdot C \tag{5}$$

where variables $p, q \geq 0$ correspond to the budget and tCPA constraints respectively, governing spending pace and actual CPA. In practice, p and q are tuned by the platform's auto-bidding system (e.g., GAVE [10]), while v_i and ppay_i are estimated via platform ML models.

For conciseness, Eq. (5) can be rewritten as:

$$\mathrm{bid}_i = \lambda_i^0 + \lambda_i^1 \cdot C \tag{6}$$

where $\lambda_i^0 = \frac{1}{p+q} \cdot v_i$ and $\lambda_i^1 = \frac{q}{p+q} \cdot \mathrm{pCTR}_i \cdot \mathrm{pCVR}_i$ are dynamically set by the platform and unobservable to the advertiser. Thus, the advertiser's only control is the tCPA target C, and our objective is to maximize GMV by hourly adjusting C.

4 Methodology

As shown in Figure 1, KBD follows the 'Predict, then Optimize' framework, and refines the task of tuning parameter C into the following formulation:

$$C = C_{ma} \cdot C_{mi} \tag{7}$$

where C_{ma} is a day-level base tCPA from the macro stage model IEFormer (Isotonic-Embedding based Transformer), and C_{mi} is an hour-level adjustment coefficient from the micro stage that combines DT model (System 2) with a PID controller (System 1) to achieve globally optimal bids.

4.1 Macro stage: IEFormer

IEFormer learns the tCPA-cost relationship to establish a near-globally optimal tCPA baseline for bid optimization. Inspired by the Knowledge Integration framework from IML [32], we incorporate human expertise into the model at the hypothesis, algorithm and data levels.

Hypothesis Level We adopt a hybrid architecture [25] comprising a Transformer encoder [31] for feature extraction and a symbolic price-volume interpreter:

- **Connectionist module.** A Transformer encoder distills historical bidding data into a dense ad embedding.
- **Symbolic module.** A price-volume interpreter maps cost to tCPA via tCPA $= \mathbf{P} \cdot \mathbf{W}$, where $\mathbf{W} \in \mathbb{R}^N$ stores marginal tCPA weights over $N = 10$ segments and $\mathbf{P} \in [0,1]^N$ is produced by an Isotonic Embedding (IE) module [27].

The Transformer personalizes the interpreter's parameters ($\mathbf{W}$), while the symbolic structure imposes inductive biases that regularize black-box learning.

A key challenge is avoiding skewed interval allocation (e.g., most samples falling into a few segments), which leads to poor estimation in sparse regions. To address this, we propose an entropy-driven adaptive partitioning strategy: for campaign i with Z_i historical samples, let z_{ij} denote the number assigned to interval j, and define empirical probability $p_{ij} = z_{ij}/Z_i$. We then seek interval boundaries that maximize total information entropy across all campaigns:

$$\max_{S^*} \sum_i \sum_{j=1}^{N} -p_{ij} \log_N(p_{ij}) \tag{8}$$

which encourages near-uniform sample distribution and improves generalization in data-sparse regimes. This optimization is efficiently solved via the Generalized Lloyd Algorithm (GLA) from data compression literature [6], repurposed here as a novel mechanism for structuring interpretable cost segments grounded in information-theoretic principles. Finally, $\mathbf{W}$ is generated from the ad embedding via a fully connected layer, enabling joint training and campaign-specific personalization.

Algorithm Level Domain expertise asserts that the tCPA–cost relationship is monotonic and smooth. We preserve both and further enforce diminishing marginal returns [35]—the rate of tCPA increase slows as cost grows.

Monotonicity is implemented by appending a Softplus activation layer after the fully connected layer in the price-volume interpreter, ensuring that $\mathbf{W}$ is non-negative. Smoothness is enforced by adding a smoothness regularizer L_{smooth} to the Mean Squared Error (MSE) loss to penalize differences between adjacent elements of $\mathbf{W}$, thereby discouraging sharp transitions

$$L_{\text{smooth}} = \sum_i \frac{(w_i - w_{i+1})^2}{w_i \cdot w_{i+1}} \tag{9}$$

To enforce the newly introduced constraint, we introduce a dedicated regularization term L_{margin}

$$L_{\text{margin}} = \sum_i ReLU\left(\frac{w_{i+1}}{step_{i+1}} - \frac{w_i}{step_i}\right) \tag{10}$$

where w_i denotes the i-th element of the weight vector $\mathbf{W}$ and $step_i$ is the width of the corresponding cost interval. This loss penalizes violations of the diminishing marginal returns condition. We incorporate the marginal utility loss L_{margin} into the total objective alongside the prediction loss L_{MSE} and smoothness regularizer L_{smooth}

$$L_{\text{IEFormer}} = L_{\text{MSE}} + \alpha_{\text{smooth}} L_{\text{smooth}} + \alpha_{\text{margin}} L_{\text{margin}} \tag{11}$$

where $\alpha_{\text{smooth}} = 0.5, \alpha_{\text{margin}} = 0.1$.

Data Level To mitigate data scarcity, we transfer knowledge from other strategies (tROI, tCPC) by converting their goals into equivalent tCPA values via the unified eCPM formula (Eq. (12)) [37]:

$$\mathbf{tCPA} = \frac{\mathbf{eCPM}}{\text{pCTR} \cdot \text{pCVR} \cdot 1000} = \frac{ppay}{\mathbf{tROI}} = \frac{\mathbf{tCPC}}{\text{pCVR}} \tag{12}$$

Augmented samples are down-weighted to 0.1 in training to preserve fidelity to native tCPA data.

4.2 Micro stage: Dual-process Control with PID & DT

At the micro stage, we perform hourly tCPA adjustments to maximize GMV. Unlike prior methods that rely on myopic, single-step optimization and ignore inter-temporal dependencies and future ad dynamics, we frame bidding as a sequential decision problem and employ a DT model [4] under the offline reinforcement learning paradigm to optimize long-horizon GMV.

DT for Sequential Bid Optimization We formulate the bid optimization problem as a 24-step Markov decision process (MDP) $\langle \mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{P} \rangle$, where $s_t \in \mathcal{S}$ denotes the system state at hour t, $a_t \in \mathcal{A}$ the bidding action, r_t the immediate reward, and $\mathcal{P}$ the state transition dynamics. The objective is to find an optimal bidding sequence $a_1, \ldots, a_{24}$ that maximizes cumulative return.

In our scenario, state s_t is a feature vector comprising campaign-level metadata and historical performance statistics (e.g., strategy type, remaining budget); action $a_t \in [0.8, 1.2]$ is a multiplicative bid coefficient; reward r_t combines normalized cumulative GMV and a spend-utilization penalty:

$$r_t = \frac{\text{GMV}_t}{\text{GMV}^*} + \min\left(0, 1 - \frac{\text{cost}_t}{B}\right) \tag{13}$$

where GMV_t and cost_t are cumulative GMV and spend up to hour t, GMV^* is the average GMV over past 7 days.

DT is trained with the following loss [28]

$$L_{\mathrm{DT}} = L_{\mathrm{MSE}}(a_{\mathrm{t,DT}},\ a_{\mathrm{t,gt}}) \tag{14}$$

where $a_{t,\mathrm{DT}}$ and $a_{t,\mathrm{gt}}$ denote predicted and ground-truth bid coefficients respectively, and a_t is the abbreviation of $a_{t,\mathrm{DT}}$.

Dual-process Control with PID & DT In real-world deployment, we observed that DT suffers performance degradation during sales promotions or new product launches, where the distributional shift between historical training data and online data renders learned policies ineffective. Inspired by dual-process theory, we propose to encode expert-derived bidding heuristics into a robust, rule-based controller that acts as System 1 to assist the deliberative DT (System 2) in out-of-distribution scenarios. Specifically, we integrate a rule-based PID controller (System 1) grounded in expert heuristics that regulates bids based on spending rate deviation

$$\begin{aligned} a_{\mathrm{t,PID}} &= k_p(\mathrm{cost}_t - \mathrm{budget}_t) \\ &+ k_i \sum_{h=1}^{t} (\mathrm{cost}_h - \mathrm{budget}_h) \end{aligned} \tag{15}$$

with k_p, k_i weighting current / cumulative cost errors, and cost_t, budget_t the actual / ideal cumulative costs up to hour t.

Drawing on insights from industrial applications of dual-process theory, DT and PID are fused in two dimensions: (1) during training, we regularize DT toward PID using a Minimum Description Length (MDL) prior [24]; (2) at inference, we dynamically blend decisions based on DT's predictive uncertainty [7].

Following the MDL principle [11], we augment the DT objective with a regularization term penalizing excessive divergence from PID

$$\begin{aligned} L_{\mathrm{DT}} &= L_{\mathrm{MSE}}(a_{\mathrm{t,DT}},\ a_{t,\mathrm{gt}}) \\ &+ \beta \cdot L_{\mathrm{MSE}}(a_{t,\mathrm{DT}},\ a_{t,\mathrm{PID}}) \end{aligned} \tag{16}$$

where $\beta = 0.1$. At decision time, we employ an uncertainty-weighted fusion mechanism. Since the PID controller is deterministic, uncertainty is quantified solely through the DT's recent prediction error. Let mape ≥ 0 denote the mean absolute percentage error of the DT's bid predictions over the past three hours, which serves as a proxy for model uncertainty. We then compute the final bid coefficient as

$$C_{mi} = \max(1 - \mathrm{mape}, 0) \cdot a_{t,\mathrm{DT}} + \min(\mathrm{mape}, 1) \cdot a_{t,\mathrm{PID}} \tag{17}$$

This ensures graceful degradation: as DT uncertainty increases, control shifts toward the robust PID policy.

5 Experiments

Several experiments are conducted for the following key Research Questions:
RQ1: Does KBD outperform state-of-the-art methods?
RQ2: What is the impact of IML on the macro stage of KBD?
RQ3: Is IEFormer robust to the segment number N?
RQ4: How does the dual-process control affect the overall performance of KBD?

Table 1. Results of click maximization on *iPinYou* dataset.

Method	$R/R^* \uparrow$	constraint satisfaction↑
ARTEO [18]	0.601	80.65%
BOCO [20]	0.618	74.10%
SPB [29]	0.699	75.32%
PUROS [21]	0.723	81.73%
GCB-safe [3]	0.606	80.77%
KBD w/o DT	0.706	81.43%
KBD w/o PID	0.721	80.32%
KBD	**0.730**	**82.78%**

5.1 Experimental Setup

Datasets and Metrics. We evaluate on the public *iPinYou* dataset (tCPC click maximization) and a private *ECA* dataset from 200 advertisers (tCPA/tROI GMV optimization, Sep–Dec 2025). We report R/R^* on *iPinYou* and wmape/perf10 on *ECA*, using Synthetic Control Method [1] for online tests. For online evaluation, we assess performance improvement via GMV lift rather than whether the merchant-set target ROI is met, because self-reported targets are often poorly calibrated—some are trivially easy to satisfy while others are arbitrarily set and practically unattainable—making them unreliable indicators of true bidding quality.

Compared Methods. On *iPinYou* [38] we compare KBD against PUROS and other baselines. On *ECA*, IEFormer is evaluated against GP and XGBoost price–volume models. Only PID-equipped methods are deployed online.

5.2 Main Results (RQ1)

Table 1 shows KBD outperforms all baselines on *iPinYou* in both R/R^* and constraint satisfaction. Online tests (Table 2) show IEFormer alone improves cost-exhaust ratio by 8.4% and DT adds a further 2.3%, confirming that a well-calibrated macro anchor enables effective micro control.

Table 2. Online test results on *ECA*: relative lifts against control group; absolute metrics redacted for commercial confidentiality.

Experimental Period	Group	Group Detail use PID	use IEFormer	use DT	cost-exhaust ratio	GMV	Campaign Duration
2025/09/28-2025/10/18	Control	✓	✗	✗	/	/	/
	Treatment	✓	✓	✗	+ 8.44%	+ 6.14%	+ 0.58%
2025/11/20-2025/12/10	Control	✓	✗	✗	/	/	/
	Treatment	✓	✓	✓	+14.55%	+ 13.01%	+ 1.36%
2025/12/11-2025/12/31	Control	✓	✓	✗	/	/	/
	Treatment	✓	✓	✓	+ 2.27%	+ 2.66%	+ 0.20%

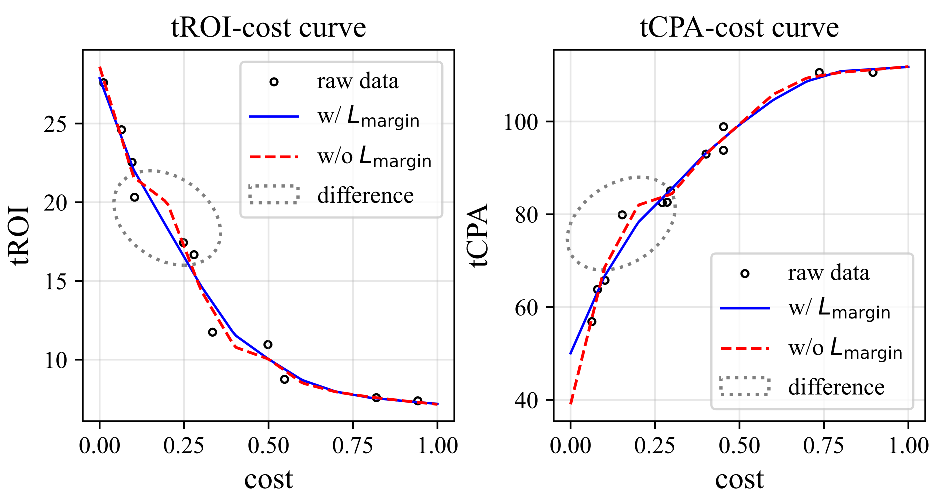

Fig. 2. Case studies on the influence of L_{margin}. As highlighted by the ellipsoidal regions, the price-volume curve is more prone to overfitting the noisy data without L_{margin}.

5.3 Further Analysis

Ablation Study of IML (RQ2) We ablate each IML level on *ECA* (Table 3): w/o IE, w/o GLA, w/o L_{margin}, and w/o tROI data. All variants degrade, confirming that knowledge integration at every level is essential; removing IE causes the largest drop.

Table 3. Ablation Study of IML on *ECA* dataset.

Method	wmape ↓	mape ↓	perf10 ↑
w/o IE	0.2588	0.2348	0.2738
w/o GLA	0.2412	0.2270	0.2793
w/o L_{margin}	0.2386	0.2268	0.2921
w/o tROI data	0.2226	0.2600	0.2976
KBD	**0.2187**	**0.2151**	**0.3214**

Ablating L_{margin} also harms performance. Figure 2 visualizes the learned curves: L_{margin} prevents overfitting and guides the model toward monotonic, smooth curves aligned with expert expectations.

Table 4 benchmarks mainstream price-volume models on *ECA*. All baselines lag far behind IEFormer, indicating that both the interpretable IE module and the powerful Transformer feature extractor are essential.

Table 4. Price-volume model performance on *ECA* dataset.

Method	wmape ↓	mape ↓
IGCM [30]	0.9979	0.9949
GP [26]	0.6414	1.8021
CMNN [19]	0.6309	2.0599
DIPN [27]	0.6308	1.7647
SMNN [17]	0.6179	2.2503
SMM [16]	0.6023	1.6785
XGBoost [5]	0.4967	1.1872
IEFormer	**0.2187**	**0.2151**

Robustness of IEFormer to Number N (RQ3) To evaluate the robustness of IEFormer to the number of segments N, we vary N from 6 to 15 and measure its impact on model performance. As shown in Figure 3, although wmape fluctuates between 21% and 27%, IEFormer consistently outperforms all baseline price-volume models in Table 4, demonstrating strong robustness of the IE module to the choice of N.

Ablation Study of Dual-process Control (RQ4) On *iPinYou* dataset, we compare the standalone DT, PID controller, and their integration via dual-process control.

As shown in Table 1, DT improves R/R^* through global sequential optimization but occasionally violates constraints, while PID reliably satisfies constraints but is myopic. The dual-process framework reconciles these trade-offs, improving both metrics by roughly 1%: DT learns PID's conservative behavior via regularization, while calibrating PID's bids for higher long-term R/R^*.

Online experiments on *ECA* (Table 2) confirm that dual-process control significantly improves cost-exhaust ratio, GMV, and campaign duration over PID alone, validating its effectiveness under distribution shifts.

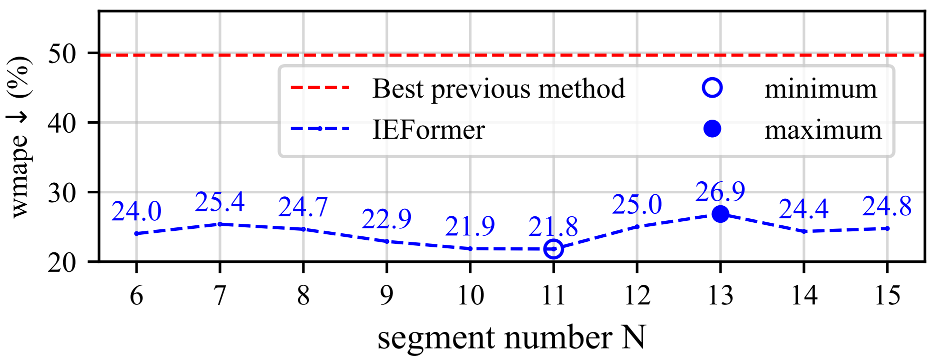

Fig. 3. Robustness of IEFormer to segment number N.

6 Conclusion

We propose KBD, a two-stage bidding framework that integrates expert knowledge into the macro-stage price–volume model via IML and optimizes long-horizon returns in the micro stage using DT. Drawing on dual-process theory, we fuse DT (System 2) with a PID controller (System 1) to balance deliberative planning and reactive control. Experiments on *iPinYou* and a real-world platform show that KBD outperforms state-of-the-art methods, with gains attributed to richer knowledge infusion and dual-process robustness under real-world distribution shifts. Future work will explore LLM-based bidding agents with transparent causal reasoning.

Acknowledgments.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Abadie, A., Diamond, A., Hainmueller, J.: Comparative politics and the synthetic control method. American Journal of Political Science **59**(2), 495–510 (2015)
2. Bennett, S.: Development of the pid controller. IEEE Control Systems Magazine **13**(6), 58–62 (1993)
3. Castiglioni, M., Nuara, A., Romano, G., Spadaro, G., Trovò, F., Gatti, N.: Safe online bid optimization with return on investment and budget constraints. In: Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1. pp. 73–81 (2025)
4. Chen, L., Lu, K., Rajeswaran, A., Lee, K., Grover, A., Laskin, M., Abbeel, P., Srinivas, A., Mordatch, I.: Decision transformer: Reinforcement learning via sequence modeling. Advances in neural information processing systems **34**, 15084–15097 (2021)
5. Chen, T.: Xgboost: A scalable tree boosting system. Cornell University (2016)
6. Chou, P.A., Lookabaugh, T., Gray, R.M.: Entropy-constrained vector quantization. IEEE Transactions on acoustics, speech, and signal processing **37**(1), 31–42 (2002)
7. Daw, N.D., Niv, Y., Dayan, P.: Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nature neuroscience **8**(12), 1704–1711 (2005)
8. Edelman, B., Ostrovsky, M., Schwarz, M.: Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. American economic review **97**(1), 242–259 (2007)
9. Elmachtoub, A.N., Grigas, P.: Smart "predict, then optimize". Management Science **68**(1), 9–26 (2022)
10. Gao, J., Li, Y., Mao, S., Jiang, P., Jiang, N., Wang, Y., Cai, Q., Pan, F., Jiang, P., Gai, K., et al.: Generative auto-bidding with value-guided explorations. In: Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 244–254 (2025)
11. Grünwald, P.D.: The minimum description length principle. MIT press (2007)
12. Guo, J., Huo, Y., Zhang, Z., Wang, T., Yu, C., Xu, J., Zheng, B., Zhang, Y.: Generative auto-bidding via conditional diffusion modeling. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. pp. 5038–5049 (2024)
13. Han, B., Kim, J., Jang, J.: A dual process vla: Efficient robotic manipulation leveraging vlm. arXiv preprint arXiv:2410.15549 (2024)
14. He, T., Liao, L., Cao, Y., Liu, Y., Liu, M., Chen, Z., Qin, B.: Planning like human: A dual-process framework for dialogue planning. arXiv preprint arXiv:2406.05374 (2024)
15. He, Y., Chen, X., Wu, D., Pan, J., Tan, Q., Yu, C., Xu, J., Zhu, X.: A unified solution to constrained bidding in online display advertising. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. pp. 2993–3001 (2021)
16. Igel, C.: Smooth min-max monotonic networks. arXiv preprint arXiv:2306.01147 (2023)
17. Kim, H., Lee, J.S.: Scalable monotonic neural networks. In: The Twelfth International Conference on Learning Representations (2024)
18. Korkmaz, B.S., Zagórowska, M., Mercangöz, M.: Safe and adaptive decision-making for optimization of safety-critical systems: The arteo algorithm. arXiv preprint arXiv:2211.05495 (2022)

19. Li, B., Pei, J., Xiao, F., Zhao, Y., Zhang, Z., Liu, D., He, H., Jia, J.: Enhancing monotonic modeling with spatio-temporal adaptive awareness in diverse marketing. arXiv preprint arXiv:2406.14132 (2024)
20. Liang, J.C.N., Lu, H., Zhou, B.: Online ad procurement in non-stationary autobidding worlds. Advances in Neural Information Processing Systems **36**, 42552–42575 (2023)
21. Luo, H., Gao, L., Huang, P., Li, T.: Puros: A cpx-ievered framework for ad procurement in autobidding worlds. In: 2024 IEEE 13th Data Driven Control and Learning Systems Conference (DDCLS). pp. 1399–1406. IEEE (2024)
22. Mei, J., Ma, Y., Yang, X., Wen, L., Cai, X., Li, X., Fu, D., Zhang, B., Cai, P., Dou, M., et al.: Continuously learning, adapting, and improving: A dual-process approach to autonomous driving. arXiv preprint arXiv:2405.15324 (2024)
23. Mitchell-Wolf, E.: Programmatic advertising forecast and ad tech trends h1 2025. Report, EMARKETER Inc. (Feb 2025)
24. Moskovitz, T., Miller, K., Sahani, M., Botvinick, M.M.: A unified theory of dual-process control. arXiv preprint arXiv:2211.07036 (2022)
25. Mumuni, F., Mumuni, A.: Improving deep learning with prior knowledge and cognitive models: A survey on enhancing explainability, adversarial robustness and zero-shot learning. Cognitive Systems Research **84**, 101188 (2024)
26. Schulz, E., Speekenbrink, M., Krause, A.: A tutorial on gaussian process regression: Modelling, exploring, and exploiting functions. Journal of mathematical psychology **85**, 1–16 (2018)
27. Shen, Y., Wang, Y., Lu, X., Qi, F., Yan, J., Mu, Y., Yang, Y., Peng, Y., Gu, J.: A framework for massive scale personalized promotion. arXiv preprint arXiv:2108.12100 (2021)
28. Su, K., Huo, Y., Zhang, Z., Dou, S., Yu, C., Xu, J., Lu, Z., Zheng, B.: Auctionnet: A novel benchmark for decision-making in large-scale games. Advances in Neural Information Processing Systems **37**, 94428–94452 (2024)
29. Su, Y., Xiang, M., Chen, Y., Li, Y., Qin, T., Zhang, H., Li, Y., Liu, X.: Spending programmed bidding: Privacy-friendly bid optimization with roi constraint in online advertising. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. pp. 5731–5740 (2024)
30. Tang, Y., Cheng, Y., Liu, X., Jiao, C., Zeng, Y., Luo, N., Yuan, P., Liu, X., Jiang, P.: Learning monotonic probabilities with a generative cost model. arXiv preprint arXiv:2506.03542 (2025)
31. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
32. Von Rueden, L., Mayer, S., Beckh, K., Georgiev, B., Giesselbach, S., Heese, R., Kirsch, B., Pfrommer, J., Pick, A., Ramamurthy, R., et al.: Informed machine learning–a taxonomy and survey of integrating prior knowledge into learning systems. IEEE Transactions on Knowledge and Data Engineering **35**(1), 614–633 (2021)
33. Weinan, W., Rong, Y., Wang, J., Zhu, T., Wang, X.: Feedback control of real-time display advertising. Proceedings of the Web Search and Data Mining (WSDM) pp. 407–416 (2016)
34. Wen, C., Xu, M., Zhang, Z., Zheng, Z., Wang, Y., Liu, X., Rong, Y., Xie, D., Tan, X., Yu, C., et al.: A cooperative-competitive multi-agent framework for auto-bidding in online advertising. In: Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining. pp. 1129–1139 (2022)

35. Wu, Z., Wang, L., Huang, F., Zhou, L., Song, Y., Ye, C., Nie, P., Ren, H., Hao, J., He, R., et al.: A framework for multi-stage bonus allocation in meal delivery platform. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. pp. 4195–4203 (2022)
36. Yang, X., Li, Y., Wang, H., Wu, D., Tan, Q., Xu, J., Gai, K.: Bid optimization by multivariable control in display advertising. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 1966–1974 (2019)
37. Zhang, K., Yuan, Z., Xiong, H.: Optimized cost per click in online advertising: A theoretical analysis. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. pp. 4232–4243 (2024)
38. Zhang, W., Yuan, S., Wang, J., Shen, X.: Real-time bidding benchmarking with ipinyou dataset. arXiv preprint arXiv:1407.7073 (2014)