

OrDA: Orthogonal Disentanglement of Access Habits Framework for Homepage Marketing Block Recommendations

Lingxiao Zhang
owen.zlx@antgroup.com
Ant Group
Hangzhou, Zhejiang, China

Xiaobo Li
yujing.lxb@antgroup.com
Ant Group
Hangzhou, Zhejiang, China

Tao Xu
tomas.xt@antgroup.com
Ant Group
Hangzhou, Zhejiang, China

Abstract

Clicks on homepage marketing blocks are driven by a dual-mechanism of content interest and access habits. However, habitual clicks often create "Pseudo-Positives" in marketing slots, where position advantage masks mediocre content quality, leading to biased recommendation ecosystems.

We propose a framework called **Orthogonal Disentanglement of Access habits (OrDA)** to purify interest signals. OrDA utilizes a dual-tower structure with a gated allocation layer to adaptively route features and minimize interference. To ensure rigorous separation, we employ orthogonal regularization to constrain the latent interest and habit manifolds to be geometrically perpendicular. OrDA performs causal intervention (*do*-calculus[10]) during inference to rank items solely by purified interest scores. Empirical online evaluations on large-scale datasets demonstrate that OrDA effectively eliminates access-habit bias, outperforming state-of-the-art methods in predictive accuracy. Online A/B test shows 5.64% user click-through rates (UCTR) improvement on the Zhima homepage marketing block, Zhima's rent-floor recommendation.

CCS Concepts

• **Information systems** → **Retrieval models and ranking.**

Keywords

Disentangled Learning, Causal Intervention, Recommender System

ACM Reference Format:

Lingxiao Zhang, Xiaobo Li, and Tao Xu. 2026. OrDA: Orthogonal Disentanglement of Access Habits Framework for Homepage Marketing Block Recommendations. In . ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 Introduction

The homepage of an internet application can be summarized as having the core function of: precise navigation and ecosystem distribution from massive content to user interests. Homepage marketing blocks (shown in Fig. 1) play the role of guiding the vast user traffic to different vertical sub-scenarios[15]. These blocks often suffer from a subtle yet pervasive form of confounding bias: the user-channel access habitual dependency. Specifically, in prominent marketing blocks—frequently accessed by "heavy users" as part of their routine digital journey—observed click data is often contaminated by "Pseudo-Positives." These are interactions driven not by

a genuine alignment between user interests and material content, but rather by the user's habitual inertia within their primary access channels. Standard models, which treat all clicks as equivalent sig-

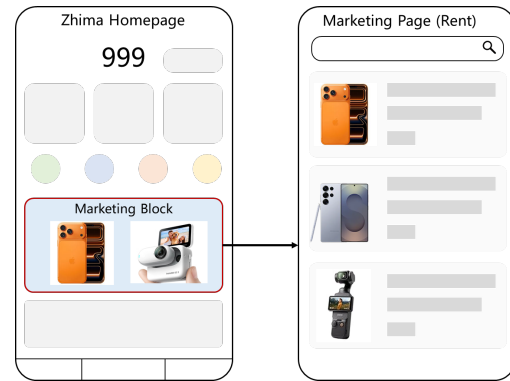


Figure 1: A demo of Zhima homepage marketing block recommendation

nals of preference, inadvertently capture this "habitual dividend," leading to a distorted representation of user interests. From a causal perspective, the user's channel affinity acts as a confounder, inflating the perceived utility of certain contents simply because they occupy the user's "path of least resistance." Consequently, such models often perform poorly with cold-start users and fail to generalize to broader, non-habitual contexts, as they struggle to distinguish between user-channel access habits and user-content interests.

In order to address this problem, we propose a **Orthogonal Disentanglement of Access habits (OrDA)** framework for homepage marketing block recommendations. The main contributions of this work are summarized as follows:

- **Problem Formulation of Access Habit Bias:** We provide a formal causal analysis of access habit bias prevalent in homepage marketing blocks. Unlike conventional exposure bias, we identify and define the phenomenon of habit-induced pseudo-positives, where users' navigational inertia at primary entry points masks their genuine content preferences, providing a new perspective on bias decomposition.
- **Orthogonal Disentanglement of Access Habits(OrDA):** We propose OrDA framework, a novel multi-tower architecture designed to decouple user intent from habitual access patterns. Our model can effectively isolate the spurious correlations between user-content interests and user-channel access habits via orthogonal regularization and *do*-calculus-based intervention strategy.

Conference'17, Washington, DC, USA

© 2026 Copyright held by the owner/author(s). Publication rights licensed to ACM. This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in , <https://doi.org/10.1145/nnnnnnn.nnnnnnn>.

- **Superior Performance on Industrial Datasets:** OrDA achieves state-of-the-art performance on Zhima homepage datasets during extensive offline experiments. Online A/B test shows 5.64% UCTR gains for Zhima's rent-floor recommendations.

2 Related Work

2.1 Debiasing in Recommendation Systems

The identification and elimination of biases, such as position bias[16] and popular bias[16], are critical for building robust recommendation systems[2]. Traditional approaches mainly include propensity-based method[8, 15], which re-weights samples based on their exposure probability, dual learning[12, 14, 19], which learns an auxiliary model to estimate biases, disentangled representation learning[16, 17], which employs contrastive learning to separate interest from conformity, and pseudo-label method[4], which aims to complete the counterfactual space by assigning estimated labels to unobserved entries. Our work differs from these by specifically focusing on access habits in homepage scenarios and performing debiasing through a geometric orthogonal constraint rather than simple score adjustment to enforce a clear causal boundary and ensure a more robust representation of user intrinsic interests.

2.2 Synergizing Multi-tower Architectures with Causal Inference

Multi-tower models (e.g., ESMM[7], MMOE[6], PLE[9]) are widely used in industrial recommendation to handle multiple tasks. In reality, these architectures provide a structural foundation for task or feature decomposition[11, 14, 15, 19]. From a causal perspective, a significant limitation of traditional multi-tower models is feature interference: since towers often share a common embedding bottom, the gradients from a biased signal (e.g., access habit) can easily contaminate the representations of other towers (e.g., content interest). Our work advances this structural decomposition by re-interpreting the multi-tower topology as a physical realization of a Structural Causal Model (SCM).

3 Methods

3.1 Problem Definition

To formally characterize the mechanism of habit-induced pseudo-positives and justify the architecture of OrDA, we employ a causal graph to represent the data generation process in homepage marketing blocks. As illustrated in Fig. 2, we define the following variables:

- X : The user-content pair.
- H : The latent navigational inertia or *access-habit* bias.
- I : The latent *interest* of a user towards the content.
- Y : The observed click outcome.
- u : User features such as demographics preferences.
- c : Content features such as category and quality.
- $u2c$: Statistical features of a user towards the content.

We define user-channel access habit and user-content interest as the main factors governing the click behavior in homepage marketing

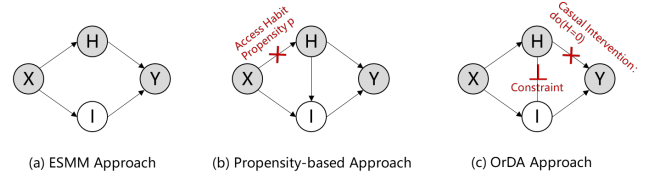


Figure 2: Causal graphs depicting training of ESMM-based Approach, Propensity-based Approach, OrDA. Hollow circles indicate latent variables, and shaded circles indicate observed variables.

blocks. The structural equations of these latent spaces are:

$$\begin{cases} I = f_I(u, c, u2c), \\ H = f_H(u) \end{cases} \quad (1)$$

We aim to get the pure user-content interest from the confusion. Fig. 2(a) shows ESMM-based approach[7] uses two towers to predict access habit probability $P(\hat{h}) = \sigma(H)$ and user-content interest probability $P(\hat{i}) = \sigma(I)$, and multiplies them to predict habit-induced user interest probability $P(\hat{y}) = \sigma(Y)$:

$$P(\hat{y}) = P(\hat{h}) * P(\hat{i}) \quad (2)$$

This approach does not enforce a clear causal boundary or relation between two towers, and thus it introduces false independent prior.

Fig. 2(b) shows propensity-based approach[11, 14, 15, 19] integrates counterfactual reasoning into multi-task learning to mitigate selection bias via inverse propensity weighting technique and *do*-calculus[10]. Habit-induced user interest probability is defined as:

$$P(\hat{y}) = P(\hat{h}) * P(\hat{i} | do(\hat{h} = 1)) \quad (3)$$

During the training stage, this approach adds another tower to predict habit-induced interest probability $P(\hat{i} | do(\hat{h} = 1)) = \frac{P(\hat{y})}{P(\hat{h})}$. While the division form of the expression relying on numerical re-weighting may suffer from high variance and instability.

OrDA (shown in Fig. 2(c)) extends the causal philosophy to the **representation level** as following equation:

$$Y = H \oplus I \quad (4)$$

where $\oplus$ denotes the joint logit fusion, representing the additive interaction between the user-content intrinsic interest and their access habit in the latent space. Our model posits that I and H are parallel causal drivers and a click occurs when the content satisfies the user's interest **or** the block triggers their access habits. Thus we enforce **Latent Orthogonal Constraint** to ensure that the interest tower is inherently invariant to habitual confounders.

3.2 Training and Counterfactual Inference

During the training stage, the loss functions of bias CTR and Habit are defined as:

$$\begin{cases} \mathcal{L}_{bCTR} = \mathcal{L}_{BCE}(y, \sigma(Y)) = \mathcal{L}_{BCE}(y, \sigma(H \oplus I)) \\ \mathcal{L}_{Habit} = \mathcal{L}_{BCE}(y, \sigma(H)) \end{cases} \quad (5)$$

where $\mathcal{L}_{BCE}$ is Binary Cross-Entropy loss:

$$\mathcal{L}_{BCE} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \quad (6)$$

where y is the ground truth and p is the probability of positive class. To ensure rigorous separation, the orthogonal cosine regularization[1, 5] is employed to constrain the latent manifolds, forcing the interest and habit vectors to be geometrically perpendicular:

$$\mathcal{L}_{Orth} = \left(\frac{\mathbf{v}_{int} \cdot \mathbf{v}_{hab}}{\|\mathbf{v}_{int}\| \cdot \|\mathbf{v}_{hab}\|} \right)^2 \quad (7)$$

where $\mathbf{v}_{int}$ is the interest tower vector and $\mathbf{v}_{hab}$ is the habit tower vector. Total loss function is as follow:

$$\mathcal{L}_{total} = \mathcal{L}_{bCTR} + \alpha * \mathcal{L}_{Habit} + \beta * \mathcal{L}_{Orth} \quad (8)$$

where α and β are hyperparameters that control the weights of the habit task and the orthogonal task.

During the online inference stage, a causal intervention on the habit variable is performed to eliminate the pseudo-positive effect. Formally, we compute the purified user-content interest score S by applying do -operator $do(H = 0)$:

$$S = \mathbb{E}[Y \mid I, do(H = 0)] = \sigma(I) \quad (9)$$

3.3 Model Architecture

As illustrated in Fig. 3, the model architecture consists of three core components: the Gated Allocation Layer (GAL), the Backbone Model (BM), and the Causal Fusion Layer (CFL).

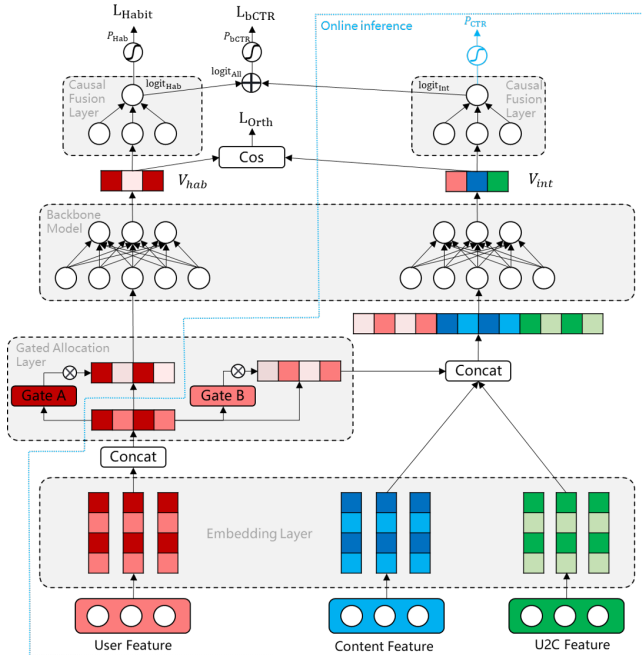


Figure 3: Model architecture of OrDA. The entire components are used for the training stage, while components within the blue box are used for the online inference stage.

3.3.1 Gated Allocation Layer (GAL). A potential pitfall in disentangled learning is information suppression, where the orthogonality constraint inadvertently erases generic user signals from shared embeddings. To address this, we introduce the Gated Allocation Layer (GAL) as a learnable router, avoiding all features to flow into both towers. For an input embedding $\mathbf{e}_i \in \mathbb{R}^d$, GAL computes a routing score $gate_i \in [0, 1]$ using a sigmoid-activated gating network:

$$gate_i = \sigma(\mathbf{w}_g^T \mathbf{e}_i + b_g) \quad (10)$$

The features are adaptively allocated into two distinct latent streams:

$$\begin{cases} \mathbf{e}_{u,hab} = gate_A \cdot \mathbf{e}_u, \\ \mathbf{e}_{u,int} = gate_B \cdot \mathbf{e}_u \end{cases} \quad (11)$$

where $gate_A$ is used for the habit tower and $gate_B$ is used for the interest tower. Note that GAL is only used in user features because we assume user-channel access habits are only affected by user features (mentioned in Eq. 1). This mechanism ensures that the orthogonality constraint only suppresses the redundant interference signals while allowing multi-faceted features (e.g., user demographics) to coexist in both manifolds through distinct projections.

3.3.2 Backbone Model (BM). The backbone model is an orthogonal dual-tower. The two feature streams are fed into the Habit Tower $\mathcal{F}_{hab}$ and the Interest Tower $\mathcal{F}_{int}$ respectively. Each tower is designed to project the filtered features into their respective latent manifolds:

$$\begin{cases} \mathbf{v}_{hab} = \mathcal{F}_{hab}(\mathbf{e}_{u,hab}) \\ \mathbf{v}_{int} = \mathcal{F}_{int}(\text{concat}(\mathbf{e}_{u,int}, \mathbf{e}_c, \mathbf{e}_{u2c})), \end{cases} \quad (12)$$

where $\mathbf{e}_{u,hab}$ and $\mathbf{e}_{u,int}$ is shown in Eq. 11. $\mathbf{e}_c$ and $\mathbf{e}_{u2c}$ are the embedding of the content features and the user2content features respectively. $\mathcal{F}_{hab}$ only processes user features.

In order to the disentanglement theorized in our causal graphs, we enforce a orthogonal constraint (Eq. 7) to ensure that $\mathbf{v}_{int} \perp \mathbf{v}_{hab}$ in the vector space, effectively "pushing" the content-driven and habit-driven signals into non-overlapping dimensions.

3.3.3 Causal Fusion Layer (CFL). The final stage is the structural combination of the two towers. Following our additive formulation of causal graphs (Eq. 4), the logits of both towers are summed via the $\oplus$ operator:

$$logit_{all} = logit_{hab} \oplus logit_{int} = \text{MLP}_{CFL}(\mathbf{v}_{hab}) + \text{MLP}_{CFL}(\mathbf{v}_{int}) \quad (13)$$

where the operator $\oplus$ denotes the joint logit fusion, representing the additive interaction between the user's intrinsic interest and their navigational habit. Our $\oplus$ operator explicitly acknowledges that the click is a superposition of two independent semantic forces. $logit_{int}$ captures the "pull" from content relevance, while $logit_{hab}$ captures the "push" from the user's ingrained access patterns on the homepage. Note that $\text{MLP}_{CFL}(\cdot)$ in Eq. 13 is a Multi-Layer Perceptron (MLP) as the mapping function from the vector space to the logit space. To ensure the isolation and additivity between the two in the logit space, we use *linear* as the activation function of the causal fusion layer.

The combined logit is then passed through a sigmoid function to produce the final click probability $P(\hat{y})$:

$$P(\hat{y}) = \sigma(logit_{all}) \quad (14)$$

Table 1: Performance comparison on production datasets.

Model	$GAUC_{all}$	$GAUC_{cold}$	$GAUC_{active}$
BASE	0.6122	0.6110	0.6497
ESMM	0.6299	0.6297	0.6366
USD	0.6241	0.6236	0.6401
Multi-IPW	0.6315	0.6314	0.6360
PAL	0.6314	0.6313	0.6345
OrDA	0.6412	0.6416	0.6304

During the training stage, both $logit_{int}$ and $logit_{hab}$ are optimized simultaneously to fit the observed click labels. During the inference stage, the $\oplus$ operator allows for the seamless removal of the habit component by setting $logit_{hab} = 0$ (Eq. 9), thereby yielding a purified interest score.

4 Experiments

4.1 Experimental Setup

4.1.1 Datasets. Production Dataset: A large-scale dataset from Zhima homepage marketing block (15-day collection), containing over 30.2 million users, 108.9 million training samples and 7.9 million evaluation samples.

4.1.2 Baselines. We compared with the standard model **BASE** (w/o debiasing components) and multiple SOTA debiasing methods: multi-task model **ESMM**[7], propensity-based models **USD**[15] and **Multi-IPW**[14], factorization model **PAL**[3].

4.1.3 Implementation Details. We use a Multi-Layer Perceptron (MLP) as $\mathcal{F}_{hab}$ and a MaskNet[13] as $\mathcal{F}_{int}$ (Eq. 12), Adam optimizer (lr=0.0001, batch size=2048), loss weights $\alpha = \beta = 1$ (Equation 8). Due to class imbalance in invalid exposure, we use GAUC [18] as primary metric: $GAUC = \frac{\sum_{u \in \mathcal{U}} w_u \times AUC_u}{\sum_{u \in \mathcal{U}} w_u}$, where $w_u = 1$, all user group, cold-start user group (click=1) and active user group (click>1) correspond to $GAUC_{all}$, $GAUC_{cold}$, $GAUC_{active}$.

4.2 Overall Performance Comparison and Ablation Study

As shown in Table 1, **BASE** model exhibits higher $GAUC_{active}$ and lower $GAUC_{cold}$ because the model tends to "memorize" the active users' long-term, predictable access habits as strong positive signals (e.g., the users always clicking the first slot upon entering the app) but does not understand the user-content interests. **Debiasing** models get higher $GAUC_{cold}$ performances since they tend to strip away the "habitual noise" and **OrDA** achieves the best performance both on $GAUC_{cold}$ and $GAUC_{all}$.

We compare OrDA with several variants as shown in Table 2 to investigate the impact of the core components. **w/o GAL** (replace GAL with a standard shared-bottom) has a performance decay, which we attribute to the information evaporation effect caused by strict orthogonality, forcing all user feature embeddings to zero or random noise. Other ablated variants: **w/o HL** (remove the habit auxiliary loss function, $\alpha = 0$), **w/o OL** (remove the cosine regularization, $\beta = 0$), **w/o doC** (use the raw $logit_{int} \oplus logit_{hab}$ for ranking) face decreases of 1.06%, 1.54%, and 2.72% in $GAUC_{all}$, respectively.

Table 2: Ablation Study of OrDA

Model	$GAUC_{all}$	$GAUC_{cold}$	$GAUC_{active}$
w/o GAL	0.6099	0.6108	0.5977
w/o HL	0.6344	0.6348	0.6234
w/o OL	0.6313	0.6311	0.6393
w/o doC	0.6237	0.6235	0.6331
OrDA	0.6412	0.6416	0.6304

4.3 Latent Space Visualization

To quantitatively and visually assess the degree of disentanglement, we compute the cosine similarity matrix between the interest vectors $\mathbf{v}_{int}$ and the habit vectors $\mathbf{v}_{habit}$ across a random subset of the evaluation set. As illustrated in Fig. 4, the similarity heatmap exhibits a near-zero distribution across both the diagonal (intra-sample) and off-diagonal (inter-sample) elements, demonstrating the two vector spaces are constrained to be mutually orthogonal, thereby confirming that the disentanglement is both sample-wise consistent and globally robust.

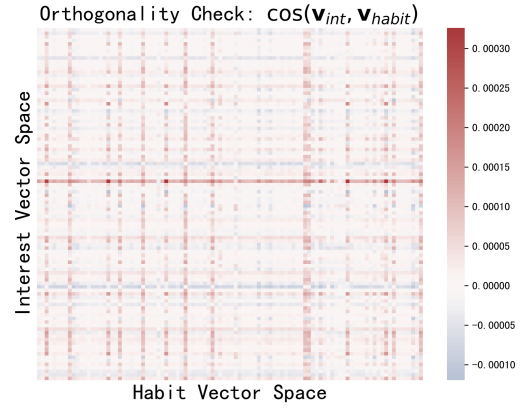


Figure 4: Visualizing the orthogonality between interest and habit vectors. The color of the heatmap is predominantly white/light, indicating that the cosine similarity values are concentrated around 0.

4.4 Online A/B Test

We run a week-long online A/B test on Zhima homepage marketing blocks: Zhima's rent-floor, where **OrDA** outperforms the baseline model (**BASE**, see Sec.4.1.2) with UCTR increased 5.64%. This substantial gain demonstrates the industrial validity of OrDA.

5 Conclusion

We propose the OrDA framework designed to disentangle user-content intrinsic interest from user-channel access habit on homepage recommendations. Extensive experiments on large-scale industrial datasets and the online A/B test demonstrate its effectiveness, effectively recovering purified interest scores. Notably, OrDA has been fully deployed in Zhima homepage marketing blocks: Zhima's rent-floor recommendation.

References

- [1] Konstantinos Bousmalis, George Trigeorgis, Nathan Silberman, Dilip Krishnan, and Dumitru Erhan. 2016. Domain separation networks. In *Proceedings of the 30th International Conference on Neural Information Processing Systems (NIPS'16)*. 343–351.
- [2] Jiawei Chen, Hande Dong, Xiang Wang, Fuli Feng, Meng Wang, and Xiangnan He. 2023. Bias and Debias in Recommender System: A Survey and Future Directions. *ACM Transactions on Information Systems* 41, 67 (Feb. 2023), 1–39. doi:10.1145/3564284
- [3] Huifeng Guo, Jinkai Yu, Qing Liu, Ruiming Tang, and Yuzhou Zhang. 2019. PAL: a position-bias aware learning framework for CTR prediction in live recommender systems. In *Proceedings of the 13th ACM Conference on Recommender Systems (RecSys '19)*. 452–456.
- [4] Jiahui Huang, Lan Zhang, Junhao Wang, Shanyang Jiang, Dongbo Huang, Cheng Ding, and Lan Xu. 2024. Utilizing Non-click Samples via Semi-supervised Learning for Conversion Rate Prediction. In *Proceedings of the 18th ACM Conference on Recommender Systems (RecSys '24)*. 350–359.
- [5] Chunjie Luo, Jianfeng Zhan, Lei Wang, and Qiang Yang. 2017. *Cosine Normalization: Using Cosine Similarity Instead of Dot Product in Neural Networks*. arXiv:1702.05870 [cs.LG]
- [6] Jiaqi Ma, Zhe Zhao, Xinyang Yi, Jilin Chen, Lichan Hong, and Ed H. Chi. 2018. Modeling Task Relationships in Multi-task Learning with Multi-gate Mixture-of-Experts. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '18)*. 1930–1939.
- [7] Xiao Ma, Liqin Zhao, Guan Huang, Zhi Wang, Zelin Hu Ram, Xiaoqiang Zhu, and Kun Gai. 2018. Entire Space Multi-Task Model: An Effective Approach for Estimating Post-Click Conversion Rate. In *Proceedings of the 41th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '18)*. 1137–1140.
- [8] Tobias Schnabel, Adith Swaminathan, Ashudeep Singh, Navin Chandak, and Thorsten Joachims. 2016. Recommendations as treatments: Debiasing learning and evaluation.. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning (ICML '16)*. 1670–1679.
- [9] Hongyan Tang, Junning Liu, Ming Zhao, and Xudong Gong. 2020. Progressive Layered Extraction (PLE): A Novel Multi-Task Learning (MTL) Model for Personalized Recommendations. In *Proceedings of the 14th ACM Conference on Recommender Systems (RecSys'20)*. 269–278.
- [10] Robert R. Tucci. 2013. *Introduction to Judea Pearl's Do-Calculus*. arXiv:1305.5506 [cs.AI]
- [11] Hao Wang, Tai-Wei Chang, Tianqiao Liu, Jianmin Huang, Zhichao Chen, Chao Yu, Ruopeng Li, and Wei Chu. 2022. ESCM2: Entire Space Counterfactual Multi-Task Model for Post-Click Conversion Rate Estimation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '22)*. 363–372.
- [12] Xiaojie Wang, Rui Zhang, Yu Sun, and Jianzhong Qi. 2019. Doubly Robust Joint Learning for Recommendation on Data Missing Not at Random. In *Proceedings of the 36th International Conference on Machine Learning (ICML '19)*. 6638–6647.
- [13] Zhiqiang Wang, Qingyun She, and Junlin Zhang. 2021. MaskNet: Introducing Feature-Wise Multiplication to CTR Ranking Models by Instance-Guided Mask. In *Proceedings of DLP-KDD 2021*.
- [14] Wenhao Zhang, Wentian Bao, Xiao-Yang Liu, Keping Yang, Quan Lin, Hong Wen, and Ramin Ramezani. 2020. Large-scale Causal Approaches to Debiasing Post-click Conversion Rate Estimation with Multi-task Learning. In *Proceedings of the Web Conference 2020 (WWW '20)*. 2775–2781.
- [15] Jiaqi Zheng, Cheng Guo, Yi Cao, Chaoqun Hou, Tong Liu, and Bo Zheng. 2025. USD: A User-Intent-Driven Sampling and Dual-Debiasing Framework for Large-Scale Homepage Recommendations. In *Proceedings of the Nineteenth ACM Conference on Recommender Systems (ReSys '25)*. 1108–1111.
- [16] Yu Zheng, Chen Gao, Xiang Li, Xiangnan He, Depeng Jin, and Yong Li. 2021. Disentangling User Interest and Conformity for Recommendation with Causal Embedding. In *Proceedings of the Web Conference 2021 (WWW '21)*. 2980–2991.
- [17] Yu Zheng, Chen Gao, Xiang Li, Xiangnan He, Yong Li, and Depeng Jin. 2021. Disentangling User Interest and Conformity for Recommendation with Causal Embedding. In *Proceedings of the Web Conference 2021 (WWW '21)*. 2980–2991.
- [18] Guorui Zhou, Xiaoqiang Zhu, Chenru Song, Ying Fan, Han Zhu, Xiao Ma, Yanghui Yan, Junqi Jin, Han Li, and Kun Gai. 2018. Deep Interest Network for ClickThrough Rate Prediction. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (SIGKDD '18)*. 1059–1068.
- [19] Feng Zhu, Mingjie Zhong, Xinxing Yang, Longfei Li, Lu Yu, Tiejia Zhang, Jun Zhou, Chaochao Chen, Fei Wu, Guanfeng Liu, and Yan Wang. 2023. DCMT: A Direct Entire-Space Causal Multi-Task Framework for Post-Click Conversion Estimation. In *Proceedings of the 39th IEEE International Conference on Data Engineering (ICDE '23)*.