TCCM: Time and Content-Aware Causal Model for Unbiased News Recommendation

Yewang Chen
Weiyao Ye
College of Science and Technology
Huaqiao University
Xiamen, China
ywchen@hqu.edu.cn
weiyao_ye@stu.hqu.edu.cn

Guipeng Xv
Chen Lin*
School of Informatics
Xiamen University
Xiamen, China
xuguipeng@stu.xmu.edu.cn
chenlin@xmu.edu.cn

Xiaomin Zhu
Strategic Assessments and Consultation Institute
Academy of Military Sciences
Beijing, China
xmzhu@nudt.edu.cn

ABSTRACT

Popularity bias significantly impacts news recommendation systems, as popular news articles receive more exposure and are often delivered to irrelevant users, resulting in unsatisfactory performance. Existing methods have not adequately addressed the issue of popularity bias in news recommendations, largely due to the neglect of the time factor and the impact of news content on popularity. In this paper, we propose a novel approach called **T**ime and **C**ontent-aware **C**ausal **M**odel, namely **TCCM** . It models the effects of three factors on user interaction behavior, i.e., the time factor, the news popularity, and the matching between news content and user interest. **TCCM** also estimates news popularity more accurately by incorporating the news content, i.e., the popularity of entity and words. Causal intervention techniques are applied to obtain debiased recommendations. Extensive experiments on well-known benchmark datasets demonstrate that the proposed approach outperforms a range of state-of-the-art techniques.

CCS CONCEPTS

• **Information systems** → **Recommender systems**.

KEYWORDS

News Recommendation; Popularity Bias; Causal Inference; Debiased Recommendation; Popularity Estimating

ACM Reference Format:
Yewang Chen, Weiyao Ye, Guipeng Xv, Chen Lin, and Xiaomin Zhu. 2023. TCCM: Time and Content-Aware Causal Model for Unbiased News Recommendation. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM '23), October 21–25, 2023, Birmingham, United Kingdom.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3583780.3615272

*Chen Lin and Weiyao Ye are the corresponding authors. Chen Lin is supported by the Natural Science Foundation of China (No.61972328). Yewang Chen is supported by Natural Science Foundation of Fujian Province, China (No.2021J01317)

CIKM '23, October 21–25, 2023, Birmingham, United Kingdom

ACM ISBN 979-8-4007-0124-5/23/10…$15.00
https://doi.org/10.1145/3583780.3615272

1 INTRODUCTION

Personalized news recommendation is a critical technique for helping users find news of interest and alleviating information overload on online news platforms [4]. Although news recommendation methods have been applied to many platforms, many existing news recommendation methods suffer from popularity bias, i.e., they prefer to recommend popular news rather than niche news that users might like [1, 29]. Popularity bias not only hampers the accuracy of personalized recommendations but also perpetuates the Matthew effect, i.e., popular news is becoming increasingly popular [10]. This can result in problematic phenomena such as echo chambers [5] and filter bubbles [20]. Therefore, addressing popularity bias in news recommendations is crucial.

Popularity bias is a common problem in recommendation systems. Existing debiased recommendation methods primarily target non-news recommendation scenarios, such as movie or product recommendations [3, 6, 8, 19, 24, 25]. Typically, these methods incorporate a popularity variable and model its influences on user interactions. Causal interventions [13] are often applied to mitigate the bias introduced by the popularity variable.

However, these methods are less effective in mitigating the popularity bias in news recommendations. (1) These methods do not consider the time factor of news, whereas the news is time-sensitive. Simply going for increased exposure to unpopular news articles, e.g., some news that are unpopular because they are outdated, may not be conducive to user experience. (2) These methods use non-content statistics to estimate popularity, e.g., the number of clicks an article receives, and the estimation does not reflect future news popularity accurately. For example, an emerging headline message can receive few clicks now but will be popular in the future. On the contrary, the news content (i.e., entities and words) is related to its future popularity. For example, hot entities such as celebrities and disasters will likely receive more public attention in the future.

To the best of our knowledge, only one previous work [14] has considered the issue of popularity bias in news recommendations. This work focuses on accurately modeling user interests by debiasing the interaction data, i.e., filtering out the news that is clicked on solely because it is popular. It fails to model the impact of popularity on the recommendation results, i.e., popular news articles are more likely to be recommended than non-popular ones.

Building on the above insights, we propose a **T**ime and **C**ontent-aware **C**ausal **M**odel (**TCCM**). Firstly, we construct a novel causal graph [12] to represent that the user interaction is affected by three factors, i.e., the news popularity, the news timeliness, and the matching between the user interest and the news content. Secondly, we design modules for learning news timeliness, news popularity, user interest, and user-news matching. Specifically, the content of the news, including the entities and words contained, is used to estimate its popularity. Finally, in the inference stage, we apply causal intervention [13] to derive debiased recommendation results.

In summary, our contributions are three-fold: (1) We consider the time factor in mitigating news popularity bias and propose a novel causal graph. (2) Based on the proposed causal graph, we propose a **T**ime and **C**ontent-aware **C**ausal **M**odel (**TCCM**) to mitigate news popularity bias, which consists of modules to more accurately estimate the news timeliness and news popularity. (3) Extensive experiments on well-known benchmark datasets demonstrate that **TCCM** outperforms a range of state-of-the-art techniques.

2 RELATED WORK

News Recommendation. Existing news recommendation methods mainly focus on ranking candidate news for a target user based on the match between news content and user interests [33]. Current works are mainly divided into: (1) Modeling of news content and user interests by various Deep Neural Networks (DNN), such as auto-encoder [11], GRU [2], CNN [23], Attention Networks [9], etc. (2) Exploring ways to match the user and news. Typical methods of user-news matching include user and news similarity [22], reinforcement learning-based approach to long-term total return optimisation [32], and so on. To our knowledge, there exists only one work [14] that considered the issue of popularity bias in news recommendations. However, this work focuses on making user modeling more accurate through debiased, and it adds news popularity to the recommendations without taking into account the impact of popularity on news exposure.

Popularity bias in recommendations. Popularity bias in non-news recommendations has recently been extensively researched [3, 6, 8, 19, 24, 25] and currently falls into three main categories: Inverse propensity weighting (IPW) [6, 19]. Causal embedding [3, 8]. and Counterfactual reasoning approaches [24, 25]. However, none of these previous methods have considered the timeliness and content of the news when mitigating popularity bias. They may recommend news that, although having low popularity, is already out of date.

Remarks. We differ from existing recommendation debiased methods in that we consider the timeliness of the news and avoid recommending outdated news when increasing the exposure of low-popularity news. At the same time, we use the content of the news to estimate its popularity, rather than simply counting the number of clicks on the whole news. In comparison to PP-Rec, we mitigate the effect of popularity on news exposure when making recommendations, whereas PP-Rec uses news popularity and does not take into account the bias it introduces.

3 METHODOLOGY

3.1 Causal View of News Recommendation

We first analyze the news recommendation from the view of causality, by using a causal graph which is a directed acyclic graph that describes the causal relationships between variables [12]. As shown in Fig. 1(a), we assume there are seven variables involved in the news recommendation process, i.e., the node N denotes a news

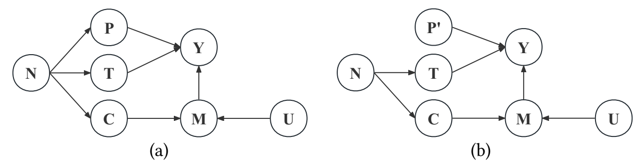

Figure 1: (a) The causal graph of the proposed TCCM . (b) After intervention on the proposed causal graph for TCCM .

article, the node U denotes a user, the node P denotes the news popularity, the node T denotes the news timeliness, the node C denotes the news content, the node M denotes user-content matching scores, the node Y denotes the interaction probability.

Since a news article has three attributes, i.e., news popularity, news timeliness, and news content, we draw the edges $N \rightarrow P$, $N \rightarrow T$, and $N \rightarrow C$. The user prefers to click on the news that matches his or her preference, i.e., $C \rightarrow M$, $U \rightarrow M$. In addition, whether a user clicks on a news article or not is influenced by its popularity and timeliness. Popular news is more likely to be exposed to users, and outdated news does not get much attention from users. So $T \rightarrow Y$, $P \rightarrow Y$, and $M \rightarrow Y$.

3.2 Learning

The variables and their relationships are learned in the recommendation systems. Suppose the historical interaction data is denoted as D, which is collected in a time sequence. Let $U = \{u_1, ..., u_{|U|}\}$ denotes all users and $N = \{n_1, ..., n_{|N|}\}$ denotes all news. Each news is associated with a set of entities and words as $\{e_1, ..., e_q\}$ and $\{w_1, ..., w_l\}$, where the entity space is denoted by $E = \{e_1, ..., e_{|E|}\}$, and the word space is denoted by $W = \{w_1, ..., w_{|W|}\}$.

As shown in the framework of **TCCM** (Fig. 2), we designed the time module, popularity module, and user-content matching module to identify and learn from these confounding influences.

Time module. For a given news n, the news timeliness is denoted as t_n. It is defined as the duration between publish time and prediction time, which is quantified in hours. We employ a time embedding layer to convert t_n into an embedding vector $\mathbf{t_n}$. The embedding layer can map the input to a vector space, which helps the model to better capture useful information and features.

After that, we apply a dense network (a feed-forward neural network variant) to $\mathbf{t_n}$ to predict the time score $t_n^{'}$. Dense networks can further extract and learn feature representations and perform feature fusion.

$$t_n^{'} = Dense(\mathbf{t_n}). \tag{1}$$

Given that news articles published earlier have a lower probability of user interaction, we compute the reciprocal value of $t_n^{'}$ and control its intensity using the parameter λ. The process is defined as follows:

$$st_n = (1/t_n^{'})^{\lambda}. \tag{2}$$

Popularity module. We believe that the popularity of news is related to its content (i.e., entities and words). Thus, given a news n, we use user interactions in recent m hours to calculate near real-time click-through rate (CTR) for each entity and word, denoted as $\{ze_1, ..., ze_q\}$ and $\{zw_1, ..., zw_l\}$. Specifically, CTR can be calculated as the ratio of the number of clicks on a particular entity or word to the number of displays it receives within a given time period of m hours. In the same way, we map the popularity of each entity

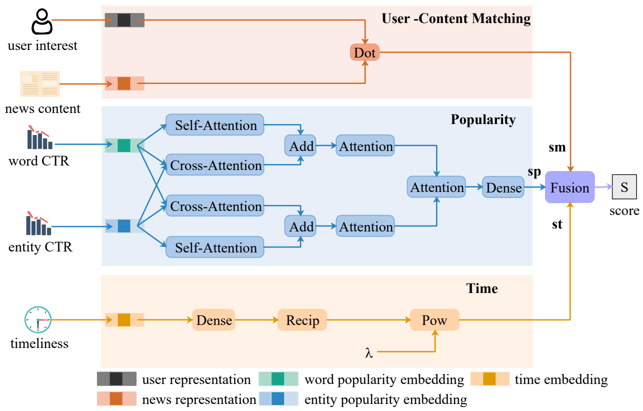

Figure 2: The overall framework of TCCM .

and word into vectors space by popularity embedding layer $\mathbf{ZE} = \{\mathbf{ze}_1, ..., \mathbf{ze}_q\}$ and $\mathbf{ZW} = \{\mathbf{zw}_1, ..., \mathbf{zw}_l\}$.

Then we utilize a multi-head self-attention network [21] (MHSA) to learn the relationship between the various entities, and a multi-head cross-attention network [14] (MHCA) to learn the relationship between individual entities and textual contexts. We formulate the popularity representation of each entity in the news as the sum of the representations learned by the MHSA and MHCA networks. Lastly, an attention network is used to aggregate the popularity representations of each entity, as follows.

$$MultiHead(Q, K, V) = Concat(head_1, ..., head_h)W^O. \quad (3)$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V). \quad (4)$$

$$z_1 = MultiHead(\mathbf{ZE}, \mathbf{ZE}, \mathbf{ZE}), z_2 = MultiHead(\mathbf{ZE}, \mathbf{ZW}, \mathbf{ZW}) \quad (5)$$

$$p_e = Attention(Add(z_1, z_2)). \quad (6)$$

Where W_i^Q, W_i^K, W_i^V, and W^O are parameter matrix.

We perform a similar operation with words, generating a word popularity representation p_w.

Finally, we obtain a unified news popularity representation by aggregating entity and word popularity representations through an attention network and turn it into a popularity score sp_n, using the dense network.

$$sp_n = Dense(Attention([p_e, p_w])). \quad (7)$$

User-content matching module. The user-content matching module calculates a match score between the news content (i.e., entities and words) and the user's preferences. In our approach, we utilize the inner product of the user embedding and the news embedding to calculate this match score.

$$sm_n = x_u \cdot x_n^T. \quad (8)$$

where x_u and x_n denote the embedding of the user and the news, we use the model from the literature [14] to learn them.

Fusion of effects. To integrate the effects of the time factor, popularity, and user interest, the outputs of the user interest module, the time module, and the popularity module were fused into a multiplicative fusion calculator, as follows.

$$s_n = (1 - \alpha)sm_n + \alpha(sp_n \cdot st_n). \quad (9)$$

where α is the trainable parameter that we use to learn the user's propensity for timeliness and popularity of news.

3.3 Training

We trained the model with BPR loss [18]. The BPR loss maximizes the difference between positive and negative samples to allow the model to better learn the features of the positive samples. Formally, the BPR loss is written as:

$$L_{bpr} = -\frac{1}{|D|}\sum_{n=1}^{|D|} \log(\sigma(s_n^i - s_n^j)). \quad (10)$$

where s_n^i and s_n^j denote the scores of the n-th news positive and negative samples respectively, D denotes the training set, and $\sigma(\cdot)$ is the sigmoid activation function. The negative samples are sampled randomly from all the news that the user has not interacted with.

3.4 Causal Intervention

Causal interventions [12, 13] are used to manipulate the potential outcome of a variable by setting it to a specific value. In the proposed **TCCM** , we intervened in the popularity module to reduce popularity bias. By setting the effect of popularity to a low level, we were able to obtain debiased recommendation results while taking into account the time factor. Let $S(y_n)$ be a function of the interaction score calculated with news n, from Eq. (7) we have:

$$s_n = S(y_n) = (1 - \alpha)sm_n + \alpha(sp_n \cdot st_n). \quad (11)$$

Intervening on popularity will cut off the relationship between news and its popularity, as shown in Fig. 1(b). After the intervention, the score can be calculated as follows:

$$S(y_n|do(P)) = S(y_n|P') = (1 - \alpha)sm_n + \alpha(sp_n' \cdot st_n). \quad (12)$$

4 EXPERIMENTS

In this section, we conduct experiments to answer the following research questions. **RQ1**: Does **TCCM** mitigate the popularity bias in news recommendations and outperform other methods? **RQ2**: How do the time module and the popularity module affect the performance?

4.1 Experimental Setup

Dataset. Our experiments are conducted on a real-world open dataset MIND[1], which is constructed from user click logs of Microsoft News[2]. The publisher randomly sampled one million users with at least five news clicks recorded over six weeks from October 12 to November 22, 2019, and generated 1 million impression logs.

Implementation. In our experiments, the time, entity popularity and word popularity embeddings are randomly initialized 200-dimensional vectors. News representations and user representations are 400-dimensional vectors obtained using the model in literature [14]. All multi-headed attention networks were set to 20 attention heads with an output dimension of 20 for each attention head. More details are provided in the source codes[3].

Metrics. To evaluate the accuracy of the recommendation, we used three metrics, i.e., Area Under Curve (*AUC*), Mean Reciprocal Rank (*MRR*), and Normalized Discounted Cumulative Gain (*NDCG*@5 & *NDCG*@10), which are commonly used evaluation metrics in [7, 17, 27, 30]. Higher values of *AUC*, *MRR*, *NDCG*@5, and *NDCG*@10 indicate more accurate recommendation results. In addition, we

[1]https://msnews.github.io
[2]https://microsoftnews.msn.com
[3]https://github.com/XFastDataLab/TCCM

Table 1: Performance of different methods

	AUC	MRR	NDCG@5	NDCG@10	PRU
NPA	0.6713	0.3290	0.3575	0.4145	0.2922
NAML	0.6722	0.3301	0.3583	0.4154	0.2298
NRMS	0.6808	0.3343	0.3618	0.4207	0.2734
LSTUR	0.6835	0.3348	0.3639	0.4216	0.2176
FUM	0.7001	0.3445	0.3752	0.4318	0.2619
CAUM	0.7004	0.3461	0.3780	0.4348	0.2728
MACR	0.6895	0.3358	0.3672	0.4265	0.2225
PP-Rec	0.7092	0.3914	0.4377	0.4996	0.1506
TCCM	**0.7220**	**0.4193**	**0.4682**	**0.5361**	**0.1128**

used *PRU* [34] to measure whether the recommendations are biased towards popular news, which is used in [31, 34]. Lower values of *PRU* values indicate less biased recommendations.

Competitors. We compared **TCCM** with two groups of methods:

Personalized news recommendation methods consist of (1-3) NPA [27], NAML [26], NRMS [28]: learn the news and user representations by different personalized attention networks; (4) LSTUR [2]: the short-term and long-term interests of users are modeled by GRU and user ID respectively; (5) FUM [15]: connects clicked news into a long document and transforms user modeling into a document modeling task; (6) CAUM [16]: incorporates candidate news into user modeling to better match candidate news and user interests.

Popularity de-biasing methods consist of: (1) MACR [25]: uses counterfactual reasoning methods to address popularity bias for the recommender system. *MACR is not specifically designed for news recommendations*; (2) PP-Rec [14]: combines news popularity to overcome cold starts and diversity issues and uses popularity-aware user encoders to remove popularity bias from user behavior. For all these competitors, we used the open-source implementation provided in the authors' paper and tuned it to the optimal result.

4.2 Comparative Performance

From Table 1, we observe the following. Comparison with personalized news recommendation methods: (1) **TCCM** consistently outperforms all competing models across all evaluation metrics, achieving the highest *MRR*, *NDCG*@5, *NDCG*@10, and the lowest *PRU*. This underscores the ability of **TCCM** to reduce popularity bias while ensuring accurate news recommendations. (2) When compared to personalized recommendation models, news popularity debiased models like **TCCM** and PP-Rec show a significant performance improvement on *MRR*, *NDCG*@5, *NDCG*@10, and *PRU*. These results highlight the crucial role of mitigating popularity bias in news recommendations.

Comparison with popularity debiased methods: Compared to MARC and PP-Rec, **TCCM** exhibits superior performance. This advantage can be attributed to **TCCM** 's consideration of both time factor and the impact of news content on popularity when alleviating popularity bias. Moreover, non-news recommendation models such as MARC, which rely on ID embedding to model news and users, may not be as effective, given that it does not incorporate content modeling. And PP-Rec fails to consider popularity may have an impact on the recommendation results.

4.3 Ablation Study

Assessing the Overall Impact of the Time Module and Popularity Module. To evaluate the collective contribution of the time

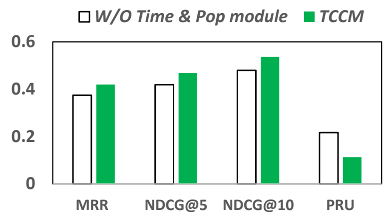

Figure 3: Performance of TCCM with and without time and popularity module.

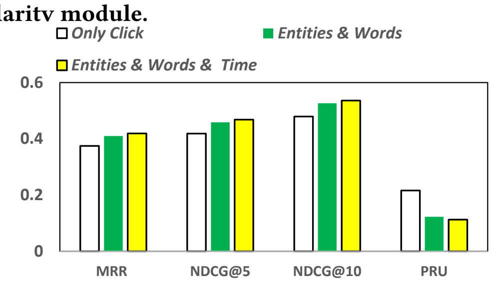

Figure 4: Performance of different implementations.

module and the popularity module, we conducted an ablation study. In this study, we removed both the disentangled operations and the content-based popularity estimation from **TCCM** , while keeping the other parts unchanged. Subsequent evaluation of these models, as presented in Fig. 3, showed an increase in recommendation accuracy and a decrease in popularity bias. These results suggest that the inclusion of time factors and estimating popularity based on news content effectively reduce popularity bias, leading to more meaningful recommendations.

Investigating the Individual Impacts of the Time Module and Popularity Module. To assess the separate contributions of the time module and the popularity module, we modified the implementation strategy for **TCCM** , while keeping all other components constant. The revised strategies are as follows: (1) *Only Click*: Here, news popularity is estimated based solely on the number of clicks on the news. (2) *Entities & Words*: This strategy combines the popularity of entities and words to estimate news popularity. (3) *Entities & Words & Time*: This approach takes into account time factors along with content-based popularity estimates.

From the results depicted in Fig. 4, we make two key observations: (1) The performance of the *Only Click* method is unsatisfactory, suggesting that estimating news popularity based solely on clicks, without considering content, is insufficient. (2) The performance of the *Entities & Words* method is outdone by the *Entities & Words & Time* approach, demonstrating that incorporating the time factor significantly improves the performance of **TCCM** in news recommendation.

5 CONCLUSION

In this paper, we propose **TCCM** , a novel method that tackles popularity bias in news recommendations, which considers the time factor and the impact of news content on popularity. It learns the effects of the time factor, popularity, and user-content matching. Through causal intervention, we generate debiased recommendation results. Experimental results demonstrate the superiority of our approach, highlighting its effectiveness in mitigating popularity bias and providing more accurate news recommendations.

REFERENCES

[1] Himan Abdollahpouri. 2020. Popularity Bias in Recommendation: A Multi-stakeholder Perspective. *ArXiv* abs/2008.08551 (August 2020).

[2] 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 *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Florence, Italy, 336–345.

[3] Stephen Bonner and Flavian Vasile. 2018. Causal Embeddings for Recommendation. In *Proceedings of the 12th ACM Conference on Recommender Systems*. Association for Computing Machinery, New York, NY, USA, 104–112.

[4] Suyu Ge, Chuhan Wu, Fangzhao Wu, Tao Qi, and Yongfeng Huang. 2020. Graph enhanced representation learning for news recommendation. In *Proceedings of The Web Conference 2020*. Association for Computing Machinery, New York, NY, USA, 2863—-2869.

[5] Yingqiang Ge, Shuya Zhao, Honglu Zhou, Changhua Pei, Fei Sun, Wenwu Ou, and Yongfeng Zhang. 2020. Understanding Echo Chambers in E-Commerce Recommender Systems. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. Association for Computing Machinery, New York, NY, USA, 2261–2270.

[6] Alois Gruson, Praveen Chandar, Christophe Charbuillet, James McInerney, Samantha Hansen, Damien Tardieu, and Ben Carterette. 2019. Offline Evaluation to Make Decisions About PlaylistRecommendation Algorithms. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*. Association for Computing Machinery, New York, NY, USA, 420–428.

[7] Jianxun Lian, Fuzheng Zhang, Xing Xie, and Guangzhong Sun. 2018. Towards Better Representation Learning for Personalized News Recommendation: a Multi-Channel Deep Fusion Approach. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18l*. International Joint Conferences on Artificial Intelligence Organization, Stockholm, Sweden, 3805–3811.

[8] Dugang Liu, Pengxiang Cheng, Zhenhua Dong, Xiuqiang He, Weike Pan, and Zhong Ming. 2020. A General Knowledge Distillation Framework for Counterfactual Recommendation via Uniform Data. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. Association for Computing Machinery, New York, NY, USA, 831–840.

[9] Danyang Liu, Jianxun Lian, Shiyin Wang, Ying Qiao, Jiun-Hung Chen, Guangzhong Sun, and Xing Xie. 2020. KRED: Knowledge-Aware Document Representation for News Recommendations. In *Proceedings of the 14th ACM Conference on Recommender Systems*. Association for Computing Machinery, New York, NY, USA, 200—-209.

[10] Yudan Liu, Kaikai Ge, Xu Zhang, and Leyu Lin. 2019. Real-time Attention Based Look-alike Model for Recommender System. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*. Association for Computing Machinery, Anchorage, AK, USA, 2765–2773.

[11] Shumpei Okura, Yukihiro Tagami, Shingo Ono, and Akira Tajima. 2017. Embedding-Based News Recommendation for Millions of Users. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery, New York, NY, USA, 1933—-1942.

[12] Judea Pearl. 2009. *Causality*. Cambridge university press, New York.

[13] Judea Pearl and Dana Mackenzie. 2018. The Book of Why: The New Science of Cause and Effect. *Science* 361, 6405 (2018), 855–855.

[14] Tao Qi, Fangzhao Wu, Chuhan Wu, and Yongfeng Huang. 2021. PP-Rec: News Recommendation with Personalized User Interest and Time-aware News Popularity. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*. Association for Computational Linguistics, Online, 5457–5467.

[15] Tao Qi, Fangzhao Wu, Chuhan Wu, and Yongfeng Huang. 2022. FUM: Fine-Grained and Fast User Modeling for News Recommendation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. Association for Computing Machinery, New York, NY, USA, 1974—-1978.

[16] Tao Qi, Fangzhao Wu, Chuhan Wu, and Yongfeng Huang. 2022. News Recommendation with Candidate-Aware User Modeling. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Managementl*. Association for Computing Machinery, New York, NY, USA, 1917––1921.

[17] Tao Qi, Fangzhao Wu, Chuhan Wu, Peijie Sun, Le Wu, Xiting Wang, Yongfeng Huang, and Xing Xie. 2022. ProFairRec: Provider Fairness-Aware News Recommendation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. Association for Computing Machinery, New York, NY, USA, 1164—-1173.

[18] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian Personalized Ranking from Implicit Feedback. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*. AUAI Press, Arlington, Virginia, USA, 452–461.

[19] 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 - Volume 48*. JMLR.org, New York, NY, USA, 1670–1679.

[20] Harald Steck. 2018. Calibrated Recommendations. In *Proceedings of the 12th ACM Conference on Recommender Systems*. Association for Computing Machinery, New York, NY, USA, 154–162.

[21] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia. Polosukhin. 2017. Attention is All You Need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Curran Associates Inc., Red Hook, NY, USA, 6000–6010.

[22] Heyuan Wang, Fangzhao Wu, Zheng Liu, and Xing Xie. 2020. Fine-grained Interest Matching for Neural News Recommendation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Online, 836—-845.

[23] Hongwei Wang, Fuzheng Zhang, Xing Xie, and Minyi Guo. 2018. DKN: Deep Knowledge-Aware Network for News Recommendation. In *Proceedings of the 2018 World Wide Web Conferencel*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 1835—-1844.

[24] Wenjie Wang, Fuli Feng, Xiangnan He, Hanwang Zhang, and Tat-Seng Chua. 2021. Clicks Can Be Cheating: Counterfactual Recommendation for Mitigating Clickbait Issue. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. Association for Computing Machinery, New York, NY, USA, 1288–1297.

[25] Tianxin Wei, Fuli Feng, Jiawei Chen, Ziwei Wu, Jinfeng Yi, and Xiangnan He. 2021. Model-Agnostic Counterfactual Reasoning for Eliminating Popularity Bias in Recommender System. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. Association for Computing Machinery, New York, NY, USA, 1791—-1800.

[26] Chuhan Wu, Fangzhao Wu, Mingxiao An, Jianqiang Huang, Yongfeng Huang, and Xing Xie. 2019. Neural News Recommendation with Attentive Multi-View Learning. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*. AAAI Press, Macao, China, 3863—-3869.

[27] Chuhan Wu, Fangzhao Wu, Mingxiao An, Jianqiang Huang, Yongfeng Huang, and Xing Xie. 2019. NPA: Neural News Recommendation with Personalized Attention. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. Association for Computing Machinery, New York, NY, USA, 2576—-2584.

[28] Chuhan Wu, Fangzhao Wu, Suyu Ge, Tao Qi, Yongfeng Huang, and Xing Xie. 2019. Neural News Recommendation with Multi-Head Self-Attention. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics, Hong Kong, China, 6389–6394.

[29] Chuhan Wu, Fangzhao Wu, Yongfeng Huang, and Xing Xie. 2021. Personalized news recommendation: A survey. *ArXiv* abs/2106.08934 (2021).

[30] Chuhan Wu, Fangzhao Wu, Tao Qi, and Yongfeng Huang. 2022. End-to-end Learnable Diversity-aware News Recommendation. *ArXiv* abs/2204.00539 (April 2022).

[31] Guipeng Xv, Chen Lin, Hui Li, Jinsong Su, Weiyao Ye, and Yewang Chen. 2022. Neutralizing Popularity Bias in Recommendation Models. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. Association for Computing Machinery, New York, NY, USA, 2623–2628.

[32] Guanjie Zheng, Fuzheng Zhang, Zihan Zheng, Yang Xiang, Nicholas Yuan, Xing Xie, and Zhenhui Li. 2018. DRN: A Deep Reinforcement Learning Framework for News Recommendation. In *Proceedings of the 2018 World Wide Web Conference*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 167–176.

[33] Qiannan Zhu, Xiaofei Zhou, Zeliang Song, Jianlong Tan, and Li Guo. 2019. DAN: Deep Attention Neural Network for News Recommendation. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence*. AAAI Press, Honolulu, Hawaii, USA, 5973–5980.

[34] Ziwei Zhu, Yun He, Xing Zhao, Yin Zhang, Jianling Wang, and James Caverlee. 2021. Popularity-Opportunity Bias in Collaborative Filtering. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*. Association for Computing Machinery, New York, NY, USA, 85––93.