PDF Download
3589335.3651472.pdf
17 January 2026
Total Citations: 5
Total Downloads: 167

Published: 13 May 2024

Citation in BibTeX format

WWW '24: The ACM Web Conference 2024
May 13 - 17, 2024
Singapore, Singapore

Conference Sponsors:
SIGWEB

SHORT-PAPER

Benchmarking News Recommendation in the Era of Green AI

QIJIONG LIU

JIEMING ZHU, Huawei Technologies Noah's Ark Lab, Hong Kong, Hong Kong

QUANYU DAI, Huawei Technologies Noah's Ark Lab, Hong Kong, Hong Kong

XIAOMING WU

Open Access Support provided by:

Huawei Technologies Noah's Ark Lab

WWW '24: Companion Proceedings of the ACM Web Conference 2024 (May 2024)
https://doi.org/10.1145/3589335.3651472
ISBN: 9798400701726

Benchmarking News Recommendation in the Era of Green AI

Qijiong Liu*
PolyU, Hong Kong
liu@qijiong.work

Jieming Zhu*
Huawei Noah's Ark Lab, China
jiemingzhu@ieee.org

Quanyu Dai
Huawei Noah's Ark Lab, China
quanyu.dai@connect.polyu.hk

Xiao-Ming Wu†
PolyU, Hong Kong
xiao-ming.wu@polyu.edu.hk

ABSTRACT

Over recent years, news recommender systems have gained significant attention in both academia and industry, emphasizing the need for a standardized benchmark to evaluate and compare the performance of these systems. Concurrently, Green AI advocates for reducing the energy consumption and environmental impact of machine learning. To address these concerns, we introduce the first Green AI benchmarking framework for news recommendation, known as GreenRec[1], and propose a metric for assessing the tradeoff between recommendation accuracy and efficiency. Our benchmark encompasses 30 base models and their variants, covering traditional end-to-end training paradigms as well as our proposed efficient only-encode-once (OLEO) paradigm. Through experiments consuming 2000 GPU hours, we observe that the OLEO paradigm achieves competitive accuracy compared to state-of-the-art end-to-end paradigms and delivers up to a 2992% improvement in sustainability metrics.

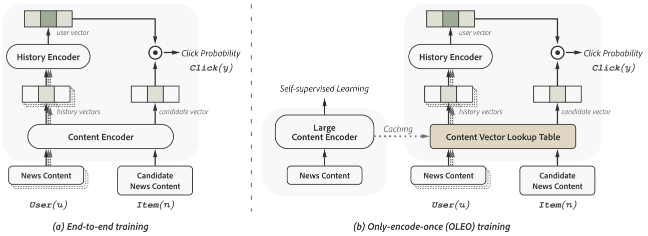
CCS CONCEPTS

• **Information systems → Recommender systems**.

KEYWORDS

news recommendation, green AI

ACM Reference Format:
Qijiong Liu, Jieming Zhu, Quanyu Dai, and Xiao-Ming Wu. 2024. Benchmarking News Recommendation in the Era of Green AI. In *Companion Proceedings of the ACM Web Conference 2024 (WWW '24 Companion), May 13–17, 2024, Singapore, Singapore.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3589335.3651472



(a) End-to-end training

(b) Only-encode-once (OLEO) training

Figure 1: Two training paradigms for news recommendation.

1 INTRODUCTION

News recommender systems have become a crucial tool in the digital age, helping users discover relevant news articles and stay informed amidst the ever-increasing number of sources available. To integrate news content knowledge into recommender systems, content encoders are employed to capture semantic news representations. Various approaches have been explored in these efforts, including simple content encoders represented by CNN [5] and Attention-based [13] models [1, 7, 16, 18], as well as emerging content encoders based on pretrained language models [8, 22].

Despite the success achieved by these studies, there is still a notable absence of standardized benchmarks and uniform evaluation protocols for news recommendation. Consequently, even when common datasets like MIND [20] are employed for evaluation, existing studies often use their own data partitions and apply unique preprocessing steps. The absence of uniformity in data preprocessing methods leads to non-reproducible and often conflicting experimental outcomes across different studies, and the use of nonstandard data preprocessing techniques makes it challenging to compare and evaluate results between papers.

Meanwhile, the environmental impact has also begun to draw attention from the machine learning community, as training large-scale networks leads to a huge increase of carbon emission [12]. Therefore, a series of studies [9, 15] explore ways to reduce energy consumption in line with Green AI principles [10].

In light of this, we establish the first Green AI Benchmark for News Recommendation, called **GreenRec**, and introduce a sustainability metric, namely unit conversion rate of carbon emission, to strike a balance between recommendation quality and efficiency. Specifically, we selected six base models and considered five different variants for each model. In total, we evaluated the recommendation accuracy and environmental impact across a total of 30 baselines. This evaluation was carried out under two distinct training paradigms: the ***conventional end-to-end training paradigm***,

*Both authors contributed equally to this research (co-first authors).
†Corresponding author.
[1] `GreenRec` has been published as part of the `Legommenders` benchmark, which is a modular framework for recommender systems. The source code and data for GreenRec can be found at https://github.com/Jyonn/Legommenders.

WWW '24 Companion, May 13–17, 2024, Singapore, Singapore

ACM ISBN 979-8-4007-0172-6/24/05
https://doi.org/10.1145/3589335.3651472

where the content encoder and other recommender components are jointly trained, and our ***proposed only-encode-once (OLEO) training paradigm***, where the content encoder and other components are trained in a decoupled manner. Our evaluation is conducted through hundreds of experiments within a standardized framework using the most widely used news recommendation datasets, MIND-small and MIND-large [20]. The experiments reveal that the conventional end-to-end paradigm encounters efficiency challenges, while our OLEO-based paradigm is much more eco-friendly. Moreover, the OLEO-based model variants maintain competitive performance with the state-of-the-art end-to-end PLM-NR variants [19], and achieve up to **2992%** improvement in our sustainability metric.

2 NEWS RECOMMENDATION

2.1 Basics of News Recommender Systems

News recommender systems are devised to predict the click probability of a user over a piece of given news article. Typically, a general news recommender system consists of three main components: content encoder, history encoder, and interaction module. As depicted in Figure 1 (a), the recommender accepts a news-user pair as input and yields the click probability. The content encoder is responsible for capturing the semantic meaning and contextual information of the news articles, while the history encoder processes the news history vectors, capturing the preferences, interests, and browsing patterns of users. Finally, the interaction module facilitates the interaction between the candidate news vector and user vector to calculate the click probability for the given news-user pair.

2.2 Training Paradigms and Complexity

Traditional models [16–19] typically follow such **end-to-end training paradigm**, as displayed in Figure 1 (a), where the three components are are trained together, guided by click-through data.

However, this traditional approach has shown limitations, particularly in terms of efficiency. A significant issue is the repetitive work done by the content encoder due to the repeated encoding of the same news articles, a consequence of their frequent interaction with various users. For instance, if there are N news articles, M users, and K instances of user-news interactions, with an average historical sequence of L interactions per user, the content encoder ends up processing $M \times L$ articles. The overall time required for training can be roughly calculated as $O(M \times L \times t_c + K \times (t_u + t_i))$, highlighting a particularly heavy load when $M \times L$ significantly exceeds either N or K, especially with advanced content encoders [19, 22] that have a much higher time complexity compared to other modules. For example, in the MIND dataset scenario, a single news article is encoded on average about 1818 times in one epoch of training, indicating a substantial amount of redundancy.

To address these inefficiencies, we draw inspiration from a new training approach known as the **only-encode-once paradigm**, as depicted in Figure 1 (b). This approach involves initially pretraining the content encoder using the news content in a self-supervised manner. After this phase, the content vectors are extracted and stored, allowing the history encoder and interaction module to be trained separately with much quicker access to the news representations via a lookup table. The pretraining phase for the content encoder has a time complexity of $O(N \times t_c)$, considered a fixed cost across various models. Meanwhile, the training of the recommendation system itself becomes significantly more efficient, with a time complexity of $O(K \times (t_u + t_i))$, marking a drastic reduction from the traditional method.

Usage	TorchRec	DeepCTR	DeepRec	RecBole	FuxiCTR	BARS	Ours
Data Preprocessing	✓	✓	✓	✓	✓	✓	✓
Content-aware CTR	×	×	×	–	–	–	✓
NewsRec Methods	×	×	×	×	×	×	✓
Decoupled Training	×	×	×	–	–	–	✓
End-to-end Training	×	×	×	×	×	×	✓
Green AI Evaluation	×	×	×	×	×	×	✓

Table 1: Comparison to exisiting recommendation benchmarks (✓ | – | × means totally | partially | not met, respectively). "Partially met" represents incomplete availability.

2.3 Adapting CTR Models to both Paradigms

While the prevailing news recommendation models adhere to a matching-based approach with simple dot product as the interaction module, preserving all positive samples and conducting negative sampling (including real or fake negatives) and trained as a classification task, there are still some studies [8, 22] that have validated the effectiveness of ranking-based CTR methods in news recommendation systems. In these approaches, all samples are retained throughout training, and the click probability is directly produced as a regression task. These CTR techniques incorporate intricate feature interaction networks, which can be utilized in the interaction module. When integrated into the content encoder, either through the end-to-end training paradigm or by initializing with pretrained news representations in the OLEO training paradigm, they gain the content-aware abilities.

2.4 Proposed GreenRec Benchmark

Here, we summarize 6 representative methods and 5 variants used in our GreenRec benchmark:

Base models. We select three representative news recommenders (i.e., NAML [16], LSTUR [1], and NRMS [18]) and three widely used CTR methods (i.e., BST [2], DCN [14], and DIN [25])

ID-based variants.[2] We remove the content encoder from the news recommenders and replace it with a ***randomly initialized*** embedding lookup table. CTR methods are inherently ID-based.

Text-based variants. We remove the news embedding lookup table of the CTR methods and replace it with a content sequence pooling layer. News recommenders are inherently text-based.

PLM-NR variants. As proposed by PLM-NR [19], we replace the content encoders of the text-based variants with pretrained language models such as BERT [3].

PREC variants. As proposed by PREC [8], we employ the OLEO training paradigm which first pretrains the content encoder and then replace the content encoders of the text-based variants with the ***informative*** news embedding lookup table.

BERT-OLEO variants. Unlike the PREC variants which requires the pretraining phase, here we directly use the pretrained BERT

[2] We use **green**, **yellow**, and **red** text to represent **OLEO-based or content-free (ID-based) training methods**, **end-to-end training variants with simple content encoders** and **end-to-end training variants with pretrained language models.**

	# News	# Users	# Interactions	# Samples	Density
MIND-small	65,238	94,057	347,727	8,381,093	0.1366%
MIND-large	104,151	750,434	3,958,501	95,447,571	0.1221%

Table 2: Dataset statistics. Density is defined as the ratio of # interactions to # all possible interactions.

model to extract and cache the news representations, and construct the ***informative*** lookup table.

3 IMPLEMENTATION AND EVALUATION

In this section, we first compare our GreenRec with other benchmarking studies. Then, we outline our evaluation procedures, including quality-based assessments and our proposed sustainability metric. Finally, we report and discuss our benchmarking results.

3.1 Comparison with Previous Benchmarks

In recent years, the issues of non-reproducibility and unfair comparisons have garnered increasing attention within the recommendation community. A series of studies [4, 11, 23, 24, 26, 27] have conducted comprehensive experiments and analyses, advocating for a unified benchmark construction. As shown in Table 1, we compare our news recommendation benchmark GreenRec with these existing efforts. It is evident that these studies primarily target in the general recommendation domain, particularly emphasizing aspects such as CTR prediction and collaborative filtering. Nevertheless, a standardized benchmarking pipeline tailored to the news recommendation domain and the environmental evaluation remains conspicuously absent. Hence, our GreenRec is the first attempt on benchmarking the news recommenders in the era of Green AI.

3.2 Dataset and Data Splitting

In this work, we mainly use two versions of the most prevalent news recommendation dataset: MIND-small and MIND-large [20]. Table 2 summarizes the data statistics information.

Due to the fact that the MIND-small dataset only provides training and validation sets, many studies have failed to specify their exact partitioning during training, and in some cases, they have directly reported results on the validation set, which leads to an unfair comparison among different models. Furthermore, the MIND-large dataset only offers online testing, which is not conducive to offline environment assessment. To address these issues, we have restructured both datasets by splitting the validation sets of both datasets in a 1:1 ratio to create new validation and testing sets. The scripts can be found in our released repository.

3.3 Evaluation Metrics

We employ three widely-used metrics for accuracy evaluation, including AUC, MRR, nDCG@5 [20]. Moreover, we also introduce one Green AI metrics: Carbon Emission [6] (CO_2E) and AUC per Carbon Emission (ApC), to evaluate the sustainability and energy utilization of model Ψ. CO_2E is computed by:

$$CO_2E = p \times t \times c, \tag{1}$$

where p is the power consumption differ by hardware types, t is the hardware running time, and c is carbon efficiency differ by power provider. ApC measures the unit conversion rate of carbon emissions on AUC, defined as:

$$ApC = \frac{AUC - 50}{CO_2E} \times 100. \tag{2}$$

Training Details. The Adam optimizer is employed for optimization. We consider various learning rates from the set $\{1e-5, 2e-5, 5e-5, 1e-4, 2e-4, 5e-4\}$ and batch sizes from the set $\{64, 128, 256, 500, 1000, 5000\}$. The embedding dimensions for all base models and their variants are fixed to 64. For the OLEO training paradigm which uses cached news vectors for initialization, we keep the pretrained embeddings fixed and use learnable transformation matrices for dimension projection. Experimental results are averaged over five runs, and all methods were trained using Nvidia GeForce RTX 3090 with 24GB memory, which has a power consumption of 350W. The carbon efficiency in our experimental region (anonymous for review) is 722g CO2-eq/kWh.

3.4 Findings

The comparative results of various models are summarized in Table 3, from which we derive the following observations.

ID- vs. Modality-based Recommender. Among the five variants, the ID-based variants exhibit the worst performance, showing the significance of news content comprehension in news recommenders. The findings in recent study [21] are in line with ours.

Enhancement of PLMs in the end-to-end training paradigm. The PLM-NR variants consistently outperforms the ID-based and Text-based ones, benefiting from the augmented understanding of news content offered by pretrained knowledge. However, they grapple with an efficiency issue, resulting in significantly higher carbon emissions compared to other variants.

Enhancement of PLMs in the OLEO training paradigm. The BERT variants maintain an environmentally sustainable level of efficiency. Nonetheless, there is a deficiency in the ability of general language models to fully grasp news content, resulting in a notable decline in accuracy compared to the PLM-NR variants. However, PREC variants achieve comparable performance to PLM-NR ones while significantly reducing carbon emissions, which represents the prime case of the tradeoff between accuracy and efficiency.

4 CONCLUSION

We have established the first standardized Green AI benchmark for news recommendation. Through extensive experiments over 30 base models and their variants, we conclude that the OLEO training paradigm successfully strikes a balance between accuracy and efficiency. We encourage other researchers to explore the OLEO paradigm and make contributions to environmental conservation.

ACKNOWLEDGEMENT

Qijiong Liu would like to thank Dr. Guanliang Chen and Mr. Haoxiang Shi for their insightful input during the brainstorming process of our paper. Additionally, the authors would like to thank the anonymous reviewers for their valuable feedback.

Dataset		MIND-small						MIND-large					
Method		Matching			CTR			Matching			CTR		
		NAML	LSTUR	NRMS	BST	DCN	DIN	NAML	LSTUR	NRMS	BST	DCN	DIN
	AUC	50.13	51.04	54.84	50.09	53.92	55.95	52.98	54.98	57.59	52.10	57.41	57.36
	MRR	23.01	22.90	26.53	22.13	25.18	25.88	24.52	25.99	27.41	24.81	26.76	26.70
ID-based	**N@5**	22.35	22.31	26.34	21.59	24.43	25.95	24.12	25.64	27.05	24.63	26.90	26.84
	CO$_2$E↓	**19**	**20**	**28**	**38**	**60**	**84**	**294**	**353**	**471**	**555**	**926**	**1294**
	ApC	0.68	5.20	17.29	0.24	6.53	7.08	1.01	1.41	1.61	0.38	0.80	0.57
	AUC	60.14	61.27	62.21	60.51	62.63	62.90	63.03	63.89	64.12	63.28	63.88	64.02
	MRR	28.93	29.64	30.19	28.59	29.73	30.06	30.40	31.24	31.77	30.73	31.65	31.98
Text-based (End-to-end)	**N@5**	29.33	30.28	31.10	29.09	30.52	30.65	31.82	32.15	32.64	31.95	32.40	33.00
	CO$_2$E↓	42	58	62	53	63	90	648	892	1010	1212	972	1386
	ApC	24.14	19.43	19.69	19.83	20.05	14.33	2.01	1.56	1.40	1.09	1.43	1.01
	AUC	62.06	**63.64**	62.53	**64.40**	63.32	**63.26**	**65.19**	**65.73**	**65.57**	**66.03**	65.42	**65.31**
	MRR	**31.66**	**31.74**	30.74	**32.21**	32.00	**31.83**	**32.74**	**33.18**	**32.94**	**33.40**	32.85	32.68
PLM-NR (End-to-end)	**N@5**	**32.25**	**32.72**	31.31	**33.34**	32.58	**32.40**	**33.77**	**34.26**	**34.13**	**34.70**	33.99	**33.70**
	CO$_2$E↓	178	202	252	505	1,752	1,839	2,527	3,032	4,043	8,086	27,036	28,329
	ApC	6.78	6.75	4.97	2.85	0.76	0.72	0.60	0.52	0.39	0.20	0.06	0.05
	AUC	60.62	61.09	60.94	60.81	62.65	62.40	63.02	63.62	63.40	62.94	64.29	63.75
	MRR	29.31	29.26	29.31	29.04	30.92	30.75	31.23	31.59	31.38	30.56	32.60	31.58
BERT (OLEO)	**N@5**	29.71	29.60	29.65	29.38	31.37	32.44	31.79	32.30	32.16	31.83	33.63	32.43
	CO$_2$E↓	22	23	33	38	62	86	353	404	505	640	956	956
	ApC	48.27	48.22	33.15	28.45	20.40	14.41	3.69	3.37	2.65	2.02	1.49	1.44
	AUC	**62.95**	62.16	**62.95**	62.43	**64.57**	63.12	64.78	64.88	64.34	65.33	**65.44**	64.53
	MRR	31.26	31.00	**31.18**	30.42	**32.60**	31.28	32.64	32.94	32.93	33.29	**33.04**	**32.72**
PREC (OLEO)	**N@5**	32.01	31.79	**32.10**	30.94	33.48	32.01	33.66	34.00	33.95	34.35	**34.03**	33.58
	CO$_2$E↓	22	23	33	38	62	86	353	404	505	640	956	956
	ApC	**58.86**	**52.87**	**39.24**	**32.71**	**23.50**	**15.25**	**4.19**	**3.68**	**2.84**	**2.40**	**1.61**	**1.52**
	ApC Imp. (%)	768%	683%	690%	1048%	2992%	2018%	598%	608%	628%	1100%	2583%	2940%

Table 3: Comparison of variants of different recommenders. "ApC Imp" represents the environmental sustainability growth rate of PREC variants compared to the SOTA PLM-NR variants. We bold the best results and underline the second-best results.

REFERENCES

[1] Mingxiao An, Fangzhao Wu, Chuhan Wu, Kun Zhang, Zheng Liu, and Xing Xie. 2019. Neural news recommendation with long-and short-term user representations. In *ACL*. 336–345.

[2] Qiwei Chen, Huan Zhao, Wei Li, Pipei Huang, and Wenwu Ou. 2019. Behavior sequence transformer for e-commerce recommendation in alibaba. In *DLP4Rec*.

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* (2018).

[4] Dmytro Ivchenko, Dennis Van Der Staay, Colin Taylor, Xing Liu, Will Feng, Rahul Kindi, Anirudh Sudarshan, and Shahin Sefati. 2022. Torchrec: a pytorch domain library for recommendation systems. In *RecSys*.

[5] Yoon Kim. 2014. Convolutional Neural Networks for Sentence Classification. In *EMNLP*. Association for Computational Linguistics, Doha, Qatar.

[6] Alexandre Lacoste, Alexandra Luccioni, Victor Schmidt, and Thomas Dandres. 2019. Quantifying the Carbon Emissions of Machine Learning. *arXiv* (2019).

[7] Jian Li, Jieming Zhu, Qiwei Bi, Guohao Cai, Lifeng Shang, Zhenhua Dong, Xin Jiang, and Qun Liu. 2022. MINER: Multi-Interest Matching Network for News Recommendation. In *ACL Findings 2022*. 343–352.

[8] Qijiong Liu, Jieming Zhu, Quanyu Dai, and Xiaoming Wu. 2022. Boosting Deep CTR Prediction with a Plug-and-Play Pre-trainer for News Recommendation. In *COLING*. International Committee on Computational Linguistics.

[9] Piotr Przybyła and Matthew Shardlow. 2022. Using NLP to quantify the environmental cost and diversity benefits of in-person NLP conferences. In *ACL Findings*.

[10] Roy Schwartz, Jesse Dodge, Noah A Smith, and Oren Etzioni. 2020. Green ai. *Commun. ACM* (2020).

[11] Weichen Shen. 2017. DeepCTR: Easy-to-use,Modular and Extendible package of deep-learning based CTR models.

[12] Giuseppe Spillo, Allegra De Filippo, Cataldo Musto, Michela Milano, and Giovanni Semeraro. 2023. Towards Sustainability-aware Recommender Systems: Analyzing the Trade-off Between Algorithms Performance and Carbon Footprint. In *RecSys*.

[13] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *arXiv* (2017).

[14] Ruoxi Wang, Bin Fu, Gang Fu, and Mingliang Wang. 2017. Deep & cross network for ad click predictions. In *ADKDD*.

[15] Steven Wilson, Walid Magdy, Barbara McGillivray, Kiran Garimella, and Gareth Tyson. 2020. Urban dictionary embeddings for slang NLP applications. In *LREC*.

[16] Chuhan Wu, Fangzhao Wu, Mingxiao An, Jianqiang Huang, Yongfeng Huang, and Xing Xie. 2019. Neural news recommendation with attentive multi-view learning. *IJCAI* (2019).

[17] Chuhan Wu, Fangzhao Wu, Mingxiao An, Jianqiang Huang, Yongfeng Huang, and Xing Xie. 2019. NPA: neural news recommendation with personalized attention. In *SIGKDD*.

[18] Chuhan Wu, Fangzhao Wu, Suyu Ge, Tao Qi, Yongfeng Huang, and Xing Xie. 2019. Neural news recommendation with multi-head self-attention. In *EMNLP-IJCNLP*.

[19] Chuhan Wu, Fangzhao Wu, Tao Qi, and Yongfeng Huang. 2021. Empowering News Recommendation with Pre-trained Language Models. *SIGIR* (2021).

[20] Fangzhao Wu, Ying Qiao, Jiun-Hung Chen, Chuhan Wu, Tao Qi, Jianxun Lian, Danyang Liu, Xing Xie, Jianfeng Gao, Winnie Wu, et al. 2020. Mind: A large-scale dataset for news recommendation. In *ACL*.

[21] Zheng Yuan, Fajie Yuan, Yu Song, Youhua Li, Junchen Fu, Fei Yang, Yunzhu Pan, and Yongxin Ni. 2023. Where to go next for recommender systems? id-vs. modality-based recommender models revisited. *arXiv* (2023).

[22] Qi Zhang, Jingjie Li, Qinglin Jia, Chuyuan Wang, Jieming Zhu, Zhaowei Wang, and Xiuqiang He. 2021. UNBERT: User-News Matching BERT for News Recommendation. In *IJCAI*.

[23] Shuai Zhang, Yi Tay, Lina Yao, Bin Wu, and Aixin Sun. 2019. Deeprec: An open-source toolkit for deep learning based recommendation. *arXiv* (2019).

[24] Wayne Xin Zhao, Yupeng Hou, Xingyu Pan, Chen Yang, Zeyu Zhang, Zihan Lin, Jingsen Zhang, Shuqing Bian, Jiakai Tang, Wenqi Sun, et al. 2022. RecBole 2.0: towards a more up-to-date recommendation library. In *CIKM*.

[25] 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 click-through rate prediction. In *SIGKDD*.

[26] Jieming Zhu, Quanyu Dai, Liangcai Su, Rong Ma, Jinyang Liu, Guohao Cai, Xi Xiao, and Rui Zhang. 2022. Bars: Towards open benchmarking for recommender systems. In *SIGIR*.

[27] Jieming Zhu, Jinyang Liu, Shuai Yang, Qi Zhang, and Xiuqiang He. 2021. Open benchmarking for click-through rate prediction. In *CIKM*.