Supplementary materials

Table 1: Performance comparison of efficiency with MF as backbone on the Yelp2018 dataset.

Model	Memory	T/Epoch	# Epoch	# T	R@20	N@20
MF	820M	2.43s	51	123.93s	0.0382	0.0338
+ IPS	820M	3.30s	21	69.30s	0.0400	0.0363
+ CausE	864M	3.37s	302	1,017.74s	0.0362	0.0335
+ MACR	830M	5.09s	42	213.78s	0.0507	0.0458
+ PDA	824M	3.93s	117	459.81s	0.0619	0.0564
+ DICE	944M	10.56s	21	221.76s	0.0449	0.0409
+ REL	822M	3.62s	43	155.66s	0.0540	0.0485
+ uCTRL	2,896M	147.11s	323	47,516.53s	0.0528	0.0327
+ PPAC	10,024M	8.62s	84	724.08s	0.0488	0.0442
+ APWCF	828M	3.56s	121	430.76s	0.0689	0.0640

Table 2: Performance comparison of efficiency with LightGCN as backbone on the Yelp2018 dataset.

Model	Memory	T/Epoch	# Epoch	# T	R@20	N@20
LightGCN	1,140M	12.12s	182	2,205.84s	0.0772	0.0710
+ IPS	1,140M	11.36s	236	2,680.96s	0.0771	0.0710
+ CausE	1,236M	62.45s	21	1,311.45s	0.0519	0.0474
+ MACR	1,148M	64.24s	211	13,554.64s	0.0784	0.0723
+ PDA	1,140M	15.26s	500	7,630s	0.0780	0.0719
+ DICE	1,550M	61.49s	39	2,398.11s	0.0540	0.0493
+ REL	1,140M	18.41s	26	478.66s	0.0494	0.0450
+ uCTRL	2,054M	116.91s	220	25,720.20s	0.0741	0.0687
+ PPAC	10,302M	11.29s	327	3,691.83s	0.0621	0.0716
+ APWCF	1,148M	14.12s	186	2626.32s	0.0786	0.0732

A Implementation detail

We implement all models using RecBole-Debias [1] [Zhao *et al.*, 2021; Zhao *et al.*, 2022; Xu *et al.*, 2023], a unified open-source framework for reproducing recommendation models. We follow the settings of classic debiased recommendation models ([Zheng *et al.*, 2021; Bonner and Vasile, 2018; Wei *et al.*, 2021; Lee *et al.*, 2023]). In order to make a more realistic evaluation of the model performance under non-IID conditions, we perform intervention processing and partitioning on the dataset. First, the original data is divided into 50% normal data and 50% candidate sets, and the candidate sets are resampled based on the inverse popularity of items to make their distribution uniform. Subsequently, the data is divided into a training set (100% normal data + 25% intervention data), a validation set (25% intervention data), and a test set (50% intervention data), with the final ratio being approximately 62.5%:12.5%:25%. In addition, following the settings of most methods, we also apply the all-ranking protocol to comprehensively assess the recommendation accuracy of various models when performing complete ranking across the entire item set. And we use Recall and NDCG, two widely used evaluation metrics, to evaluate the effectiveness of debiased recommendation.

[1] https://github.com/RUCAIBox/RecBole

B Training efficiency analysis

In order to evaluate the effectiveness of the proposed APWCF method in training compared with the baseline, we use MF and LightGCN as the backbone, test on the Yelp2018 dataset and record the memory overhead, training time of a single epoch, total epochs required for training, total training time, and performance indicators Recall@20 and NDCG@20 of each method. We report the results in Tables 1 and 2. According to the results, we find that under the same method, although the training time of LightGCN as the backbone is greatly increased compared with MF, its performance is far better than the latter. This may be because LightGCN can more comprehensively capture the potential association between users and items through high-order neighbor interaction propagation, providing more accurate feature representation for the debiasing method, making the debiasing process more effective. In addition, APWCF outperforms most methods in terms of space overhead. Moreover, compared with other baselines, APWCF also shows competitive results in terms of training time while maintaining the optimal performance indicators, especially when LightGCN is used as the backbone.

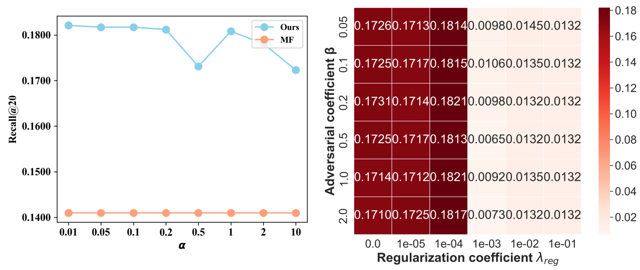

Figure 1: The performance of difference α and β of our method on KuaiRec dataset.

C Parameters sensitivity study

The effect of α and β. In APWCF, α controls the weight of the adversarial loss in the total loss and determines the strength of the adversarial learning module. β controls the strength of the gradient reversal and affects the ability of the bias discriminator to separate bias information. We fine-tune α on the KuaiRec dataset to 0.01, 0.05, 0.1, 0.2, 0.5, 1, 2, 10 and use MF as the backbone. We report the results of the validation set in Fig. 1. We find that when α is small, the model performance is limited, and the model performs best between 0.1 and 1. Further increasing α may lead to performance degradation. In addition, we fine-tune the regularization coefficient λ_{reg} in 0, 1e-05, 1e-04, 1e-03, 1e-02, 1e-01 and fine-tune β in 0.05, 0.1, 0.2, 0.5, 1.0, 2.0, and report the experimental results in Fig. 1. Compared with the parameter λ_{reg}, the hyperparameter β of APWCF shows stronger stability. Under the same λ_{reg}, the experimental results do not change much, but under the same β, changing λ_{reg} will cause a large fluctuation in the results.

References

[Bonner and Vasile, 2018] Stephen Bonner and Flavian Vasile. Causal embeddings for recommendation. In Sole Pera, Michael D. Ekstrand, Xavier Amatriain, and John O'Donovan, editors, *Proceedings of the 12th ACM Conference on Recommender Systems, RecSys 2018, Vancouver, BC, Canada, October 2-7, 2018*, pages 104–112. ACM, 2018.

[Lee *et al.*, 2023] Jae-woong Lee, Seongmin Park, Mincheol Yoon, and Jongwuk Lee. uctrl: Unbiased contrastive representation learning via alignment and uniformity for collaborative filtering. In Hsin-Hsi Chen, Wei-Jou (Edward) Duh, Hen-Hsen Huang, Makoto P. Kato, Josiane Mothe, and Barbara Poblete, editors, *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2023, Taipei, Taiwan, July 23-27, 2023*, pages 2456–2460. ACM, 2023.

[Wei *et al.*, 2021] Tianxin Wei, Fuli Feng, Jiawei Chen, Ziwei Wu, Jinfeng Yi, and Xiangnan He. Model-agnostic counterfactual reasoning for eliminating popularity bias in recommender system. In Feida Zhu, Beng Chin Ooi, and Chunyan Miao, editors, *KDD '21: The 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, Singapore, August 14-18, 2021*, pages 1791–1800. ACM, 2021.

[Xu *et al.*, 2023] Lanling Xu, Zhen Tian, Gaowei Zhang, Junjie Zhang, Lei Wang, Bowen Zheng, Yifan Li, Jiakai Tang, Zeyu Zhang, Yupeng Hou, Xingyu Pan, Wayne Xin Zhao, Xu Chen, and Ji-Rong Wen. Towards a more user-friendly and easy-to-use benchmark library for recommender systems. In *SIGIR*, pages 2837–2847. ACM, 2023.

[Zhao *et al.*, 2021] Wayne Xin Zhao, Shanlei Mu, Yupeng Hou, Zihan Lin, Yushuo Chen, Xingyu Pan, Kaiyuan Li, Yujie Lu, Hui Wang, Changxin Tian, Yingqian Min, Zhichao Feng, Xinyan Fan, Xu Chen, Pengfei Wang, Wendi Ji, Yaliang Li, Xiaoling Wang, and Ji-Rong Wen. Recbole: Towards a unified, comprehensive and efficient framework for recommendation algorithms. In *CIKM*, pages 4653–4664. ACM, 2021.

[Zhao *et al.*, 2022] Wayne Xin Zhao, Yupeng Hou, Xingyu Pan, Chen Yang, Zeyu Zhang, Zihan Lin, Jingsen Zhang, Shuqing Bian, Jiakai Tang, Wenqi Sun, Yushuo Chen, Lanling Xu, Gaowei Zhang, Zhen Tian, Changxin Tian, Shanlei Mu, Xinyan Fan, Xu Chen, and Ji-Rong Wen. Recbole 2.0: Towards a more up-to-date recommendation library. In *CIKM*, pages 4722–4726. ACM, 2022.

[Zheng *et al.*, 2021] Yu Zheng, Chen Gao, Xiang Li, Xiangnan He, Yong Li, and Depeng Jin. Disentangling user interest and conformity for recommendation with causal embedding. In Jure Leskovec, Marko Grobelnik, Marc Najork, Jie Tang, and Leila Zia, editors, *WWW '21: The Web Conference 2021, Virtual Event / Ljubljana, Slovenia, April 19-23, 2021*, pages 2980–2991. ACM / IW3C2, 2021.