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

Huixiang Luo, Longyu Gao, Qianqian Chen, Tianning Li, Yaqi Liu, Pingchun Huang

Alibaba Health • AdKDD 2026

AliHealth
阿里健康

Outline

- Part 1. Motivation & Problem Formulation
- Part 2. Methodology — KBD Framework
- Part 3. Experiments & Results
- Part 4. Conclusion & Future Work

01 Motivation & Problem Formulation

Motivation: Bid Optimization Challenges

The Challenge

Auto-bidding dominates digital advertising (>$200B in 2026). Advertisers set high-level goals (tROI, tCPA), and platforms handle per-impression bidding.

Why Black-box ML Falls Short

(1) Poor generalization in data-sparse cases

(2) Myopic single-step optimization

(3) Fragile under distribution shifts

Our Approach: KBD

Embed human expertise + global sequential optimization + robust dual-process control.

Mainstream Auto-bidding Strategies

Rule-based

RTBControl, MPID

Reinforcement Learning

USCB, MAAB

Generative Models

DiffBid, GAVE

How to combine expert knowledge with data-driven models?

Problem Formulation

Bid Optimization under GSP Auction (tCPA auto-bidding)

$$\max_{\{x_i\}} \sum_{i=1}^{M} x_i \cdot v_i \quad \text{s.t.} \quad \sum_{i=1}^{M} x_i \cdot \mathrm{wp}_i \le B \quad \frac{\sum_{i=1}^{M} x_i \cdot \mathrm{wp}_i}{\sum_{i=1}^{M} x_i \cdot \mathrm{pCTR}_i \cdot \mathrm{pCVR}_i} \le C \quad x_i \in \{0,1\}, \quad \forall i.$$

$v_i = \mathrm{pCTR}_i \cdot \mathrm{pCVR}_i \cdot \mathrm{ppay}_i$ (expected conversion value), $x_i \in \{0,1\}$

Optimal Bid (MPID Framework)

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

λ^0_i, λ^1_i set by platform (unobservable)
C: the advertiser's only control lever

KBD Decomposition

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

C_ma: day-level tCPA (IEFormer) C_mi: hourly coeff (DT+PID)

Macro Stage (Daily)

IEFormer → base tCPA (C_ma)

→

Micro Stage (Hourly)

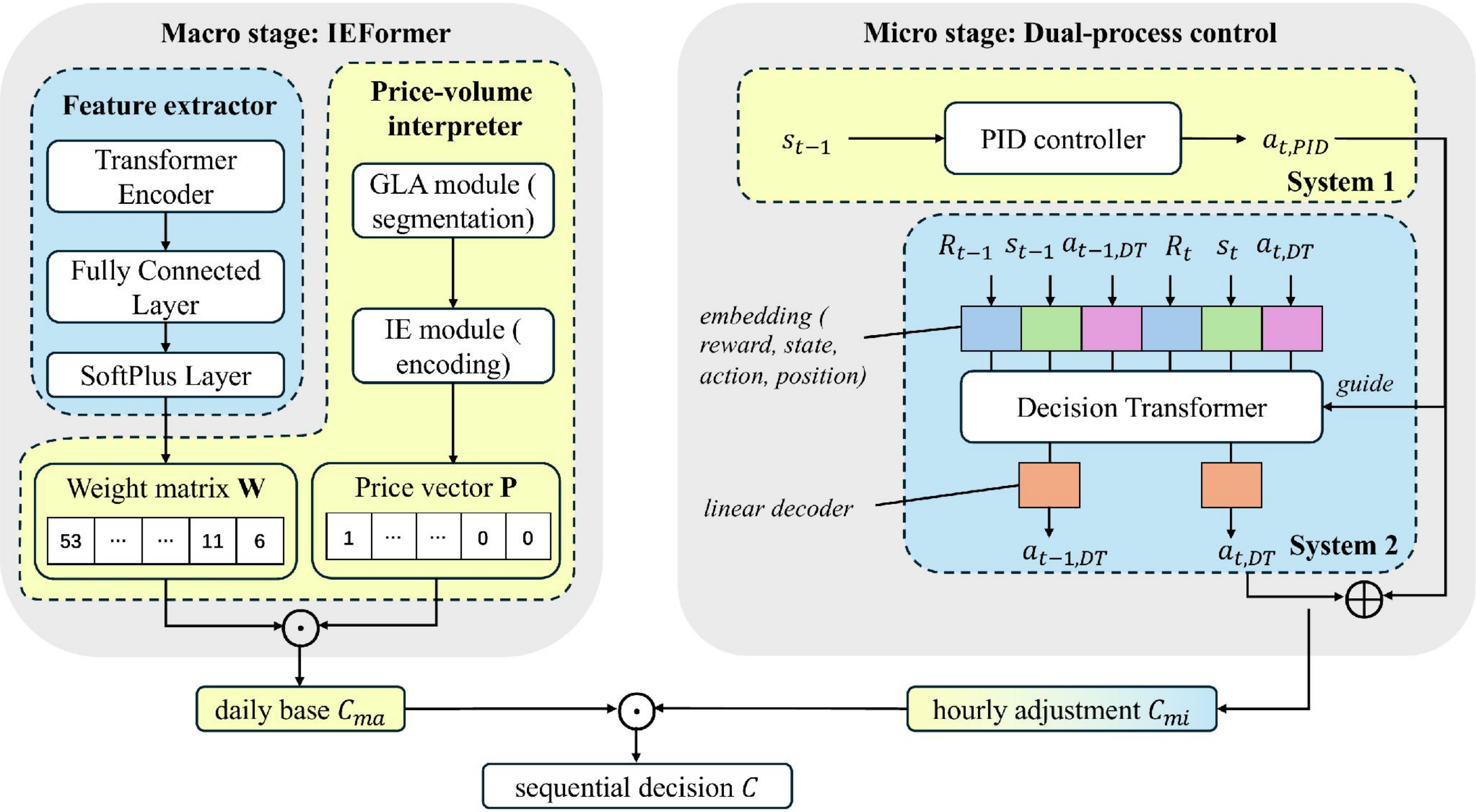
DT (System 2) + PID (System 1) → hourly coeff (C_mi)

02

Methodology: KBD Framework

AliHealth 阿里健康

KBD Architecture Overview



AliHealth 阿里健康

Macro Stage: IEFormer

Hypothesis Level

- Transformer encoder for feature extraction
- Symbolic price-volume interpreter: tCPA = P · W
- Isotonic Embedding ensures monotonicity
- Entropy-driven partitioning via GLA

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

Algorithm Level

- Monotonicity: Softplus activation (W ≥ 0)
- Smoothness: penalize sharp weight transitions
- Diminishing returns: L_margin constraint
- Combined loss ensures expert-aligned curves

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

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

Data Level

- Cross-strategy transfer: tROI/tCPC ⟶ tCPA
- Unified eCPM formula for goal conversion
- Augmented samples weighted at 0.1
- Mitigates data scarcity problem

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

L_IEFormer = L_MSE + α_smooth · L_smooth + α_margin · L_margin   (α_smooth = 0.5, α_margin = 0.1)

AliHealth 阿里健康

Micro Stage: Dual-process Control (DT + PID)

System 2: Decision Transformer (DT)

- 24-step MDP: state = campaign features
- Action a_t ∈ [0.8, 1.2]: multiplicative bid coefficient
- Reward: normalized GMV + spend-utilization penalty
- Optimizes long-horizon cumulative return
- Loss: $L_{\mathrm{DT}} = L_{\mathrm{MSE}}(a_{\mathrm{t,DT}},\ a_{\mathrm{t,gt}})$

System 1: PID Controller

- Rule-based: spending rate deviation control
- $a_{\mathrm{t,PID}} = k_p(\mathrm{cost}_t - \mathrm{budget}_t) + k_i \sum_{hi=1}^{t} (\mathrm{cost}_{hi} - \mathrm{budget}_{hi})$
- Deterministic, low-latency, no training needed
- Robust to distribution shifts
- Provides expert heuristic guidance

Fusion Mechanism

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

Training:

$$L_{\mathrm{DT}} = L_{\mathrm{MSE}}(a_{\mathrm{t,DT}},\ a_{t,\mathrm{gt}}) + \beta \cdot L_{\mathrm{MSE}}(a_{t,\mathrm{DT}},\ a_{t,\mathrm{PID}})$$

As DT uncertainty ↑, control shifts → PID

03 Experiments & Results

Experimental Setup

Datasets & Metrics

- iPinYou (public): tCPC click max., metric R/R*
- ECA (private): 200 advertisers, tCPA/tROI
- Evaluation: Sep–Dec 2025, 3 test periods
- Online eval: Synthetic Control Method, GMV lift

Research Questions

RQ1	Does KBD outperform SOTA?
RQ2	Impact of IML on macro stage?
RQ3	Robustness to segment number N?
RQ4	Effect of dual-process control?

Compared Methods

iPinYou: ARTEO, BOCO, SPB, PUROS, GCB-safe • ECA: GP, XGBoost, CMNN, DIPN, SMNN, SMM, IGCM

RQ1: Offline Results on iPinYou Dataset

Method	R/R* ↑	Constraint Sat. ↑
ARTEO	0.601	80.65%
BOCO	0.618	74.10%
SPB	0.699	75.32%
PUROS	0.723	81.73%
GCB-safe	0.606	80.77%
KBD w/o DT	0.706	81.43%
KBD w/o PID	0.721	80.32%
KBD (full)	**0.730**	**82.78%**

Key Findings

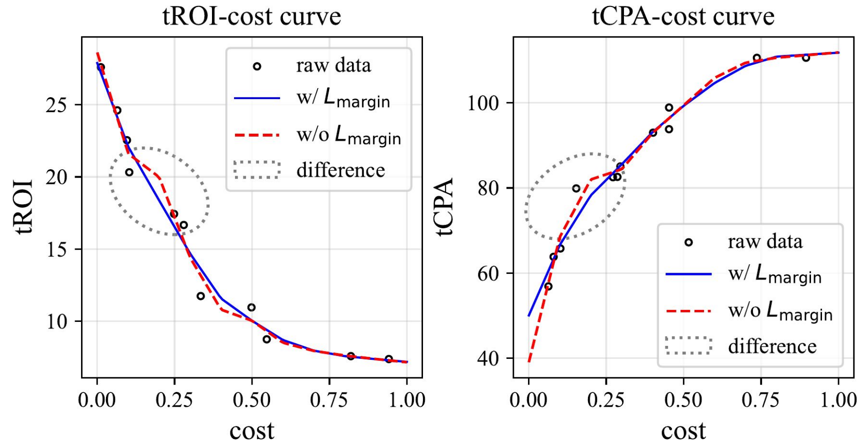
- KBD outperforms all baselines on both metrics
- DT: +0.7% R/R* via global sequential optimization
- PID: +2.5% constraint satisfaction via reactive control
- Dual-process fusion: best of both worlds

RQ1: Online A/B Tests on ECA Platform

Period	IEFormer	DT	Cost-exhaust	GMV	Duration
Sep 28 – Oct 18	✓	✗	+8.44%	+6.14%	+0.58%
Nov 20 – Dec 10	✓	✓	+14.55%	+13.01%	+1.36%
Dec 11 – Dec 31	**✓**	**✓**	**+2.27%**	**+2.66%**	**+0.20%**

All groups use PID. Lifts relative to control. Absolute metrics redacted for confidentiality.

Effect of L_margin on Price-Volume Curves



Takeaways

- IEFormer alone: +8.4% cost-exhaust, +6.1% GMV
- Full KBD: +14.6% cost-exhaust, +13.0% GMV
- L_margin prevents overfitting in price-volume curves
- Dual-process robust across all 3 test periods

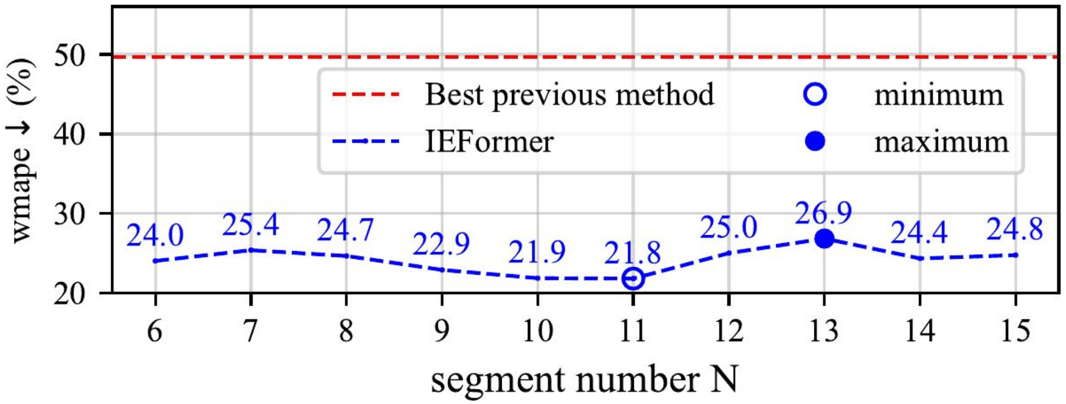
阿里健康

Further Analysis (RQ2–RQ4)

RQ2: IML Ablation on ECA

Variant	wmape ↓	perf10 ↑
w/o IE	0.2588	0.2738
w/o GLA	0.2412	0.2793
w/o L_margin	0.2386	0.2921
w/o tROI	0.2226	0.2976
KBD (full)	**0.2187**	**0.3214**

RQ3: Robustness to Segment Number N



wmape 21–27% for N∈[6,15]; consistently outperforms all baselines

RQ4: Dual-process Effect

DT improves R/R* via global optimization.
PID ensures constraint compliance.
Fusion: ~1% gain on both metrics vs either alone.

AliHealth 阿里健康

Q&A: DT — Behavior Cloning or Optimization?

Q: Is the Decision Transformer doing behavior cloning or true optimization?

Training Paradigm: Expert Demonstration Learning

- DT learns from carefully curated high-quality bidding trajectories rather than arbitrary historical data.

Trajectory Selection Criteria (past 30 days):

- Campaign duration ≥ 12 hours
- Cost exhaustion ratio ≥ 0.9 (budget fully utilized)
- ROI ranks in top 50% among campaigns with the same budget tier

Why This Matters:

- The selected trajectories represent near-optimal bidding strategies. DT generalizes these expert policies to unseen campaigns, achieving optimization through imitation of proven winners.

Key Clarification

Not random cloning

Trajectories are rigorously filtered for quality, ensuring DT learns from near-optimal expert policies.

Offline → Online

DT's offline-learned policy is further refined at inference via the MAPE-based fusion with PID (System 1).

Result: $R/R^* = 0.730$

Outperforms all baselines including pure RL approaches.

AliHealth 阿里健康

Q&A: Statistical Significance & SCM Methodology

Why Synthetic C

- **DID:** violated — ex
- **PSM:** unable to fin
- **DDD:** non-algo-ma

on gaps too large.

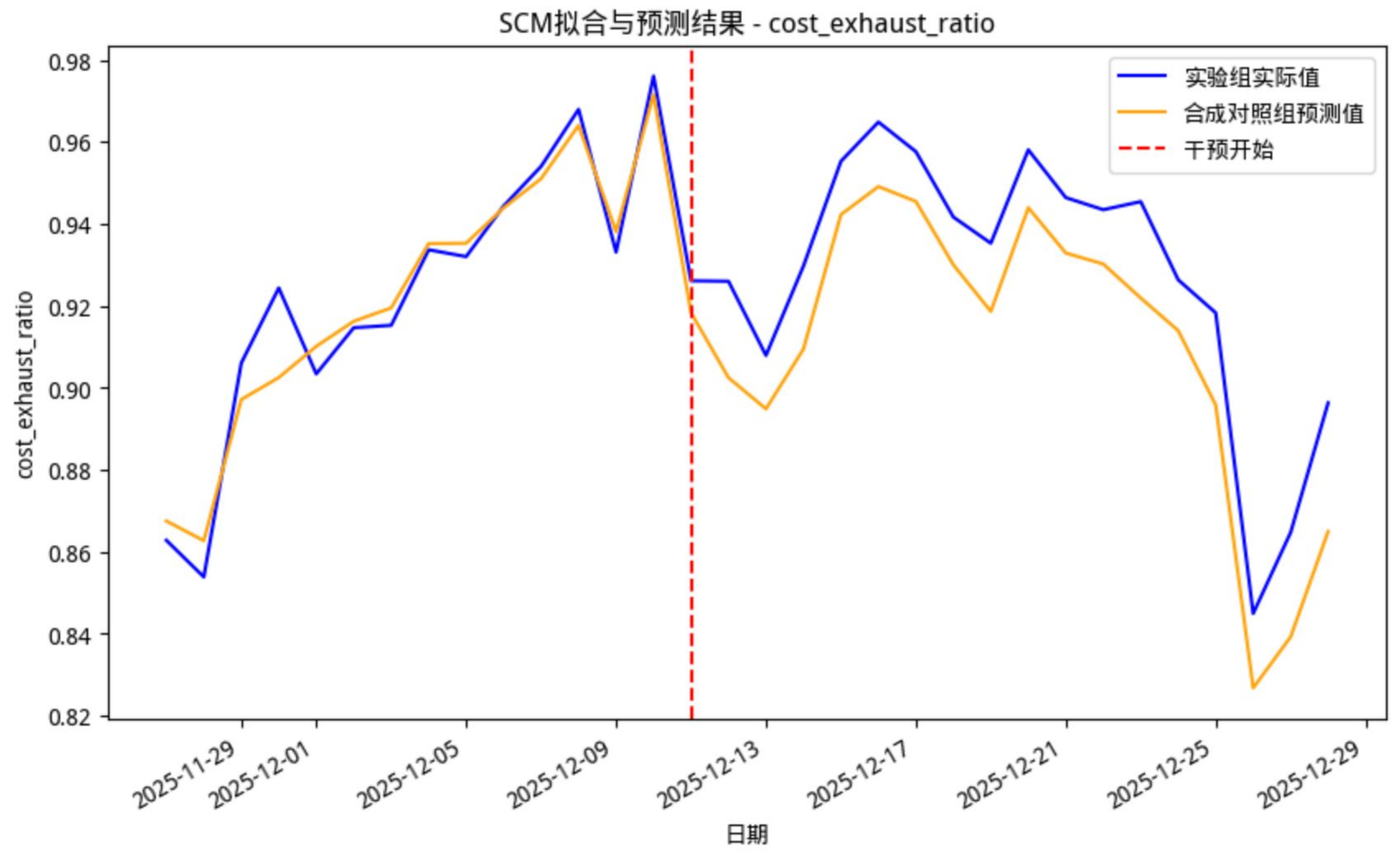
SCM Evaluation Pip

Step 1: In

Split all campaigns into and control groups such cost, budget, and camp are as close as possible

gnificance

inst
AB
significant
ed.



Q&A: Why Not Report Online CPA Compliance?

The Problem with Self-reported Targets

Merchants set their own target CPA/ROI, which are often poorly calibrated:

- Some targets are trivially easy to satisfy → no discriminating power
- Others are arbitrarily set and practically unattainable → unfair penalty
- Wide variance across advertisers makes aggregation unreliable

Using these as evaluation metrics would conflate bidding quality with target-setting quality.

Our Evaluation Strategy

GMV Lift

Measures real business value improvement, independent of target calibration.

Cost Exhaustion Ratio

Measures budget utilization efficiency, reflecting bidding aggressiveness quality.

Summary

We assess performance improvement via GMV lift and cost exhaustion rather than CPA target compliance. This is because self-reported targets are unreliable indicators of true bidding quality — some trivially easy, others arbitrarily strict. GMV lift captures the actual business impact of bidding decisions.

AliHealth 阿里健康

Conclusion

1 **KBD Framework**

Two-stage bid optimization: macro IEFormer (daily tCPA) + micro DT (hourly adjustment). Embeds expert knowledge at hypothesis, algorithm, and data levels.

2 **Dual-process Control**

System 1 (PID) guides System 2 (DT) during training via MDL regularization. Uncertainty-weighted fusion ensures graceful degradation under distribution shifts.

3 **Strong Empirical Results**

Outperforms SOTA on iPinYou (R/R* = 0.730). Online A/B tests: +14.6% cost-exhaust ratio, +13.0% GMV lift with full KBD.

Future Work: LLM-based bidding agents with transparent causal reasoning

阿里健康

Thank You!

AliHealth
阿里健康