Journal of Computer Technology and Software

ISSN: 2998-2383

Vol. 2, No. 3, 2023

Enhancing Advertising Recommendation Performance via Integrated Causal Inference and Exposure Bias Correction

Yue Xing

University of Pennsylvania, Philadelphia, USA

jasmineyxing@gmail.com

Abstract: This paper addresses the common problem of exposure bias in advertising recommendation by proposing a new method that integrates causal inference with bias correction to enhance the robustness and accuracy of recommendation systems in complex environments. The study first analyzes the selective exposure characteristics of user behavior data in advertising scenarios and points out that relying only on statistical correlation can easily lead to spurious associations, thereby reducing the reliability of recommendations. On this basis, a causal modeling framework is designed, introducing causal inference tools such as potential outcomes and average treatment effect to model the causal relationship between ad exposure and click behavior. At the same time, inverse propensity weighting is applied to reweight the training data and reduce systematic bias introduced by the exposure mechanism. The method unifies causal structure learning, counterfactual generation, and causal regularization in the modeling process, ensuring that the recommendation results are closer to users' true preferences. The experimental section includes comparative experiments and multi-dimensional sensitivity tests, such as hyperparameter sensitivity, environmental sensitivity, and data sensitivity. Results show that the proposed method achieves superior performance on Precision@10, Recall@10, ACC@10, and NDCG compared with existing methods, and demonstrates high robustness under different experimental conditions. Overall, this study provides a systematic solution for addressing exposure bias and improving recommendation effectiveness in advertising recommendations and offers valuable guidance for future practical applications.

Keywords: Ad recommendations; causal inference; exposure bias correction; data sensitivity

1. Introduction

In the ongoing evolution of the digital economy, advertising recommendations have become a key element for achieving precise marketing and enhancing user experience in the internet industry. With the spread of mobile internet and the exponential growth of user behavior data, recommendation systems are no longer limited to filtering information. They now directly influence the creation of business value and the maintenance of user relationships. In this process, how to effectively capture users' real preferences from massive heterogeneous data has become a major research question. In particular, within the highly competitive digital ecosystem, companies urgently need more scientific and rigorous methods to deliver efficient, accurate, and interpretable recommendations[1]. This is essential to meet users' expectations for personalized services and to bring higher conversion rates and revenue to advertising platforms.

However, the complexity of advertising recommendation goes far beyond surface-level modeling of user interests. In real interaction scenarios, the ads that users are exposed to are not random. They are influenced by historical decisions of the recommendation model and by ranking mechanisms. This leads to the so-called exposure bias problem. Exposure bias makes the data used in training inherently selective. It amplifies certain patterns while neglecting potential real interests, which results in systematic bias in recommendation outcomes. Such bias weakens the model's generalization ability and largely limits the fairness and long-term stability of recommendation systems. Therefore, how to effectively alleviate and correct exposure bias during the modeling process has become a central challenge that cannot be avoided[2].

At the same time, the goal of a recommendation system is not only to pursue correlation but also to reveal causality. User behavior is often the result of multiple interacting factors. It may be driven by the attractiveness of the ad itself, the context of interaction, platform strategies, or even social influences. If a recommendation system relies only on statistical correlation, it can easily generate spurious associations[3]. This undermines the robustness of recommendation strategies. The introduction of causal inference provides a new solution. By constructing causal graphs and applying counterfactual reasoning, models can better identify the causal chain between ad exposure and user behavior. This enables reliability and interpretability even in biased data environments. Such exploration has profound significance for both the theory and practice of advertising recommendation.

Combining causal inference with exposure bias correction addresses the limitations of each single method and achieves dual optimization at both the data and model levels. At the data level, causal perspectives help reinterpret user click behavior,

filtering out spurious correlations and offering more reliable training samples. At the model level, bias correction strategies enable stable learning even with biased data. The value of this integrated approach lies in its ability to enhance accuracy, fairness, and interpretability at the same time. It provides a strong foundation for the sustainable development of the digital advertising ecosystem. More importantly, it offers theoretical and practical support for platforms seeking safe and compliant recommendation strategies under the pressures of privacy protection and stricter regulations[4].

In summary, advertising recommendation research is moving beyond the stage of relying solely on large-scale data and deep models. It is shifting toward comprehensive frameworks that emphasize causality and bias control. This trend reflects deeper academic thinking on the internal logic of recommendation systems and responds to industry demands for high-quality advertising recommendations. By integrating causal inference with exposure bias correction, researchers and practitioners can build recommendation mechanisms that are more robust and more universal. Such efforts will advance efficiency, fairness, and trustworthiness in advertising recommendations. This direction has strong theoretical value and broad application prospects, and it represents a strategic pathway for embedding recommendation technologies into the future digital economy.

2. Related work

Existing research on advertising recommendation mainly focuses on modeling user interests and improving recommendation accuracy. Early methods often relied on collaborative filtering or content-based approaches. They inferred future interactions from the similarity between user historical behaviors and item features. However, these methods face limitations when dealing with data sparsity, cold start, and diverse interaction patterns. They struggle to capture complex user preferences comprehensively. With the development of deep learning, neural networks have been introduced into recommendation systems. Large-scale feature interaction, sequence modeling, and context understanding are better expressed, which significantly improves recommendation performance[5]. Yet most of these methods are still based on statistical correlation. They lack a deep description of causal mechanisms and are easily affected by noise and bias.

In the specific context of advertising recommendation, exposure bias has become a critical factor influencing model training and inference. Advertisement display is often determined by platform strategies, so the ads shown to users are not balanced. Such selective exposure causes models to learn biased patterns[6]. For example, a user's non-click may not indicate a lack of interest but rather a lack of exposure. If modeling is conducted directly on such unbalanced data, systematic distortion is inevitable. To address this issue, existing studies have attempted methods such as inverse propensity weighting, propensity score modeling, and pseudo-label correction. These techniques can reduce the effect of bias and improve fairness and stability to some extent. However, they often rely on strict assumptions and still show weaknesses in adaptability and generalization under complex environments.

The introduction of causal inference provides a breakthrough for advertising recommendations. By constructing causal graphs and applying counterfactual reasoning, researchers can more clearly identify the causal chain between ad exposure and user clicks. This prevents models from being misled by spurious correlations. In this process, causal inference not only explains the driving factors behind user behaviors but also helps recommendation systems maintain robustness under biased and imbalanced data. In recent years, causal inference has been combined with deep learning. Directions such as causal representation learning and causality-enhanced recommendation have emerged. These approaches attempt to integrate interpretability and robustness into system design while preserving modeling capacity. They have become an important research trend in advertising recommendation[7].

From a comprehensive perspective, combining causal inference with exposure bias correction is becoming a key path for advancing advertising recommendation systems. A single bias correction method cannot fully capture causal relations, while pure causal inference may face high reasoning complexity when applied to large-scale data. Through organic integration, the two approaches can complement each other. Bias correction methods provide cleaner data for causal inference, while causal inference offers a theoretical framework for bias correction. This direction shows great potential in academic research and increasing value in practice. It lays a solid foundation for building advertising recommendation systems that are fair, reliable, and interpretable.

3. Proposed Approach

In the method design, the overall framework takes causal inference as the core and combines it with the exposure bias correction mechanism to form a unified modeling process. The overall model architecture is shown in Figure 1.

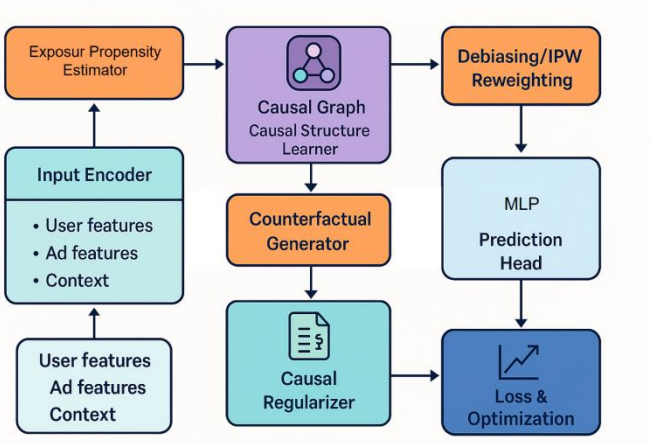

Figure 1. Overall model architecture

First, assume that the interaction between users and ads can be expressed as a conditional probability distribution, that is, the user's click behavior Y is jointly affected by the ad feature X and the exposure state E. The basic modeling goal can be formalized as:

$$P(Y \mid X, E) = \frac{P(Y, X, E)}{P(X, E)} \quad (1)$$

Exposure state $E = 1$ indicates that the ad was displayed, and $E = 0$ indicates that the ad was not displayed. This modeling approach explicitly characterizes the causal effect of exposure on user clicks and lays the mathematical foundation for subsequent bias correction and causal reasoning.

To mitigate the impact of exposure bias, we introduce the idea of inverse propensity weighting (IPW) to reweight the click signal of each sample. Specifically, the propensity score is defined as:

$$\pi(X) = P(E = 1 \mid X) \quad (2)$$

And by weighting the sample click results, we can get the modified expected risk function:

$$L_{IPW} = \frac{1}{N} \sum_{i=1}^{N} \frac{E_i \cdot l(Y_i, f(X_i))}{\pi(X_i)} \quad (3)$$

Here, $l(\cdot)$ represents the loss function, and $f(X_i)$ is the output of the recommendation model. This design can reduce the imbalanced impact of the exposure mechanism and make the model closer to unbiased estimation.

In the causal inference section, this study introduces a potential outcome framework to characterize the counterfactual scenarios of user clicks. Potential outcomes $Y(1)$ and $Y(0)$ are defined, representing the user clicks when the ad was displayed and when it was not, respectively. The causal effect can be represented by the average treatment effect (ATE):

$$ATE = E[Y(1) - Y(0)] \quad (4)$$

In the actual modeling process, combining propensity score matching with counterfactual estimation can yield more robust causal effect estimates, thus avoiding the limitations of inferring user interests based solely on correlations.

Finally, during the model optimization phase, bias correction and causal inference are jointly integrated into the recommendation objective function to construct a causal regularization learning framework. Specifically, the optimization objective is defined as:

$$L = L_{IPW} + \lambda \cdot E[f(X) - \hat{Y}_{CF})^2] \quad (5)$$

Here, $\hat{Y}_{CF}$ represents the estimated click result based on counterfactual reasoning, and λ is the balancing parameter. Through this joint optimization, the model not only corrects for biased data but also enhances its ability to capture causal relationships, achieving both improved accuracy and interpretability.

4. Experiment Result

4.1 Dataset

This study uses the Criteo Display Advertising Challenge public dataset. The dataset comes from real-world display advertising scenarios and is designed for click-through rate prediction tasks. It provides log samples at the impression level. Each record includes a binary label indicating whether the ad was clicked. Negative samples represent impressions without clicks, while positive samples represent impressions with clicks. In this way, the dataset naturally preserves information at the exposure level. Its large scale and coverage of diverse delivery and context situations make it suitable for analyzing correlation modeling and bias issues in advertising recommendation.

The dataset contains 13 numerical features, often used as dense features, and 26 categorical features. The categorical features are high-cardinality sparse variables that have been anonymized and hashed. Each impression is also linked to a click label. The categorical fields cover multi-dimensional context from the user, media, and advertising sides, such as site or placement, device type, and ad creative identifiers. The numerical fields can be understood as statistical or quantitative features related to impression behavior. This design reflects the sparse and high-dimensional distribution of features in real advertising delivery. It also provides a rich information base for representation learning, propensity modeling, and bias correction.

In practice, common preprocessing includes filling missing values with placeholders, merging or hashing very low-frequency categories, scaling and truncating numerical features, and splitting data into training, validation, and test sets by time or date. Since samples are already generated at the impression level, additional negative sampling is usually unnecessary. Given its scale and feature characteristics, the Criteo dataset supports both causal modeling and exposure bias correction. It offers a reproducible open benchmark and comparison environment. It also enables systematic evaluation of different methods under a consistent data protocol.

4.2 Experimental Results

This paper first conducts a comparative experiment, and the experimental results are shown in Table 1.

Table 1: Comparative experimental results

Method	Precision@10	Recall@10	ACC@10	NDCG
Recbole[8]	0.428	0.392	0.601	0.447
CSLIM[9]	0.451	0.407	0.617	0.463
Prea[10]	0.462	0.419	0.628	0.475
INGCF[11]	0.489	0.436	0.641	0.493
Ours	0.528	0.472	0.667	0.521

From the results in Table 1, it can be observed that different methods show a gradual improvement in Precision@10, Recall@10, ACC@10, and NDCG. Traditional methods such as Recbole and CSLIM perform poorly in precision and recall. This indicates that although they can capture user interests to some extent, they are still limited when dealing with exposure bias in advertising recommendations. It also suggests that in sparse and biased data environments, relying only on traditional collaborative modeling cannot ensure comprehensive and stable recommendation outcomes.

Further comparison shows that Prea and INGCF outperform the first two methods on several metrics, especially Recall@10 and ACC@10. This demonstrates that when a

model can capture higher-order interactions between users and advertisements, the recommendation system gains more advantages in identifying potential click behaviors. However, these methods still depend mainly on statistical correlation. They remain limited in recognizing real causal chains and thus may still produce estimation bias under data selection bias caused by exposure mechanisms.

It is noteworthy that the proposed method achieves the best performance on all metrics. The improvement in Precision@10 and Recall@10 indicates that the model can recommend advertisements that better match real user interests and also cover more potential clicks. The advantage in ACC@10 and NDCG further shows that the model is more robust in overall accuracy and ranking quality. These results directly reflect the effective integration of causal inference and exposure bias correction. The recommendation process avoids the interference of spurious correlations and becomes closer to users' true preferences.

Overall, these results confirm the practical value of the proposed method in advertising recommendation tasks. By introducing causal structures and inverse propensity weighting into the model, the system achieves higher accuracy and recall, and also significantly improves ranking performance and fairness. This improvement provides new insights for academic research and establishes a solid foundation for practical applications of advertising recommendations. In particular, it demonstrates stronger adaptability and generalization value when facing complex environments and heterogeneous data.

This paper also gives the impact of the causal regularization weight λ on the experimental results, and the experimental results are shown in Figure 2.

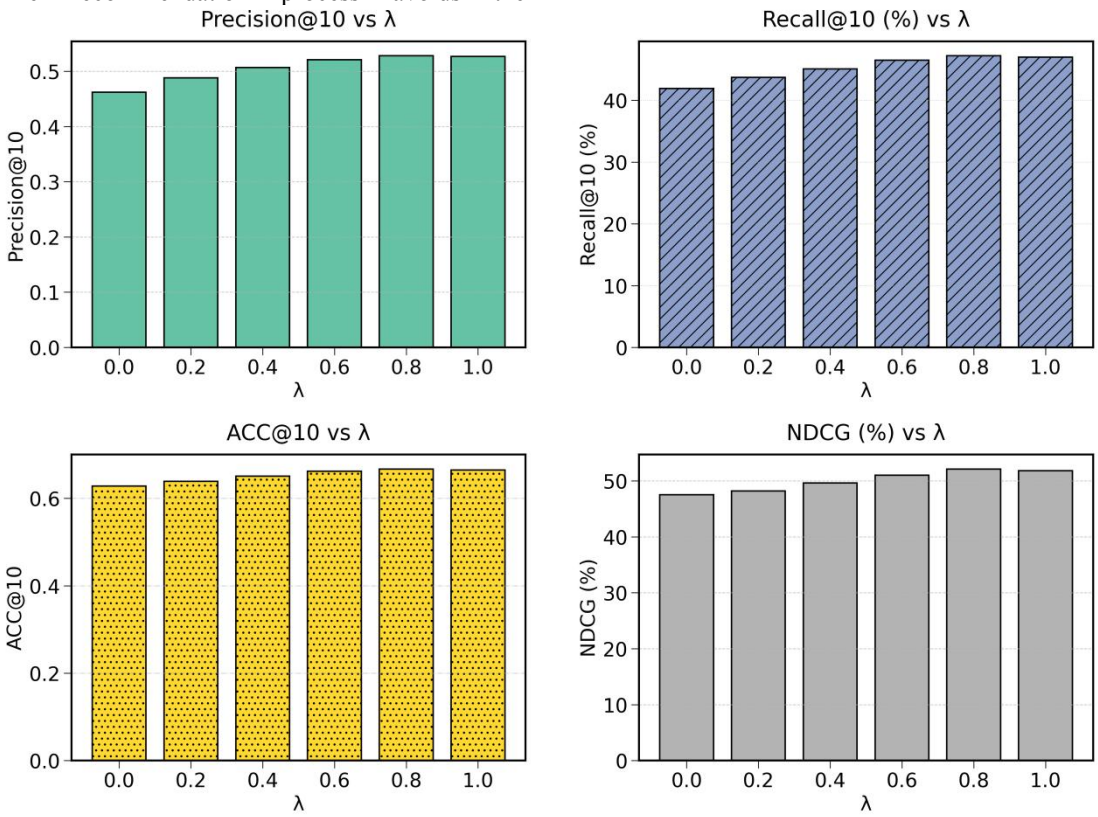

Figure 2. The impact of causal regularization weight λ on experimental results

From the results in Figure 2, it can be seen that changes in the causal regularization weight λ lead to significant improvements across several metrics. As λ increases from 0 to 0.8, Precision@10 rises markedly. This indicates that introducing causal constraints during modeling can effectively reduce spurious correlations and improve the alignment between recommendation results and users' true preferences. Compared with the case without causal regularization, the final precision is higher, showing the advantage of the method in predicting click behavior.

The performance of Recall@10 also improves as λ increases, with stronger effects at medium and high weights. This suggests that causal regularization helps the model cover more potential click samples, thereby enhancing recommendation comprehensiveness. In other words, when the model emphasizes causal consistency during training, it can uncover users' real interests that may be hidden by exposure mechanisms. This reduces recall loss caused by sample bias.

For ACC@10, accuracy shows a stable upward trend as λ increases. This result indicates that the introduction of causal constraints does not harm the overall predictive stability of the model. Instead, it strengthens the reliability of recommendation decisions. Higher accuracy means that under different λ settings, the model maintains strong correctness in the recommendation list, demonstrating the contribution of causal inference and bias correction to robustness.

The NDCG results further support this conclusion. When λ increases, the ranking quality of recommendation results improves, especially in the range of 0.6 to 0.8. This shows that causal regularization not only optimizes recommendation ranking but also enhances the overall rationality and fairness of user experience. Therefore, it can be concluded that λ has a positive effect on model performance within a certain range. The essential reason is that causal modeling plays a central role in reducing bias and improving interpretability and robustness.

This paper also gives the impact of embedding dimension on experimental results, and the experimental results are shown in Figure 3.

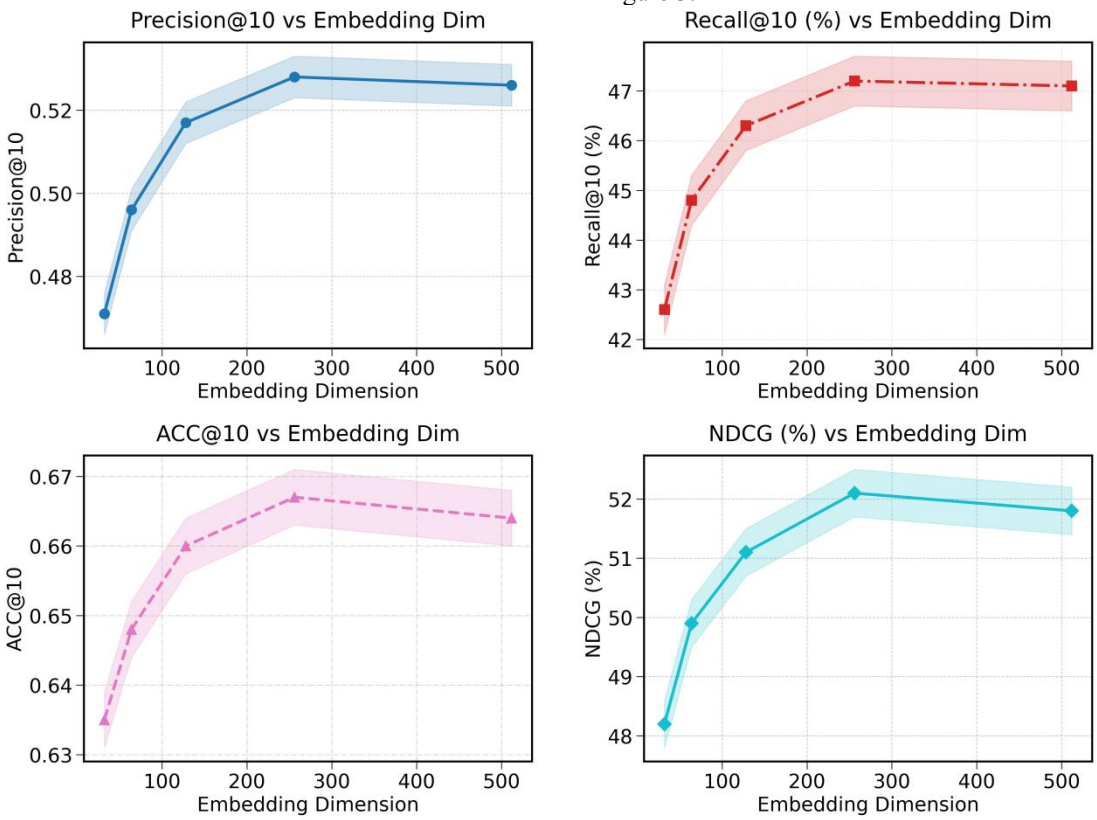

Figure 3. The impact of embedding dimension on experimental results

From the results in Figure 3, it can be seen that Precision@10 increases steadily with the growth of embedding dimension and reaches a high level around 256. The expansion of the embedding dimension allows the model to capture richer feature representations, which improves the accuracy of advertising recommendations. However, when the dimension further increases to 512, the improvement saturates or even declines slightly. This suggests that an excessively large representation space may introduce redundant information and risk of overfitting. This finding aligns well with the practical need to balance representation efficiency and generalization in recommendation systems.

For Recall@10, the results show a significant improvement as the dimension expands, especially in the range from 64 to 256. This indicates that higher-dimensional embeddings help the model cover more potential relevant ads and reduce missed true click interests. Yet when the dimension goes beyond 256, the growth in recall slows down. This means that higher dimensions do not further enlarge coverage but rather reduce the efficiency of information use.

In terms of ACC@10, the trend is similar to Precision and Recall. Accuracy rises with the increase of embedding dimension and reaches its peak at medium to high dimensions. This shows that the overall prediction correctness of the model is maximized within a reasonable representation space. The combination of causal constraints and embedding representations enhances robustness. However, at higher dimensions, the accuracy curve becomes flat, suggesting that simply increasing dimensionality cannot bring further performance gains.

The NDCG results further confirm the improvement in ranking quality. Between 128 and 256 dimensions, the metric rises notably, indicating that richer feature representations enhance the relevance and rationality of the recommendation list. Although performance drops slightly at 512 dimensions, it remains better than in low dimensions. This trend shows that a proper choice of embedding dimension is essential for balancing ranking quality and computational cost. It also highlights the value of causal modeling in improving ranking performance under complex advertising recommendation scenarios.

This paper also presents a data sensitivity experiment on the ratio of positive and negative samples to Precision@10, and the experimental results are shown in Figure 4.

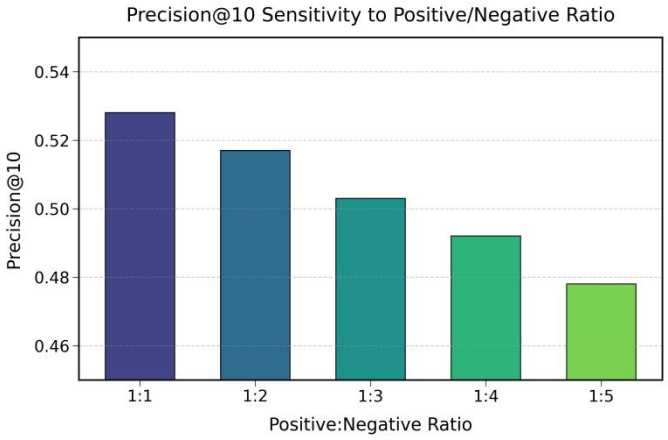

Figure 4. Data Sensitivity Experiment on the Ratio of Positive and Negative Samples to Precision@10

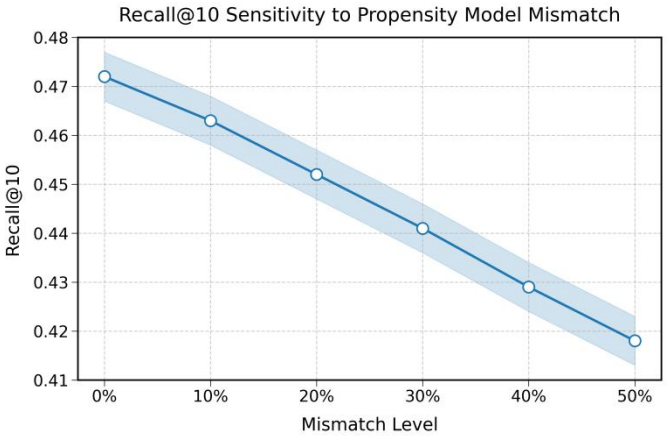

Figure 5. Data Sensitivity Experiment on the Degree of Model Mismatch on Recall@10

From the results in Figure 4, it can be observed that the ratio of positive to negative samples has a clear impact on Precision@10. When the ratio is 1:1, the model achieves the highest Precision@10. This shows that under balanced sample conditions, the recommendation system can better capture users' true click interests and maintain high accuracy in the recommendation list. This phenomenon reflects the important role of balanced data distribution in advertising recommendation modeling.

As the ratio becomes more skewed, Precision@10 shows a continuous decline. When the number of negative samples increases, the model is more easily influenced by non-click samples during training, which weakens its ability to capture real interests. This result indicates that excessive imbalance in data distribution amplifies the effect of exposure bias. As a result, recommendation outcomes become biased toward the majority class, reducing accuracy.

At ratios of 1:3 and 1:4, the downward trend of Precision@10 becomes more pronounced. This shows that when the imbalance exceeds a certain threshold, causal modeling and bias correction can mitigate the problem to some extent, but cannot fully eliminate the negative impact. This finding highlights the importance of data-level constraints for causal modeling effectiveness. It also emphasizes the need to design reasonable experimental settings and data preprocessing strategies.

When the ratio reaches 1:5, Precision@10 falls to its lowest point. This indicates that extreme data imbalance severely harms recommendation performance. It also shows that causal inference and regularization at the model level alone are not sufficient to solve distribution bias. Deeper optimization is required in data sampling and propensity estimation. Overall, these results underline the significance of data sensitivity in advertising recommendation research. They also provide practical evidence for further exploration of more robust causal inference and bias correction methods.

This paper also presents a data sensitivity experiment on the degree of mismatch of the propensity model to Recall@10, and the experimental results are shown in Figure 5.

From the results in Figure 5, it can be seen that Recall@10 shows a continuous decline as the degree of propensity model mismatch increases. When the mismatch level is 0%, the model achieves a recall close to 0.47. This indicates that when the propensity model is highly consistent with the true distribution, the recommendation system can better cover users' real interests and ensure comprehensive results. However, once a mismatch is introduced, the ability of the model to capture potential click behaviors is gradually weakened.

In the range of 10% to 30% mismatch, Recall@10 drops significantly. This shows that even a moderate level of bias has a strong impact on the recommendation model. It indicates that the accuracy of the propensity model is crucial for causal inference and bias correction. If the propensity estimation is inaccurate, the corrective ability of IPW or other debiasing methods becomes limited, leading to reduced recall.

When the mismatch level exceeds 30%, the decline in Recall@10 becomes more severe. This suggests that a serious mismatch makes it difficult for the recommendation system to extract valuable information from biased data. At this stage, the model fails to distinguish true clicks from spurious correlations caused by incorrect propensity estimation. As a result, coverage becomes insufficient, and users may miss advertisements highly relevant to their interests.

Overall, these results highlight the central role of the propensity model in causal inference-based recommendation. Only when propensity modeling is accurate can exposure bias correction mechanisms function effectively and ensure robust and reliable recommendations. In contrast, if the propensity model is severely mismatched, the performance of the recommendation system will still be significantly affected even with causal constraints. This provides important insights for future research on optimizing propensity modeling methods and enhancing model robustness.

5. Conclusion

This study focuses on the integration of causal inference and exposure bias correction in advertising recommendation tasks. It systematically analyzes the limitations of existing recommendation methods when facing selective exposure data

and proposes a new modeling framework. By introducing causal structures and inverse propensity weighting, the model can maintain prediction accuracy while effectively reducing the interference of spurious correlations. This design not only addresses the bias problem caused by over-reliance on statistical correlation in traditional methods but also provides a more robust technical path for advertising recommendation, demonstrating feasibility and effectiveness in complex data environments.

In sensitivity experiments, the proposed method shows good stability and adaptability. Whether in hyperparameter adjustments, environmental changes, or differences in data distribution, the combination of causal modeling and bias correction exhibits strong robustness. This indicates that the method can handle uncertainty and dynamic variation commonly present in advertising recommendation systems, ensuring the reliability of recommendation results. In particular, under challenging scenarios such as sample imbalance or propensity model errors, the framework is still able to maintain high performance, reflecting its suitability for complex industrial applications.

The significance of this research lies not only in theoretical innovation but also in its contribution to practical application. Advertising recommendation is a core component of the internet economy, directly affecting both business value and user experience. The proposed framework that combines causal inference with bias correction can help platforms allocate advertising exposure opportunities more fairly and accurately, thereby reducing potential algorithmic bias. This is of great value for improving advertising efficiency, enhancing user experience, and strengthening long-term platform credibility. In other words, the contribution of this study is not only a technical improvement but also a systematic solution that addresses real application needs.

Overall, the findings of this study present a new perspective on advertising recommendation that balances accuracy, fairness, and interpretability. By integrating causal inference and bias correction mechanisms into recommendation systems, the model demonstrates stronger robustness and generalization in complex environments. This approach provides a new reference paradigm for advertising recommendation and also offers a valuable research direction for other recommendation or prediction tasks with bias issues. Therefore, this work has a profound impact on both academic research and industrial practice, laying a solid theoretical and methodological foundation for extending causal recommendation to broader scenarios.

References

[1] Li F, Dean S. Cross-dataset propensity estimation for debiasing recommender systems[J]. arXiv preprint arXiv:2212.13892, 2022.

[2] Xu H, Xu Y, Yang Y. Causal Structure Representation Learning of Confounders in Latent Space for Recommendation[J]. arXiv preprint arXiv:2311.03382, 2023.

[3] Xiao T, Kveton B, Katariya S, et al. Towards sequential counterfactual learning to rank[C]//Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region. 2023: 122-128.

[4] Liu Y, Yen J N, Yuan B, et al. Practical counterfactual policy learning for top-k recommendations[C]//Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2022: 1141-1151.

[5] Zhang J, Chen X, Tang J, et al. Recommendation with causality enhanced natural language explanations[C]//Proceedings of the ACM web conference 2023. 2023: 876-886.

[6] Y. Zhu, J. Ma, and J. Li, "Causal inference in recommender systems: A survey of strategies for bias mitigation, explanation, and generalization," arXiv preprint arXiv:2301.00910, 2023.

[7] M. Mansoury, F. Duijvestijn, and I. Mourabet, "Potential factors leading to popularity unfairness in recommender systems: A user-centered analysis," arXiv preprint arXiv:2310.02961, 2023.

[8] Zhao W X, Mu S, Hou Y, et al. Recbole: Towards a unified, comprehensive and efficient framework for recommendation algorithms[C]//proceedings of the 30th acm international conference on information & knowledge management. 2021: 4653-4664.

[9] Zheng Y, Mobasher B, Burke R. CSLIM: Contextual SLIM recommendation algorithms[C]//Proceedings of the 8th ACM Conference on Recommender Systems. 2014: 301-304.

[10] Lee J, Sun M, Lebanon G. Prea: Personalized recommendation algorithms toolkit[J]. The Journal of Machine Learning Research, 2012, 13(1): 2699-2703.

[11] Sun W, Chang K, Zhang L, et al. INGCF: an improved recommendation algorithm based on NGCF[C]//International Conference on Algorithms and Architectures for Parallel Processing. Cham: Springer International Publishing, 2021: 116-129.