SHORT-PAPER

Multi-Interest Matching for Personalized News Recommendation with Large Language Models

HONGXIANG LIN, Peking University, Beijing, China

YIXIAO ZHOU, Peking University, Beijing, China

HUIYING HU, Peking University, Beijing, China

XIAOQING LYU, Peking University, Beijing, China

SIGIR '25: Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (July 2025)
https://doi.org/10.1145/3726302.3730233
ISBN: 9798400715921

Multi-Interest Matching for Personalized News Recommendation with Large Language Models

Hongxiang Lin
Wangxuan Institute of Computer Technology, Peking University
Beijing, China
Meituan Inc.
Beijing, China
linhx0@hotmail.com

Yixiao Zhou
Wangxuan Institute of Computer Technology, Peking University
Beijing, China
chnzyx2016@gmail.com

Huiying Hu
Wangxuan Institute of Computer Technology, Peking University
Beijing, China
190001779@pku.edu.cn

Xiaoqing Lyu*
Wangxuan Institute of Computer Technology, Peking University
Beijing, China
State Key Laboratory of Multimedia Information Processing, Peking University
Beijing, China
lvxiaoqing@pku.edu.cn

Abstract

Personalized news recommendation plays a vital role in mitigating information overload, yet challenges persist in accurately capturing user preferences and fine-grained interests. Leveraging the semantic understanding and extraction capabilities of large language models (LLMs), we propose a Multi-Interest Personalized News Recommendation (MIPNR) model to address these issues. MIPNR separately models user interests at the user, news, and entity levels. Specifically, we introduce a Category-Guided Interest-News Matching (CGIN-Matching) method to identify potential interests, a Local News Entity Graph (LNEG) to model subtle entity relationships, and an entity-wise attention mechanism to extract fine-grained interests. In addition, LLMs are used to generate explicit textual descriptions of user preferences. Extensive experiments on real-world datasets demonstrate the effectiveness of our approach.

CCS Concepts

• **Information systems → Retrieval models and ranking**.

Keywords

News recommendation, Large language model, User modeling

*Corresponding Author

SIGIR '25, Padua, Italy

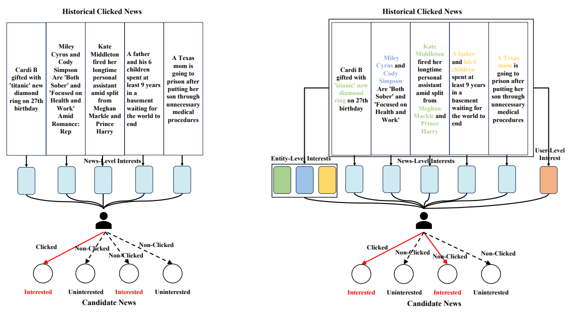
ACM ISBN 979-8-4007-1592-1/2025/07
https://doi.org/10.1145/3726302.3730233

ACM Reference Format:
Hongxiang Lin, Yixiao Zhou, Huiying Hu, and Xiaoqing Lyu. 2025. Multi-Interest Matching for Personalized News Recommendation with Large Language Models. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '25), July 13–18, 2025, Padua, Italy.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3726302.3730233

1 Introduction

The exponential growth of online news has exacerbated information overload, compelling platforms to deploy personalized recommendation systems that analyze user behavior through historical interactions [2, 5]. Conventional approaches typically aggregate news-level interests from clicked history into a unified representation for candidate news matching (Fig. 1). While effective, these methods often overlook nuanced user preferences embedded in nonclicked content and fine-grained semantic patterns. To bridge this gap, we propose a Multi-Level Interest (MLI) framework that systematically models user preferences through three complementary perspectives: (1) news-level interests derived from clicked history, (2) entity-level interactions extracted from implicit entity groups, and (3) explicit user-level preferences inferred via semantic analysis.

Despite advancements in personalized recommendation techniques [8, 16], critical limitations persist. First, existing methods predominantly rely on click-through signals to infer user interests, neglecting potential preferences in non-clicked news caused by time constraints or information saturation. Second, the granularity of interest modeling remains coarse, as entity-level semantic relationships within news content are insufficiently exploited. Third, user profiles are implicitly constructed through behavioral patterns rather than explicit preference representations, limiting interpretability and accuracy. To address these challenges, we present MIPNR, a Multi-Interest Personalized News Recommendation model



(a) Regular methods (b) Our method

Figure 1: (a) Regular methods and our method not only pays attention to the news-level interests, but also covers user-level and entity-level interests.

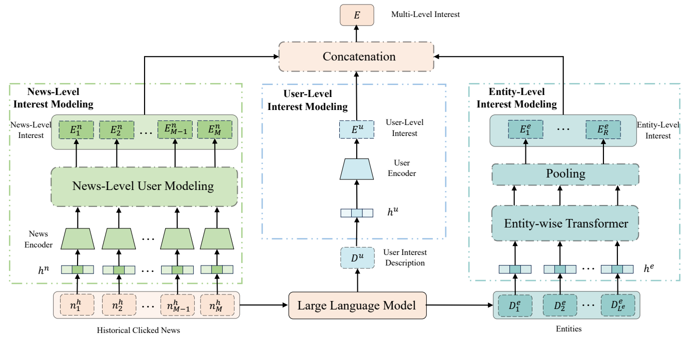


Figure 2: The user modeling in the framework of MIPNR.

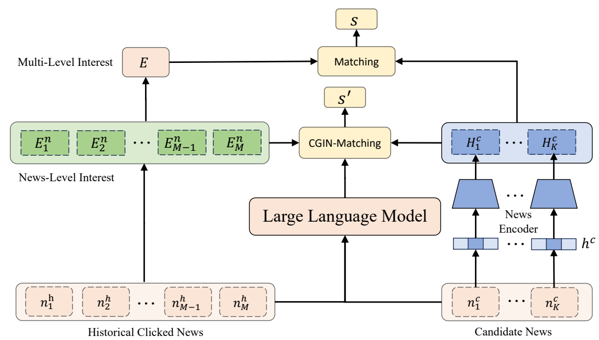


Figure 3: Two matching methods for MIPNR.

enhanced by large language models (LLMs). Our framework introduces three key innovations. First, a Category-Guided Interest-News Matching (CGIN-Matching) mechanism leverages LLMs to enrich sparse category labels (e.g., on the MIND dataset [15]), enabling the identification of latent interests in both clicked and non-clicked news through category intersection analysis. Second, a Local News Entity Graph (LNEG) constructs entity-group relationships from LLM-extracted entities, refining entity-level interests via attention-based feature enhancement. Third, LLMs directly generate explicit user-level interest profiles by semantically encoding historical interactions. Extensive experiments validate the superiority of MIPNR over state-of-the-art baselines, demonstrating its capability to mitigate label sparsity, capture fine-grained semantics, and improve recommendation transparency.

Our contributions include: (1) a unified multi-interest recommendation framework integrating LLM-enhanced news, entity, and user-level modeling; (2) a category-guided matching strategy addressing potential interest discovery; (3) an entity interaction graph with attention mechanisms for fine-grained preference extraction; and (4) empirical validation on real-world benchmarks.

2 The Proposed Model

2.1 Problem Formulation

In personalized news recommendation, each user u is associated with certain news, including a set of historical clicked news $N^h = \{n_1^h, n_2^h, \cdots, n_M^h\}$, and a set of candidate news $N^c = \{n_1^c, n_2^c, \cdots n_K^c\}$, where M and K are the number of historical clicked news and candidate news, respectively. In the candidate news N^c, n_1^c is the clicked news, while the others are non-clicked news. Each news $n_i^h \in N^h$ and $n_j^c \in N^c$ respectively has textual information T_i^h and T_j^c, as well as category sets c_i^h and c_j^c. Each textual information T_i^h or T_j^c contains at most L_n words. Our goal is to model the user interest and then compute the matching score $\boldsymbol{s}$ between user u and candidate news N^c based on the user interest. As shown in Figure 2, the interest has three levels, including user-level interest $\boldsymbol{E}^u \in \mathbb{R}^d$, news-level interest $\boldsymbol{E}^n = [\boldsymbol{E}_1^n; \boldsymbol{E}_2^n; \cdots; \boldsymbol{E}_M^n] \in \mathbb{R}^{M \times d}$, and entity-level interest $\boldsymbol{E}^e = [\boldsymbol{E}_1^e; \boldsymbol{E}_2^e; \cdots; \boldsymbol{E}_R^e] \in \mathbb{R}^{R \times d}$, where R represents the number of entity groups extracted from historical clicked news N^h, $[;]$ denotes row-wise concatenation operation, and d denotes the dimension. Finally, we fuse $\boldsymbol{E}^u$, $\boldsymbol{E}^n$, and $\boldsymbol{E}^e$ into a multi-level interest $\boldsymbol{E} \in \mathbb{R}^d$, which is used to compute the matching score $\boldsymbol{s}$ with candidate news.

2.2 Category-Guided Interest-News Matching

Current personalized news recommendation methods focus on the match between users and clicked news, typically using NCE loss [4] for training to boost the matching scores of clicked news and lower those of non-clicked news. However, in reality, users may not click on news of interest due to various reasons, leading to overly low matching scores for such non-clicked news by existing methods. To tackle this, we propose a CGIN-Matching loss to elevate the matching scores of candidate news that users might click on in the future.

We determine whether a historical clicked news n_i^h and a candidate news n_j^c belong to the same category attribute to ascertain if they match (category overlap signals), thereby obtaining supervision labels for CGIN-Matching. We obtain the label $\boldsymbol{y} \in \mathbb{R}^{M \times K}$ by determining whether there is an intersection between c_i^h and c_j^c, i.e., if there is an intersection between c_i^h and c_j^c, then $\boldsymbol{y}_{ij} = 1$, otherwise $\boldsymbol{y}_{ij} = 0$.

However, in the current dataset, e.g. MIND [15], each news has only one category, which results in correspondingly sparse supervision labels. To tackle this problem, we fully exploit the powerful

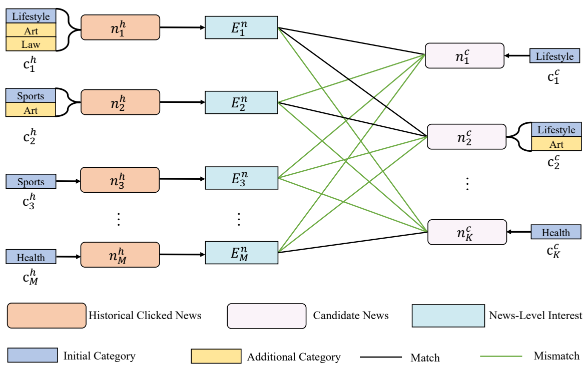


Figure 4: Design of supervision labels for CGIN-Matching loss by category overlap signals.

text information extraction capability of LLMs. Based on the additional self-defined category set, we use LLMs to select up to k categories for the historical click news n_i^h and candidate news n_j^c, which are then added to the category sets c_i^h and c_j^c, respectively. For the MIND dataset, each initial category set c_i^h and c_j^c contains only one element. After expansion, the number of elements in the category sets c_i^h and c_j^c is between 1 and $k+1$.

As shown in Figure 4, the category set c_1^h of the historical clicked news n_1^h include category nouns ($Lifestyle$, Art, Law), while the category set c_1^c and c_2^c of the candidate news n_1^c and n_2^c respectively include category nouns $Lifestyle$ and ($Lifestyle, Art$). c_1^h intersects with both c_1^c and c_2^c, thus the news-level interest E_1^n extracted from the historical clicked news n_1^h matches with n_1^c and n_2^c, i.e., $\boldsymbol{y}_{11} = 1$ and $\boldsymbol{y}_{12} = 1$.

2.2.1 Feature Encoder. We adopt a two-layer self-attention model as news encoder ϕ^n and user encoder ϕ^u, and the self-attention mechanism is defined as follows:

$$Attention(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \text{Softmax}(\frac{\boldsymbol{Q}\boldsymbol{K}^T}{\sqrt{d}})\boldsymbol{V}, \tag{1}$$

where $\boldsymbol{Q}$, $\boldsymbol{K}$, and $\boldsymbol{V}$ are query, key and value embeddings.

After obtaining the labels for CGIN-Matching, we also need to encode the historical clicked news items N^h and candidate news items N^c for the user. As shown in Figures 2 and 3, we obtain the word embeddings $\boldsymbol{h}_i^n \in \mathbb{R}^{L^n \times d}$ and $\boldsymbol{h}_j^c \in \mathbb{R}^{L^n \times d}$ of the text information T_i^h and T_j^c for each news $n_i^h \in N^h$ and $n_j^c \in N^c$ respectively. Subsequently, we input $\boldsymbol{h}^n \in \mathbb{R}^{M \times L^n \times d}$ and $\boldsymbol{h}^c \in \mathbb{R}^{K \times L^n \times d}$ into the news encoder ϕ^n and perform average pooling to obtain feature representations $\boldsymbol{H}^n \in \mathbb{R}^{M \times d}$ and $\boldsymbol{H}^c \in \mathbb{R}^{K \times d}$.

2.2.2 News-Level User Modeling. We adopt the user modeling component in FUM [6], CAUM [7], NRMS [12], and MINER [2] as our news-level user modeling module Φ to validate the effectiveness of our MIPNR model in various personalized news recommendation scenarios. We take the hidden feature $\boldsymbol{H}^n$ of clicked news N^h as input and output multiple news-level interest vectors $\boldsymbol{E}^n \in \mathbb{R}^{M \times d}$ representing user interests by $\boldsymbol{E}^n = \Phi(\boldsymbol{H}^n)$.

2.2.3 CGIN-Matching Loss. For all users in the training dataset $\mathcal{D}_{tr}$, we obtain the total news-level interests $\boldsymbol{E}_{tr}^n \in \mathbb{R}^{|\mathcal{D}_{tr}| \times M \times d}$ and the total training label $\boldsymbol{y}^{tr} \in \mathbb{R}^{|\mathcal{D}_{tr}| \times M \times K}$. As shown in Figure 3, we calculate CGIN-Matching scores $\boldsymbol{s}' \in \mathbb{R}^{|\mathcal{D}_{tr}| \times M \times K}$ with the candidate news representation $\boldsymbol{H}^c \in \mathbb{R}^{K \times d}$:

$$\boldsymbol{s}' = \boldsymbol{E}_{tr}^n (\boldsymbol{H}^c)^T, \tag{2}$$

where $\boldsymbol{s}'_{ijk}$ represents the matching score between the news-level interest $\boldsymbol{E}_j^n$ extracted from the j-th historical clicked news n_j^h of user u_i and the k-th candidate news n_k^c.

After obtaining the matching scores between the news-level interests and the candidate news, in order to enhance the matching scores for the candidate news of interest, we define the CGIN-Matching loss as follows:

$$\mathcal{L}_{cgin} = -\frac{1}{|\mathcal{D}_{tr}|} \sum_{i=1}^{|\mathcal{D}_{tr}|} \sum_{j=1}^{M} \log \frac{\sum_{k=1}^{K} \exp(\mathbb{1}_{\boldsymbol{y}_{ijk}^{tr}=1} \boldsymbol{s}'_{ijk})}{\sum_{l=1}^{K} \exp(\boldsymbol{s}'_{ijl})}, \tag{3}$$

where $\mathbb{1}_{\boldsymbol{y}_{ijk}^{tr}=1} \in \{0, 1\}$ evaluate to 1 when the supervision label $\boldsymbol{y}_{ijk}^{tr}$ equals to 1, which means the k-th candidate news n_k^c is a positive sample for the j-th news-level interest vector $\boldsymbol{E}_j^n$ of user u_i.

2.3 Local News Entity Graph

We extract user interests from both individual news articles and entities across different news. As shown in Figure 1, historical clicked news contains three entity groups: (1) Family life ("A father," "his 6 children," "A Texas mom"), (2) Romantic relationship ("Miley Cyrus," "Cody Simpson"), and (3) Celebrity life ("Kate Middleton," "Meghan Markle," "Prince Harry," "'Titanic' new diamond ring"). However, it is challenging to directly extract these associated entities from historical clicked news. Due to the power of LLMs, we utilize LLMs to extract L^e entities $D^e = \{D_1^e, D_2^e, \cdots, D_{L^e}^e\}$, and divide them into R entity groups based on whether they belong to the same interest, each representing an entity-level interest.

To model relationships between entities, we introduce a Local News Entity Graph (LNEG), where entities become nodes. Nodes in the same group (e.g., "Celebrity life") are fully connected. To link different interests, we randomly add edges between nodes of different groups with probability p. For the entity-level attention, we define the adjacency matrix $\boldsymbol{A} \in \mathbb{R}^{L^e \times L^e}$ of LNEG, where $\boldsymbol{A}_{ij} = 1$ if nodes i and j are connected, and 0 otherwise.

2.3.1 Entity-wise Attention. The entity-wise attention uses a two-layer self-attention network ψ. We compute entity embeddings $\boldsymbol{h}^e \in \mathbb{R}^{L^e \times d}$ from word embeddings, then obtain entity features $\boldsymbol{H}^e = \psi(\boldsymbol{h}^e)$. Since entities in the same group share similar interests, we employ local attention (first layer) to capture these relationships, followed by global attention (second layer, Eq. 1) for improved performance. The local attention is defined as follows:

$$Attention(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \text{Softmax}(\frac{\boldsymbol{Q}\boldsymbol{K}^T}{\sqrt{d}} \cdot \boldsymbol{A})\boldsymbol{V}. \tag{4}$$

2.3.2 Pooling. The index $\boldsymbol{I} \in \mathbb{R}^{L^e}$ of entity features $\boldsymbol{H}^e$ is defined as $\boldsymbol{I} = [1, 1, \cdots, R, R]$, where $\boldsymbol{I}_i$ means the i-th entity is in the $\boldsymbol{I}_i$-th interest group. We obtain the entity-level interest $\boldsymbol{E}_r^e \in \mathbb{R}^d$ of the

Table 1: Performance comparison with different baselines.

	MIND-small				MIND-large			
	AUC	MRR	nDCG@5	nDCG@10	AUC	MRR	nDCG@5	nDCG@10
NAML [11]	65.53	30.37	33.10	39.39	66.51	32.32	35.19	40.88
LSTUR [1]	64.37	29.43	31.92	38.15	68.39	33.25	36.33	42.04
FIM [10]	65.05	30.29	32.95	39.15	67.81	33.28	36.24	41.89
HieRec [9]	67.95	32.87	36.36	42.53	69.03	33.89	37.08	43.01
Fastformer+PLM-NR* [14]	70.03	35.96	38.12	43.98	72.68	37.45	41.51	46.84
FUM [6]	66.14	30.52	34.05	40.46	69.51	33.81	37.38	43.38
FUM-MIPNR-L	67.83	31.97	36.15	42.32	71.64	35.41	39.29	45.26
FUM-MIPNR-C	68.01	32.13	36.32	42.48	71.89	35.53	39.56	45.47
CAUM [7]	66.15	31.07	34.18	40.57	69.09	33.53	37.31	43.72
CAUM-MIPNR-L	67.80	32.76	36.04	41.94	71.10	35.34	39.49	45.60
CAUM-MIPNR-C	68.23	33.56	36.84	42.10	71.30	35.58	39.73	45.92
NRMS [12]	64.93	30.04	32.56	38.96	68.08	33.34	36.26	41.95
NRMS-UniLM [13]	68.54	33.73	36.02	42.87	70.64	35.39	38.71	44.38
NRMS-ONCE [3]	68.74	36.66	38.60	44.37	-	-	-	-
NRMS-MIPNR-L	68.30	34.70	37.13	42.61	70.72	35.59	38.87	44.49
NRMS-MIPNR-C	68.04	34.61	37.04	42.48	70.33	35.32	38.65	44.22
MINER [2]	69.61	33.97	37.62	43.90	71.51	36.18	39.72	45.34
MINER-ONCE [3]	68.92	36.74	38.72	44.48	-	-	-	-
MINER-MIPNR-L	70.63	36.55	38.76	44.57	72.49	37.16	40.76	46.57
MINER-MIPNR-C	**71.03**	**36.92**	**38.98**	**44.85**	**73.08**	**37.95**	**41.93**	**47.02**

r-th group by pooling the attention outputs:

$$E_r^e = \frac{\sum_{i=1}^{L^e} \mathbb{1}_{I_i=r} H_i^e}{\sum_{i=1}^{L^e} \mathbb{1}_{I_i=r}}. \tag{5}$$

The entity-level interest matrix $E^e \in \mathbb{R}^{R \times d}$ is defined as $E^e = [E_1^e; E_2^e; \cdots; E_R^e]$ where $[;]$ denotes row-wise concatenation operation.

2.4 Multi-Level Interest

In the process of user modeling, we extract the interest vectors of users at three levels. However, at the user level, due to the need for privacy protection, existing methods cannot directly obtain text descriptions about users, which express user interests more explicitly and specifically. Thus, we adopt LLMs to extract text descriptions of user interests based on their news click history, thereby improving the accuracy of recommendations.

For user-level interest modeling, as shown in Figure 2, we first employ LLMs to extract the user interest description D^u. Subsequently, we perform a decomposition of the interest description D^u into multiple words. After that, we extract the first L^u words, obtain the word embedding of each word, and ultimately derive the overall embedding $h^u \in \mathbb{R}^{L^u \times d}$ that represents the interest description D^u. Finally, we input h^u into the user encoder ϕ^u and perform average pooling to obtain the user-level interest vector $E^u \in \mathbb{R}^d$.

We concatenate the three-level interests and perform pooling of the interests to obtain the multi-level interest:

$$E' = [E^u || E^n || E^e]. \tag{6}$$

$$E = \sum_{i=1}^{R+M+1} \frac{\exp(E_i' W + b)}{\sum_{j=1}^{R+M+1} \exp(E_j' W + b)} E_i', \tag{7}$$

where $W \in \mathbb{R}^{d \times d}$ and $b \in \mathbb{R}^d$ are learnable. $[||]$ denotes column-wise concatenation operation.

3 Experiments

3.1 Datasets

Due to the scarcity of news recommendation datasets, we adopt the only public large-scale dataset, MIND [15]. MIND dataset consists of two distinct versions, namely MIND-small and MIND-large.

Table 2: Experimental results of component ablation.

	AUC	MRR	nDCG@5	nDCG@10
MINER-MIPNR-C (w/o U)	70.21	35.47	38.00	44.13
MINER-MIPNR-C (w/o L)	70.25	35.68	37.90	44.37
MINER-MIPNR-C (w/o C)	70.10	35.06	37.25	44.04
MINER-MIPNR-C (w/o EA)	70.42	35.99	37.89	44.31
MINER-MIPNR-C (w/o E)	70.24	36.02	38.05	44.39
MINER-MIPNR-C	**71.03**	**36.92**	**38.98**	**44.85**

3.2 Comparison Baselines

To comprehensively evaluate the performance of MIPNR, we conduct an in-depth comparison with the LLM4Rec method and personalized news recommendation methods. LLM4Rec methods include: ONCE [3]. Personalized news recommendation methods include: NAML [11], LSTUR [1], FIM [10], NRMS [12], UniLM [13], HieRec [9], Fastformer [14], MINER [2], FUM [6], CAUM [7].

3.3 Overall Performance Comparison

We experimented on the publicly available MIND dataset, with results in Table 1 (percentages without the "%" symbol, best in bold). Applying MIPNR on FUM, CAUM, NRMS, and MINER led to significant improvement. Baselines like HieRec and FIM, using multilevel interest modeling, have limitations: insufficient consideration of non-clicked news user interest and under-utilization of LLMs' information extraction. MIPNR outperforms the LLM4Rec baseline ONCE in recommendation. Comparing LLaMA2-7b and ChatGPT (gpt-3.5-turbo) in MIPNR, ChatGPT-based MIPNR is better, mainly due to its stronger information generalization and extraction.

3.4 Ablation Study

We conducted ablation experiments on MINER-MIPNR-C in the MIND-small dataset (results in Table 2) to explore MIPNR components. (w/o U, L, C, EA, E) represent removing different elements. Removing modules decreased performance, with (w/o C) having the sharpest drop, showing users' interest in non-clicked news. (w/o L) degradation indicates sparse labels harm CGIN-Matching loss training.

4 Conclusion

In this paper, we propose a multi-interest personalized news recommendation model. To consider potential interests of users, we introduce a CGIN-Matching method. To capture subtle relationships between news entities, we introduce a LNEG and propose an entity-wise attention mechanism to extract fine-grained interests. In addition, we adopt LLMs to generate textual descriptions of user interests to directly express user preferences.

5 Acknowledgement

This work is supported by the projects of Beijing Science and Technology Program (Z231100007423011) and National Natural Science Foundation of China (No. 62376012), which is also a research achievement of State Key Laboratory of Multimedia Information Processing and Key Laboratory of Science, Technology and Standard in Press Industry (Key Laboratory of Intelligent Press Media Technology).

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 *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL)*. 336–345.

[2] 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 *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL)*. 343–352.

[3] Qijiong Liu, Nuo Chen, Tetsuya Sakai, and Xiao-Ming Wu. 2023. ONCE: Boosting Content-based Recommendation with Both Open- and Closed-source Large Language Models. *arXiv preprint arXiv:2305.06566* (2023).

[4] Andriy Mnih and Yee Whye Teh. 2012. A fast and simple algorithm for training neural probabilistic language models. In *Proceedings of International Coference on International Conference on Machine Learning (ICML)*. 419–426.

[5] Tao Qi, Fangzhao Wu, Chuhan Wu, and Yongfeng Huang. 2021. Personalized news recommendation with knowledge-aware interactive matching. In *Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*. 61–70.

[6] Tao Qi, Fangzhao Wu, Chuhan Wu, and Yongfeng Huang. 2022. FUM: fine-grained and fast user modeling for news recommendation. In *Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*. 1974–1978.

[7] Tao Qi, Fangzhao Wu, Chuhan Wu, and Yongfeng Huang. 2022. News recommendation with candidate-aware user modeling. In *Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*. 1917–1921.

[8] 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 International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*. 1164–1173.

[9] 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 *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL)*. 5446–5456.

[10] Heyuan Wang, Fangzhao Wu, Zheng Liu, and Xing Xie. 2020. Fine-grained interest matching for neural news recommendation. In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL)*. 836–845.

[11] 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 International Joint Conference on Artificial Intelligence (IJCAI)*. 3863–3869.

[12] 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 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 6389–6394.

[13] Chuhan Wu, Fangzhao Wu, Tao Qi, and Yongfeng Huang. 2021. Empowering news recommendation with pre-trained language models. In *Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*. 1652–1656.

[14] Chuhan Wu, Fangzhao Wu, Tao Qi, Yongfeng Huang, and Xing Xie. 2021. Fastformer: Additive attention can be all you need. *arXiv preprint arXiv:2108.09084* (2021).

[15] 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 *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL)*. 3597–3606.

[16] Qi Zhang, Qinglin Jia, Chuyuan Wang, Jingjie Li, Zhaowei Wang, and Xiuqiang He. 2021. Amm: Attentive multi-field matching for news recommendation. In *Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*. 1588–1592.