News Recommendation with Candidate-aware User Modeling

Tao Qi
Department of Electronic Engineering,
Tsinghua University
taoqi.qt@gmail.com

Fangzhao Wu*
Microsoft Research Asia
wufangzhao@gmail.com

Chuhan Wu
Department of Electronic Engineering,
Tsinghua University
wuchuhan15@gmail.com

Yongfeng Huang
Department of Electronic Engineering,
Tsinghua University
yfhuang@tsinghua.edu.cn

ABSTRACT

News recommendation aims to match news with personalized user interest. Existing methods for news recommendation usually model user interest from historical clicked news without the consideration of candidate news. However, each user usually has multiple interests, and it is difficult for these methods to accurately match a candidate news with a specific user interest. In this paper, we present a candidate-aware user modeling method for personalized news recommendation, which can incorporate candidate news into user modeling for better matching between candidate news and user interest. We propose a candidate-aware self-attention network that uses candidate news as clue to model candidate-aware global user interest. In addition, we propose a candidate-aware CNN network to incorporate candidate news into local behavior context modeling and learn candidate-aware short-term user interest. Besides, we use a candidate-aware attention network to aggregate previously clicked news weighted by their relevance with candidate news to build candidate-aware user representation. Experiments on real-world datasets show the effectiveness of our method in improving news recommendation performance.

CCS CONCEPTS

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

KEYWORDS

News Recommendation, Candidate-aware User Model, Single-Tower
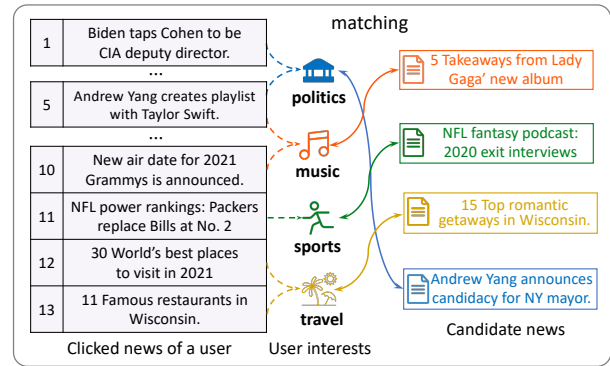
*The corresponding author.

Figure 1: The matching between candidate news and user interest inferred from historical clicked news.

1 INTRODUCTION

Personalized news recommendation is a critical technique for online news platforms to improve user experience [5, 11, 14, 26, 27, 32, 33]. Accurate modeling of user interest on candidate news is important for personalized news recommendation [2, 4, 7, 13, 29, 30]. Many existing methods first model user interests and candidate news content separately and then use their representations for interest matching [23]. For instance, An et al. [1] used a GRU network and ID embeddings to learn user interest representations from clicked news. Wu et al. [20] applied an attention network to learn user interest representations by aggregating user's clicked news. Both of them modeled the relevance between user interests and candidate news based on the dot product of their representations. In these methods, user interests are modeled in a candidate-agnostic way. However, each user usually has multiple interests [15], and it may be difficult to accurately match candidate news with a specific user interest if candidate news is not considered in user modeling [19].

Our work is motivated by the following observations. First, users usually have multiple interests. For instance, as shown in Fig. 1, we can infer that the example user is interested in many different fields, such as politics, music, sports, and travel, from her clicked news. However, a candidate news usually only matches a small part of user interests. For instance, the 4th candidate news only matches user interests in politics, and it has low relevance to other interests of this user like music and sports. Thus, it may be difficult to accurately match the candidate news if candidate news information is not considered in user modeling. Second, local contexts of users' news

click behaviors are useful for inferring short-term user interests. For example, as shown in Fig. 1, we can infer the user's recent interests on travel in Wisconsin from the relatedness between the 12th and 13th news clicks. Third, long-range relatedness between users' historical clicks also provides rich information to model long-term user interests. For example, we can infer the long-term user interests in music from the long-range relatedness between the 5th and 10th clicks. Thus, understanding both short- and long-term user interests is important for accurate news recommendation [1].

In this paper, we propose a candidate-aware user modeling framework for personalized news recommendation (*CAUM*), which can incorporate candidate news information into user modeling for accurate interest matching. We propose a candidate-aware self-attention network to learn candidate-aware global user interest representations. It uses candidate news representation to guide the modeling of global relatedness between historical clicked news. In addition, we propose a candidate-aware CNN network to learn candidate-aware short-term user interest representations. It incorporates candidate news information into the modeling of local contexts of click behaviors. Besides, we adopt a candidate-aware attention network to weight clicked news based on their relevance with candidate news to learn candidate-aware user interest representation for better matching with candidate news. Experimental results on two real-world datasets verify that *CAUM* can improve the performance of user modeling for news recommendation.

2 METHODOLOGY

2.1 Candidate-aware User Modeling

Next, we introduce the candidate-aware user interest modeling framework, which can exploit candidate news to guide user interests modeling. It takes representations $\{\mathbf{c}_i\}_{i=1}^N$ of user's recent N clicks $\{c_i\}_{i=1}^N$ and representation $\mathbf{n}_c$ of candidate news n_c as inputs. (Refer to Section 2.2.). Fig. 2 shows it contains three modules.

Candi-SelfAtt: Long-range contexts of news clicks are usually informative for inferring global user interests. Besides, different long-range behavior contexts usually have different importance to capture different global user interests. For example, Fig. 1 shows the relatedness between the 1st and 5th click can help infer user interests in politics while the relatedness between the 5th click and 10th click can help infer user interests in music. Thus, modeling long-range behavior contexts with candidate news information may better model global user interests to match candidate news. Motivated by these observations, we propose a candidate-aware self-attention network (*Candi-SelfAtt*), which can use candidate news information to guide global behavior contexts modeling. The core of *Candi-SelfAtt* is to adjust attention weights of behavior contexts via candidate news to select important ones. We first apply multiple self-attention heads [17] to model click relatedness:

$$\hat{r}_{i,j}^k = \mathbf{q}_i^T \mathbf{W}_k^r \mathbf{c}_j, \quad \mathbf{q}_i^T = \mathbf{Q}_u \mathbf{c}_i, \tag{1}$$

where $\hat{r}_{i,j}^k$ denotes the attention score generated by the k-th attention head to model relatedness between the i-th and j-th click, $\mathbf{Q}_u$ is the projection matrix, and $\mathbf{W}_k^r$ is parameters of the k-th attention head. Note that $\{\hat{r}_{i,j}^k\}_{j=1}^N$ model the relatedness between the i-th
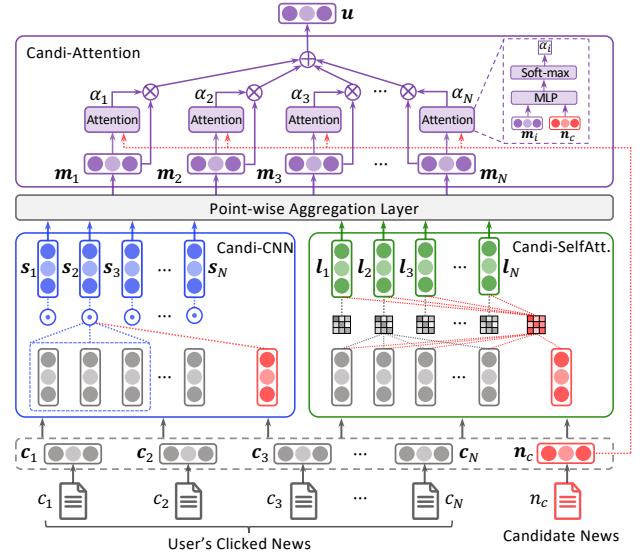

Figure 2: Framework of our *CAUM* method.

clicks and other user's clicks. We further adaptively select informative long-range relatedness for modeling user interest in candidate n_c based on their relevance with candidate news:

$$r_{i,j}^k = \hat{r}_{i,j}^k + \mathbf{q}_c^T \mathbf{W}_k^r \mathbf{c}_j, \quad \mathbf{q}_c^T = \mathbf{Q}_c \mathbf{n}_c, \tag{2}$$

where $r_{i,j}^k$ is the candidate-aware attention score, and $\mathbf{Q}_c$ is a projection matrix. Then we learn the representation $\mathbf{l}_i^k$ generated by the k-th head for the i-th click based on attention weights $\{\gamma_j^k\}_{j=1}^N$:

$$\mathbf{l}_i^k = \mathbf{W}_o^k \sum_{j=1}^N \gamma_j^k \mathbf{c}_j, \quad \gamma_j^k = \frac{\exp(r_{i,j}^k)}{\sum_{p=1}^N \exp(r_{i,p}^k)}, \tag{3}$$

where $\mathbf{W}_o^k$ is the projection matrix of the k-th attention head. Finally, we learn the global contextual representation $\mathbf{l}_i$ for i-th click by contacting $\{\mathbf{l}_i^k\}_{k=1}^K$, where K is the number of attention heads.

Candi-CNN: Besides global user interests, short-term user interests are also important for matching candidate news [1, 3]. Short-term user interests can usually be effectively modeled from local contexts between adjacent user behaviors [1]. In addition, incorporating candidate news information into local behavior contexts modeling also has the potential to better model short-term interest in candidate news. Thus, we propose a candidate-aware CNN network, which can capture local contexts between adjacent clicks with candidate news information. We apply multiple filters to capture the potential patterns between local contexts of adjacent clicks and candidate news: $\mathbf{s}_i = \mathbf{W}_c [\mathbf{c}_{i-h}; ...; \mathbf{c}_i; ...; \mathbf{c}_{i+h}; \mathbf{n}_c]$, where $\mathbf{s}_i$ represents local contextual representation of the i-th click, $2h+1$ is the window size of the CNN network, and $\mathbf{W}_c$ represents parameters of filters in the *Candi-CNN* network. Similarly, we can learn local contextual representations $[\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_N]$ of all clicked news. These local contextual representations of clicked news encode candidate-aware short-term user interests. Then, we learn unified contextual representation $\mathbf{m}_i$ for the i-th click based on the aggregation of $\mathbf{l}_i$ and $\mathbf{s}_i$: $\mathbf{m}_i = \mathbf{P}_m [\mathbf{s}_i; \mathbf{l}_i]$, where $\mathbf{P}_m$ is the projection matrix.

Table 1: Performance comparisons. The improvement of *CAUM* over baseline methods is significant at level $p \le 0.001$.

	MIND				NewsApp			
	AUC	MRR	nDCG@5	nDCG@10	AUC	MRR	nDCG@5	nDCG@10
GRU [10]	65.69±0.15	31.47±0.06	33.96±0.07	39.70±0.07	63.23±0.37	27.83±0.26	31.84±0.31	37.41±0.34
NAML [20]	66.49±0.19	32.38±0.13	35.17±0.15	40.84±0.14	64.52±0.35	29.02±0.20	33.35±0.30	38.90±0.33
NPA [21]	66.56±0.18	32.42±0.10	35.20±0.11	40.87±0.13	64.39±0.14	28.93±0.10	33.31±0.11	38.83±0.11
NRMS [24]	68.04±0.20	33.31±0.07	36.23±0.15	41.92±0.12	65.36±0.28	29.47±0.21	33.96±0.27	39.49±0.19
LSTUR [1]	68.36±0.22	33.30±0.11	36.30±0.16	42.00±0.14	65.18±0.23	29.28±0.21	33.71±0.23	39.28±0.22
KRED [8]	67.73±0.13	32.87±0.11	35.81±0.13	41.43±0.15	65.45±0.14	29.56±0.09	34.11±0.11	39.65±0.12
DKN [19]	66.32±0.18	32.13±0.14	34.86±0.13	40.47±0.18	62.86±0.37	28.00±0.23	32.12±0.29	37.68±0.28
HiFi-Ark [9]	67.93±0.25	32.87±0.07	35.77±0.08	41.47±0.10	64.91±0.15	29.10±0.12	33.52±0.18	38.98±0.14
FIM [18]	67.84±0.12	33.26±0.06	36.18±0.10	41.86±0.11	65.39±0.10	29.63±0.11	34.14±0.12	39.60±0.10
GNewsRec [3]	68.36±0.22	33.41±0.10	36.36±0.13	42.01±0.14	65.31±0.22	29.40±0.14	33.92±0.16	39.48±0.16
CAUM	**70.04**±0.08	**34.71**±0.08	**37.89**±0.07	**43.57**±0.07	**66.44**±0.07	**30.07**±0.10	**34.69**±0.12	**40.23**±0.10

Candi-Att: Since the importance of clicked news for modeling user interest in candidate news may be different, we apply a candidate-aware attention network to model the importance of clicked news from their relevance with candidate news n_c and further build the candidate-aware user interest representation $\mathbf{u}$:

$$\mathbf{u} = \sum_{i=1}^{N} \alpha_i \mathbf{m}_i, \ \alpha_i = \frac{\exp(\Phi(\mathbf{m}_i, \mathbf{n}_c))}{\sum_{j=1}^{N} \exp(\Phi(\mathbf{m}_j, \mathbf{n}_c))}, \tag{4}$$

where α_i is the weight of the i-th click and Φ is an MLP network. In this way, user interests relevant to the candidate news can be effectively encoded into $\mathbf{u}$ to improve the accuracy of interest matching.

2.2 News Modeling

In *CAUM* we model news based on previous methods. Motivated by previous works[8, 12], we apply self-attention networks to learn title representation $\mathbf{n}^t$ and entity representation $\mathbf{n}^e$ for a news n from its title and entities, individually. Besides, following Wu et al. [20], we derive representation $\mathbf{n}^v$ of news topic via a topic embedding layer. Finally we formulate news representation $\mathbf{n}$ as the aggregation of these representations: $\mathbf{n} = \mathbf{n}^t + \mathbf{n}^e + \mathbf{n}^v$.

2.3 Interest Matching and Model Learning

Based on the news modeling and candidate-aware user interest modeling method, we can learn representation $\mathbf{n}_c$ of candidate news n_c, and the corresponding user representation $\mathbf{u}$. We further match them to measure user interest in candidate news and calculate the matching score: $\hat{y} = \mathbf{n}_c^T \cdot \mathbf{u}$. Motivated by previous works [22, 25, 28], we adopted BPR loss [16] for model learning: $\mathcal{L} = -\frac{1}{H} \sum_{i=1}^{H} \log \phi(\hat{y}_i^p - \hat{y}_i^n)$, where H is the training dataset size, ϕ is sigmoid function, $\hat{y}_i^p$ and $\hat{y}_i^n$ is the matching score of the i-th positive and negative sample, and negative samples are randomly sampled for each positive sample from the same impression.

3 EXPERIMENT

3.1 Dataset and Experimental Settings

We conduct extensive experiments on two real-world datasets. The first one is a public news recommendation dataset (*MIND*) [31]. The second one is *NewsApp*, consisting of user logs collected from the

news feeds app of Microsoft from January 23 to April 01, 2020 (13 weeks). It contains 110,000 and randomly selected from the first ten weeks for training, and 100,000 impressions randomly selected from the last three weeks to construct the test set.

The data processing of *CAUM* follows Wu et al. [24]. In *CAUM*, dimensions of both news and user interest representations are set to 400. *Candi-SelfAtt* contains 20 attention heads, and output vectors of each head are 20-dimensional. *Candi-CNN* contains 400 filters and window size is set to 3. We train *CAUM* 3 epochs via Adam [6] with 5×10^{-5} learning rate. All hyper-parameters of *CAUM* and other baseline methods are selected based on the validation dataset. Codes are in https://github.com/taoqi98/CAUM. Following Wu et al. [31], we adopted AUC, MRR, nDCG@5, and nDCG@10 for evaluation.

3.2 Performance Comparison

We compare *CAUM* with several SOTA baseline methods: (1) *GRU* [10]: model user interest via a GRU network. (2) *DKN* [19]: apply a candidate-aware attention network to learn user representation. (3) *NAML* [20]: learn user representation via a attention network. (4) *NPA* [21]: propose a personalized attention network to model user interests. (5) *HiFi-Ark* [9]: learn user representation from multiple archives of user interests via a candidate-aware attention network. (6) *LSTUR* [1]: use GRU network and ID embeddings to model long- and short- term user interests. (7) *NRMS* [24]: apply self-attention network to model user interests. (8) *KRED* [8]: model news from title and entities via a KGAT network. (9) *GNewsRec* [3]: adopt GRU and GNN network to model user interests. (10) *FIM* [18]: utilize 3D-CNN to model user interests in candidate news based on word-level similarities between clicks and candidate.

We repeat each experiment 5 times and report average results and standard deviations in Table 1. First, we find that *CAUM* can significantly outperform other baseline methods which model user interests in a candidate-agnostic manner, such as *NRMS* and *LSTUR*. This is because users usually have multiple interests and a candidate news usually only matches a specific user interest. Modeling user interests in a candidate-agnostic manner makes these methods cannot effectively capture user interests that are relevant to a specific candidate news, and maybe sub-optimal for the interest matching. Second, our *CAUM* method also outperforms baseline methods with candidate-aware attention network, such as *DKN*, *HiFi-Ark*
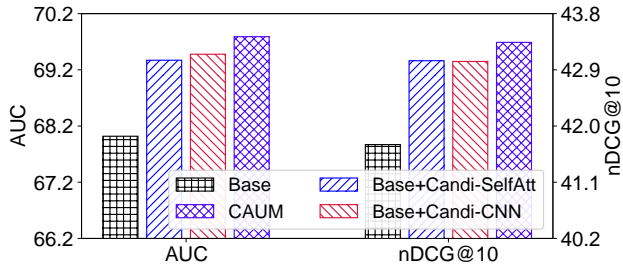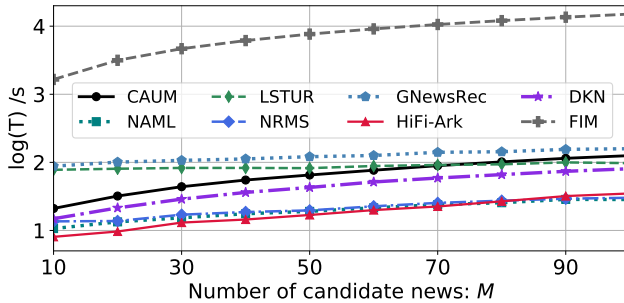
Figure 3: Ablation study of *CAUM*.

Figure 4: Normalized inference time of different methods.

GNewsRec. This is because different contexts of user's news clicks usually contain various clues to infer different user interests. Incorporating candidate news information into the behavior contexts modeling can help capture more relevant user interests for matching the candidate news. However, in these methods, user behavior contexts are ignored (e.g., *DKN*) or modeled in a candidate-agnostic way (e.g., *HiFi-Ark*), which are sub-optimal for the interest matching. Different from these methods, we propose *Candi-SelfAtt* and *Candi-CNN* to exploit candidate news to guide the user modeling from both long-range and local behavior contexts.

3.3 Ablation Study

We conduct an ablation study to verify the effectiveness of *Candi-SelfAtt* and *Candi-CNN* by adding them to the base model of *CAUM* (named *Base*). *Base* is a variation of *CAUM* that replaces *Candi-SelfAtt*, *Candi-CNN* and *Candi-Att* network with self-attention, CNN, and attention network individually. Results are shown in Fig. 3 and we have several findings. First, adding one of *Candi-SelfAtt* and *Candi-CNN* significantly improves the performance of *Base*. This result verifies that *Candi-SelfAtt* and *Candi-CNN* can effectively exploit behavior contexts to capture global and short-term user interests that are informative for matching the candidate news, respectively. Second, *CAUM* outperforms both *Base+CandiCNN* and *Base+Candi-SelfAtt*. This is because candidate-aware user interest modeling is important for both global and short-term user interest.

3.4 Analysis on Model Efficiency

We will present some efficiency analysis and comparisons on *CAUM* and other user modeling methods. First, in Table 2, we show time

Table 2: Method time complexity (multiplication operation) of calculating matching scores of M candidate news. News and user representation are d-dimensional.

NAML	$O(Md + Nd^2)$	GRU	$O(Md + Nd^2)$
LSTUR	$O(Md + Nd^2)$	NRMS	$O(3Nd^2 + N^2d + Md)$
CAUM	$O((3N + M)d^2 + (N^2 + MN)d)$		

complexities of *CAUM* and candidate-agnostic methods for calculating matching scores of M candidate news for a user.[1] A notable result is that although *CAUM* needs to calculate different user representations for different candidate news, the time complexity of *CAUM* is not M times that of other methods. This is because, in *CAUM*, many operations only need to be performed once for different candidate news such as calculating self-attention scores $\hat{r}_{i,j}^k$ between different clicked news. Thus, by avoiding executing duplicated calculations, the efficiency of *CAUM* can be significantly improved. Besides, in general, the number of candidate news M is usually in a small scale (e.g., 100) in real-world recommender systems and it is comparable with the number of users' clicked news N used for interest modeling (e.g., 50). Thus, in practical settings, *CAUM* can achieve comparable time complexity with *NRMS*. In addition, although *GRU* and *LSTUR* have smaller time complexity than *NRMS* and *CAUM*, it is difficult to speed up these RNN based methods via parallel computations and they usually cost more time in real applications. Second, as shown in Fig. 4, we compare running time T of different methods for calculating matching scores of M candidate news for 100,000 users. Different methods are executed in the same experimental environment (a Nvidia 1080 Ti GPU). We find that *CAUM* can achieve comparable speeds with many candidate-agnostic methods (e.g., *NAML* and *NRMS*) and outperform some candidate-agnostic methods (e.g., *LSTUR*). These results further verify that the efficiency of *CAUM* is satisfied like candidate-agnostic methods.

4 CONCLUSION

In this paper, we propose a candidate-aware user modeling framework for personalized news recommendation (*CAUM*), which can incorporate candidate news information into user modeling for more accurate interest matching. We propose a candidate-aware self-attention network to exploit candidate news information as clue to model global user interests in candidate news. In addition, we also propose a candidate-aware CNN network to incorporate candidate news information into local click behavior contexts modeling to match short-term user interests with the candidate news. Extensive experiments on two real-world datasets demonstrate that *CAUM* can significantly outperform many baseline methods and improve the accuracy of user modeling.

---

[1]These methods can directly exploit news representations calculated in advance.

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] Suyu Ge, Chuhan Wu, Fangzhao Wu, Tao Qi, and Yongfeng Huang. 2020. Graph enhanced representation learning for news recommendation. In *WWW*. 2863–2869.

[3] Linmei Hu, Chen Li, Chuan Shi, Cheng Yang, and Chao Shao. 2020. Graph neural news recommendation with long-term and short-term interest modeling. *IP&M* (2020), 102142.

[4] Linmei Hu, Siyong Xu, Chen Li, Cheng Yang, Chuan Shi, Nan Duan, Xing Xie, and Ming Zhou. 2020. Graph neural news recommendation with unsupervised preference disentanglement. In *ACL*. 4255–4264.

[5] Dhruv Khattar, Vaibhav Kumar, Vasudeva Varma, and Manish Gupta. 2018. Weave&Rec: A word embedding based 3-D Convolutional network for news recommendation. In *CIKM*. 1855–1858.

[6] Diederik P Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *ICLR*.

[7] Dongho Lee, Byungkook Oh, Seungmin Seo, and Kyong-Ho Lee. 2020. News Recommendation with Topic-Enriched Knowledge Graphs. In *CIKM*. 695–704.

[8] 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 *RecSys*. 200–209.

[9] Zheng Liu, Yu Xing, Fangzhao Wu, Mingxiao An, and Xing Xie. 2019. Hi-Fi ark: deep user representation via high-fidelity archive network. In *IJCAI*. 3059–3065.

[10] Shumpei Okura, Yukihiro Tagami, Shingo Ono, and Akira Tajima. 2017. Embedding-based news recommendation for millions of users. In *KDD*. 1933–1942.

[11] Tao Qi, Fangzhao Wu, Chuhan Wu, and Yongfeng Huang. 2021. Personalized News Recommendation with Knowledge-aware Interactive Matching. In *SIGIR*. 61–70.

[12] Tao Qi, Fangzhao Wu, Chuhan Wu, and Yongfeng Huang. 2021. PP-Rec: News Recommendation with Personalized User Interest and Time-aware News Popularity. In *ACL*. 5457–5467.

[13] Tao Qi, Fangzhao Wu, Chuhan Wu, Yongfeng Huang, and Xing Xie. 2020. Privacy-Preserving News Recommendation Model Learning. In *EMNLP: Findings*. 1423–1432.

[14] Tao Qi, Fangzhao Wu, Chuhan Wu, Yongfeng Huang, and Xing Xie. 2021. Uni-FedRec: A Unified Privacy-Preserving News Recommendation Framework for Model Training and Online Serving. In *EMNLP: Findings*. 1438–1448.

[15] Tao Qi, Fangzhao Wu, Chuhan Wu, Peiru Yang, Yang Yu, Xing Xie, and Yongfeng Huang. 2021. HieRec: Hierarchical User Interest Modeling for Personalized News Recommendation. In *ACL*. 5446–5456.

[16] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian personalized ranking from implicit feedback. In *UAI*. 452–461.

[17] 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 *NIPS*. 6000–6010.

[18] Heyuan Wang, Fangzhao Wu, Zheng Liu, and Xing Xie. 2020. Fine-grained interest matching for neural news recommendation. In *ACL*. 836–845.

[19] Hongwei Wang, Fuzheng Zhang, Xing Xie, and Minyi Guo. 2018. DKN: Deep knowledge-aware network for news recommendation. In *WWW*. 1835–1844.

[20] Chuhan Wu, Fangzhao Wu, Mingxiao An, Jianqiang Huang, Yongfeng Huang, and Xing Xie. 2019. Neural news recommendation with attentive multi-view learning. *IJCAI* (2019), 3863–3869.

[21] Chuhan Wu, Fangzhao Wu, Mingxiao An, Jianqiang Huang, Yongfeng Huang, and Xing Xie. 2019. Npa: Neural news recommendation with personalized attention. In *KDD*. 2576–2584.

[22] Chuhan Wu, Fangzhao Wu, Mingxiao An, Yongfeng Huang, and Xing Xie. 2019. Neural News Recommendation with Topic-Aware News Representation. In *ACL*. 1154–1159.

[23] Chuhan Wu, Fangzhao Wu, Mingxiao An, Tao Qi, Jianqiang Huang, Yongfeng Huang, and Xing Xie. 2019. Neural news recommendation with heterogeneous user behavior. In *EMNLP*. 4876–4885.

[24] Chuhan Wu, Fangzhao Wu, Suyu Ge, Tao Qi, Yongfeng Huang, and Xing Xie. 2019. Neural news recommendation with multi-head self-attention. In *EMNLP*. 6390–6395.

[25] Chuhan Wu, Fangzhao Wu, Yongfeng Huang, and Xing Xie. 2021. User-as-Graph: User Modeling with Heterogeneous Graph Pooling for News Recommendation. In *IJCAI*. 1624–1630.

[26] Chuhan Wu, Fangzhao Wu, Tao Qi, and Yongfeng Huang. 2020. SentiRec: Sentiment Diversity-aware Neural News Recommendation. In *AACL*. 44–53.

[27] Chuhan Wu, Fangzhao Wu, Tao Qi, and Yongfeng Huang. 2020. User modeling with click preference and reading satisfaction for news recommendation. In *IJCAI*. 3023–3029.

[28] Chuhan Wu, Fangzhao Wu, Tao Qi, and Yongfeng Huang. 2021. Empowering News Recommendation with Pre-trained Language Models. In *SIGIR*. 1652–1656.

[29] Chuhan Wu, Fangzhao Wu, Tao Qi, Jianxun Lian, Yongfeng Huang, and Xing Xie. 2020. PTUM: Pre-training User Model from Unlabeled User Behaviors via Self-supervision. In *EMNLP: Findings*. 1939–1944.

[30] Chuhan Wu, Fangzhao Wu, Xiting Wang, Yongfeng Huang, and Xing Xie. 2021. FairRec:Fairness-aware News Recommendation with Decomposed Adversarial Learning. In *AAAI*.

[31] 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*. 3597–3606.

[32] Jingwei Yi, Fangzhao Wu, Chuhan Wu, Ruixuan Liu, Guangzhong Sun, and Xing Xie. 2021. Efficient-FedRec: Efficient Federated Learning Framework for Privacy-Preserving News Recommendation. In *EMNLP*. 2814–2824.

[33] Guanjie Zheng, Fuzheng Zhang, Zihan Zheng, Yang Xiang, Nicholas Jing Yuan, Xing Xie, and Zhenhui Li. 2018. DRN: A deep reinforcement learning framework for news recommendation. In *WWW*. 167–176.