

HieRec: Hierarchical User Interest Modeling for Personalized News Recommendation

Tao Qi¹, Fangzhao Wu², Chuhan Wu¹, Peiru Yang¹, Yang Yu², Xing Xie² and Yongfeng Huang¹

¹Department of Electronic Engineering & BNRist, Tsinghua University, Beijing 100084, China

²Microsoft Research Asia, Beijing 100080, China

{taoqi.qt, wufangzhao, wuchuhan15, peiruyang17}@gmail.com
yfhuang@mail.tsinghua.edu.cn
{t-yyu, xing.xie}@microsoft.com

Abstract

User interest modeling is critical for personalized news recommendation. Existing news recommendation methods usually learn a single user embedding for each user from their previous behaviors to represent their overall interest. However, user interest is usually diverse and multi-grained, which is difficult to be accurately modeled by a single user embedding. In this paper, we propose a news recommendation method with hierarchical user interest modeling, named *HieRec*. Instead of a single user embedding, in our method each user is represented in a hierarchical interest tree to better capture their diverse and multi-grained interest in news. We use a three-level hierarchy to represent 1) overall user interest; 2) user interest in coarse-grained topics like sports; and 3) user interest in fine-grained topics like football. Moreover, we propose a hierarchical user interest matching framework to match candidate news with different levels of user interest for more accurate user interest targeting. Extensive experiments on two real-world datasets validate our method can effectively improve the performance of user modeling for personalized news recommendation.

1 Introduction

Recently, massive people are habituated to reading news articles on online news platforms, such as Google News and Microsoft News (Khattar et al., 2018; Das et al., 2007). To help users efficiently obtain their interested news information, personalized news recommendation technique that aims to recommend news according to user interests, is widely used by these platforms (Wu et al., 2020a; Liu et al., 2010; Lin et al., 2014).

User interest modeling is a critical step for personalized news recommendation (Wu et al., 2021; Zheng et al., 2018; Wu et al., 2020c). Existing methods usually learn a single representation vector

ID	Click	Topic	Subtopic	Title
1	✓	Movies	Celebrity	Ben Affleck breaks silence after "slip" in sobriety.
2	✓	Sports	Football	Myles Garrett suspended indefinitely by the NFL.
3	✓	Sports	Football	Jaguars veteran cornerback Josh Robinson retires suddenly.
4	✗	Sports	Basketball	Trey Lyles off to a good start with Spurs.
5	✗	Sports	Golf	Can an amateur win again on the PGA Tour?
6	✓	Finance	Stocks	3M is a dog of the Dow -- and it may not get better in 2020.
7	✓	Finance	Taxes	New Trump tax documents show major inconsistencies.
8	✓	Health	Fitness	This guy lost 30 pounds and gained a rock-hard pack.
9	✗	Lifestyle	Food	Candy Corn has tired to the Metro East.

Figure 1: Click and non-click logs of an example user.

to model overall user interests from users' clicked news (Okura et al., 2017; Wu et al., 2020b; An et al., 2019). For example, Okura et al. (2017) used a GRU network to model user interests from clicked news. They used the latest hidden state of GRU as the user interest representation. Wu et al. (2019e) used multi-head self-attention network to capture user interests, and used an attentive pooling network to obtain a unified user representation. However, user interest is usually diverse and multi-grained. For example, as shown in Fig. 1, a user may have interest in movies, sports, finance and health at the same time. In addition, for users who are interested in sports, some of them may have general interest in this area, while other users like the example user in Fig. 1 may only have interest in a specific sport like football. However, it is difficult for these methods to accurately model the diverse and multi-grained user interest for news recommendation via a single user embedding.

In this paper, we propose a personalized news recommendation approach with hierarchical user interest modeling, named *HieRec*, which can effectively capture the diverse and multi-grained user interest. Our approach contains three levels of user interest representations to model user interests in different aspects and granularities. The first one is subtopic-level, which contains multiple interest representations to model fine-grained user interests in different news subtopics (e.g., interest in football and golf). They are learned from embeddings of subtopics and the clicked news in the correspond-

ing subtopics. The second one is topic-level, which contains multiple interest representations to capture coarse-grained user interests in major news topics (e.g., interest in sports and finance). They are learned from embeddings of news topics and their subordinate subtopic-level interest representations. The third one is user-level, which contains an interest representation to model overall user interests. It is learned from topic-level interest representations. Besides, we propose a hierarchical user interest matching framework to match candidate news with different levels of interest representations to target user interests more accurately. Extensive experiments on two real-world datasets show that *HieRec* can effectively improve the accuracy of user interest modeling and news recommendation.

2 Related Work

Personalized news recommendation is an important intelligent application and is widely studied in recent years (Bansal et al., 2015; Wu et al., 2019c; Qi et al., 2020; Ge et al., 2020). Existing methods usually model news from its content, model user interest from user’s clicked news, and recommend candidate news based on their relevance with user interests (Okura et al., 2017). For example, Okura et al. (2017) utilized an auto-encoder to learn news representations from news bodies. They applied a GRU network to capture user interests from the sequence of users’ historical clicks and used the last hidden state vector of GRU as user interest representation. Besides, they proposed to model relevance between user interest and candidate news based on the dot product of their representations. Wu et al. (2019a) learned news representations from news titles, bodies, categories, and subcategories based on an attentive multi-view learning framework. They build user interest representation based on the attentive aggregation of clicked news representations. An et al. (2019) used a CNN network to learn news representations from news titles and categories. They applied a GRU network to user’s clicked news to build a short-term user interest representation and applied user ID embedding to learn long-term user interest representation. They further learned a unified user interest representation based on the aggregation of short- and long-term user interest representation. Liu et al. (2020) proposed to learn news representations from news titles and entities via a knowledge graph attention network. They also obtained

user interest representation from representations of clicked news via an attention network. Besides, all of these three methods adopted the inner product for matching candidate news. Most existing methods learn a single user embedding to represent the overall user interests (Wang et al., 2018; Wu et al., 2019e,b). However, user interests are usually very diverse and multi-grained, which are difficult to be accurately modeled by a single user embedding. Different from these methods, we propose a hierarchical user interest modeling framework to model user interests in different aspects and granularities. In addition, we propose a hierarchical user interest matching framework to understand user interest in candidate news from different interest granularities for more accurate user interest targeting.

3 HieRec

In this section, we first give a problem formulation of personalized news recommendation. Then we introduce our *HieRec* method in detail.

3.1 Problem Formulation

Given a candidate news n_c and a target user u , the goal is calculating an interest score o to measure the interest of this user in the candidate news. Each news n has a title, a topic t and a subtopic s . The title is composed of a text sequence $\mathbf{T} = [w_1, w_2, \dots, w_T]$ and an entity sequence $\mathbf{E} = [e_1, e_2, \dots, e_E]$, where w_i and e_i respectively denote the i -th word and entity in news title, T and E respectively denote the number of words and entities. We assume the user has M clicked news. In *HieRec*, we further divide these clicks based on their topics and subtopics for hierarchical user interest modeling. More specifically, we build a clicked topic set $\{t_i | i = 1, \dots, m\}$ from topics of user’s clicks, where t_i is the i -th clicked topic and m is the number of clicked topics. We can further obtain a clicked subtopic set $\{s_j^i | j = 1, \dots, d\}$ subordinate to each clicked topic t_i , where s_j^i is the j -th clicked subtopic subordinate to topic t_i and d is the size of the set. Finally, user’s clicked news in topic t_i and subtopic s_j^i are divided into the same click group $\mathcal{N}_j^i = \{n_k^{i,j} | k = 1, \dots, l\}$, where $n_k^{i,j}$ denotes the k -th clicked news in this group and l is the number of clicked news in the group.

3.2 Hierarchical User Interest Modeling

In general, user interest is usually very diverse and multi-grained. For example, according to Fig. 1,

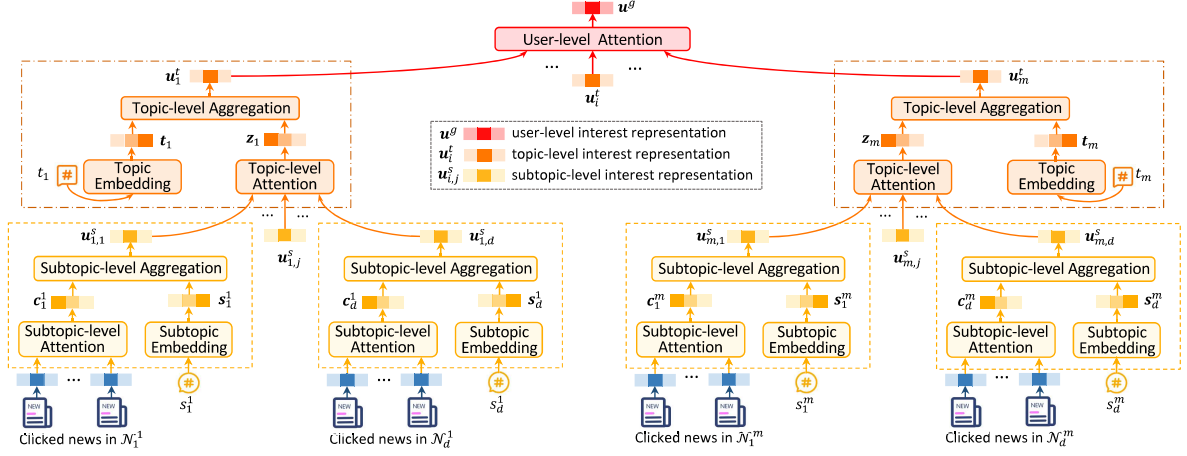


Figure 2: Framework of hierarchical user interest modeling in *HieRec*.

the example user has interests in many different aspects at the same time, such as sports, movies, and finance. Besides, for users who are interested in sports, some of them may have general interests in this area and may read news in different kinds of sports, such as basketball, football, golf, and so on. While other users (like the example user in Fig. 1) may only have interest in a specific sport like football. Understanding user interest in different aspects and granularities has the potential to model user interests more accurately. Thus, we propose a hierarchical user interest modeling framework, which learns a hierarchical interest tree to capture diverse and multi-grained user interest. As shown in Fig. 2, *HieRec* represents user interests via a three-level hierarchy.

First, we learn multiple subtopic-level interest representations to model fine-grained user interests in different news subtopics (e.g. football and golf). The subtopic-level interest representation for subtopic s_j^i is learned from \mathcal{N}_j^i that is composed of user's clicked news in subtopic s_j^i . Since clicked news may have different informativeness for modeling user interest, we adopt a subtopic-level attention network to select informative clicked news for modeling user interest in subtopic s_j^i :

$$\mathbf{c}_j^i = \sum_{k=1}^l \gamma_k \mathbf{n}_k^{i,j}, \quad \gamma_k = \frac{\exp(\phi_s(\mathbf{n}_k^{i,j}))}{\sum_{p=1}^l \exp(\phi_s(\mathbf{n}_p^{i,j}))}, \quad (1)$$

where γ_k denotes the attention weight of the k -th clicked news $\mathbf{n}_k^{i,j}$ in \mathcal{N}_j^i , $\mathbf{n}_k^{i,j}$ is the representation of news $\mathbf{n}_k^{i,j}$ (Section. 3.4 introduces how to obtain it) and $\phi_s(\cdot)$ denotes a dense network. Besides, we also adopt a subtopic embedding layer to capture semantic information of different subtopics, from

which we can obtain the embedding vector \mathbf{s}_j^i of subtopic s_j^i . Finally, we learn the subtopic-level user interest representation $\mathbf{u}_{i,j}^s$ based on the combination of \mathbf{c}_j^i and \mathbf{s}_j^i , i.e., $\mathbf{u}_{i,j}^s = \mathbf{c}_j^i + \mathbf{s}_j^i$. Similarly, we also learn subtopic-level interest representations for other subtopics clicked by the user.

Second, we learn multiple topic-level interest representations to model coarse-grained user interests in major news topics (e.g. sports and finance). The topic-level interest representation for a clicked topic t_i is learned from subtopic-level interest representations $\{\mathbf{u}_{i,j}^s | j = 1, \dots, d\}$ of subtopics $\{s_j^i | j = 1, \dots, d\}$ subordinate to the topic t_i . More specifically, user interests in different subtopics may have different importance for modeling user interest in a specific topic. Besides, the number of clicked news on a subtopic may also reflect its importance for modeling topic-level user interest. Thus, we utilize a topic-level attention network to select important subtopic-level user interest representations to model user interest in topic t_i :

$$\mathbf{z}_i = \sum_{j=1}^d \beta_j \mathbf{u}_{i,j}^s, \quad \beta_j = \frac{\exp(\phi_t(\mathbf{v}_{i,j}^s))}{\sum_{k=1}^d \exp(\phi_t(\mathbf{v}_{i,k}^s))}, \quad (2)$$

where $\mathbf{v}_{i,j}^s = [\mathbf{u}_{i,j}^s; \mathbf{r}_j^i]$, \mathbf{r}_j^i is the embedding vector for the number of clicked news on subtopic s_j^i , $[\cdot; \cdot]$ is the concatenation operation, β_j is the attention weight of $\mathbf{u}_{i,j}^s$, and $\phi_t(\cdot)$ is a dense network. Besides, we also use a topic embedding layer to model semantic information of different topics and drive the embedding vector \mathbf{t}_i for topic t_i . Finally, we aggregate \mathbf{z}_i and \mathbf{t}_i to learn the topic-level user interest representation \mathbf{u}_i^t in topic t_i : $\mathbf{u}_i^t = \mathbf{z}_i + \mathbf{t}_i$. Similarly, we also learn topic-level interest representations for other clicked topics.

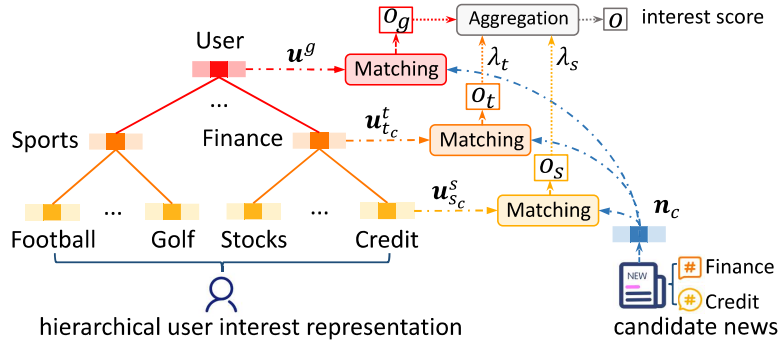


Figure 3: Framework of hierarchical user interest matching in *HieRec*.

Third, we learn a user-level interest representation \mathbf{u}^g to model overall user interests. It is learned from topic-level interest representations. Similarly, we adopt a user-level attention network to model relative importance of topic-level user interests to learn user-level interest representation:

$$\mathbf{u}^g = \sum_{i=1}^m \alpha_i \mathbf{u}_i^t, \quad \alpha_i = \frac{\exp(\phi_g(\mathbf{v}_i^t))}{\sum_{j=1}^m \exp(\phi_g(\mathbf{v}_j^t))}, \quad (3)$$

where $\mathbf{v}_i^t = [\mathbf{u}_i^t; \mathbf{r}_i]$, \mathbf{r}_i is the embedding vector for the number of user's clicked news on topic t_i , α_i denotes the attention weight of the i -th topic-level interest representation, and $\phi_g(\cdot)$ denotes a dense network for calculating attention scores.

3.3 Hierarchical User Interest Matching

Matching between candidate news and user interests at different granularities can provide various clues for user interest targeting. For example, according to Fig. 1, although all of the 3rd, 4th, and 5th news are about sports, the user only clicks the 3rd news probably because of her fine-grained interests in football rather than basketball and golf. This implies that the matching between candidate news and fine-grained user interests is useful for personalized news recommendation. Besides, not all candidate news can match with fine-grained user interests. For instance, a news on subtopic baseball cannot match any fine-grained interests of the example user in Fig. 1. Fortunately, the coarse-grained user interests (i.e., interest in sports) and overall user interests can match with this candidate news. This implies that matching candidate news with coarse-grained user interests and overall user interests is also important. Thus, we propose a hierarchical user interest matching framework, which models user interests in candidate news from different interest granularities. As shown in Fig. 3, it

takes candidate news (including its representation \mathbf{n}_c , topic t_c and subtopic s_c) and hierarchical user interest representation as input. First, we match candidate news with overall user interests and calculate a user-level interest score o_g based on the relevance between \mathbf{n}_c and \mathbf{u}^g : $o_g = \mathbf{n}_c \cdot \mathbf{u}^g$.

Second, topic-level interest representation $\mathbf{u}_{t_c}^t$ models coarse-grained user interests in the topic t_c of candidate news. It can provide coarse-grained information to understand user interest in candidate news. Thus, we match topic-level interest representation $\mathbf{u}_{t_c}^t$ with candidate news \mathbf{n}_c as: $\hat{o}_t = \mathbf{n}_c \cdot \mathbf{u}_{t_c}^t$. Besides, we can infer users may be more interested in topics that they have clicked more. Thus, we weights \hat{o}_t based on the ratio w_{t_c} of topic t_c in historical clicked news and obtained topic-level interest score o_t : $o_t = \hat{o}_t * w_{t_c}$. Besides, if the candidate news does not belong to any user's clicked topics, we set o_t as zero directly.

Third, subtopic-level interest representation $\mathbf{u}_{s_c}^s$ models fine-grained user interest in the subtopic s_c of candidate news and can be used to capture fine-grained user interests in candidate news. Thus, we match subtopic-level interest representation $\mathbf{u}_{s_c}^s$ and candidate news \mathbf{n}_c as: $\hat{o}_s = \mathbf{n}_c \cdot \mathbf{u}_{s_c}^s$. Similarly, we weights \hat{o}_s based on the ratio w_{s_c} of subtopic s_c in user's clicked news and obtain the subtopic-level interest score: $o_s = \hat{o}_s * w_{s_c}$.

Finally, interest scores of three different levels are aggregated to an overall interest score o :

$$o = \lambda_s o_s + \lambda_t o_t + (1 - \lambda_s - \lambda_t) o_g, \quad (4)$$

where $\lambda_t, \lambda_s, \in \mathbb{R}^+$ are hyper-parameters for controlling the relative importance of interest scores of different levels. Besides, we have $\lambda_t + \lambda_s < 1$.

3.4 News Representation

We introduce how to obtain news representation from texts and entities of news titles. As shown in

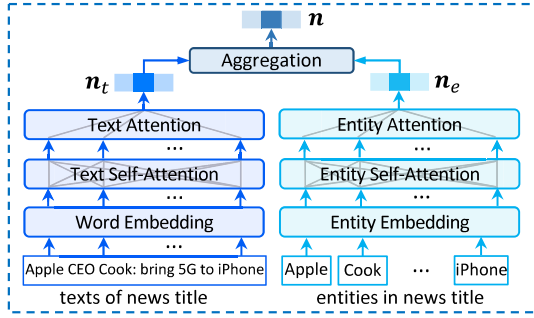


Figure 4: News representation learning framework.

Fig. 4, we first use a text encoder to model news texts. It first applies a word embedding layer to enrich semantic information of the model. Next, it adopts a text self-attention network (Vaswani et al., 2017) to learn word representations from contexts of news texts. Then, it uses a text attention network to learn text representation \mathbf{n}_t by aggregating word representations. Besides texts, knowledge graphs can also provide rich information for understanding news content via entities in news (Wang et al., 2018). Thus, we apply an entity encoder to learn entity representation of news. We first use an entity embedding layer to incorporate information from knowledge graphs into our model. We further apply an entity self-attention network to capture relatedness among entities. Next, we utilize an entity attention network to learn entity representation \mathbf{n}_e of news by aggregating entities. Finally, we build representation \mathbf{n} of news as: $\mathbf{n} = \mathbf{W}_t \mathbf{n}_t + \mathbf{W}_e \mathbf{n}_e$, where \mathbf{W}_t and \mathbf{W}_e are parameters.

3.5 Model Training

Following (Wu et al., 2019d), we utilize the NCE loss for model optimization. Given a positive sample n_i^+ (a clicked news) in the training dataset \mathcal{O} , we randomly select K negative samples $[n_i^1, \dots, n_i^K]$ (non-clicked news) for it from the same news impression displayed to the user u . The NCE loss \mathcal{L} requires the positive sample should be assigned a higher interest score o_i^+ than other negative samples $[o_i^1, \dots, o_i^K]$ and is formulated as:

$$\mathcal{L} = - \sum_{i=1}^{|\mathcal{O}|} \log \frac{\exp(o_i^+)}{\exp(o_i^+) + \sum_{j=1}^K \exp(o_i^j)}. \quad (5)$$

4 Experiment

4.1 Experimental Datasets and Settings

We conduct extensive experiments on two real-world datasets to evaluate the effectiveness of *Hi-*

	# News	# Topics	# Subtopics	# Users	# Clicks
<i>MIND</i>	65,238	18	270	94,057	347,727
<i>Feeds</i>	1,126,508	28	-	50,605	473,697

Table 1: Statistic information of the two datasets.

eRec. The first one is the public *MIND* dataset (Wu et al., 2020d)¹. It is constructed by user behavior data collected from Microsoft News from October 12 to November 22, 2019 (six weeks), where user data in the first four weeks was used to construct users’ reading history, user data in the penultimate week was used for model training and user data in the last week was used for evaluation. Besides, *MIND* contains off-the-shelf topic and subtopic label for each news. The second one (named *Feeds*) is constructed by user behavior data sampled from a commercial news feeds app in Microsoft from January 23 to April 01, 2020 (13 weeks). We randomly sample 100,000 and 10,000 impressions from the first ten weeks to construct training and validation set, and 100,000 impressions from the last three weeks to construct test data. Since *Feeds* only contains topic label of news, we implement a simplified version of *HieRec* with only user- and topic-level interest representations on *Feeds*. Besides, following Wu et al. (2020d), users in *Feeds* were anonymized via hash algorithms and de-linked from the production system to protect user privacy. Detailed information is summarized in Table 1.

Next, we introduce experimental settings and hyper-parameters of *HieRec*. We use the first 30 words and 5 entities of news titles and users’ recent 50 clicked news in experiments. We adopt pre-trained glove (Pennington et al., 2014) word embeddings and TransE entity embeddings (Bordes et al., 2013) for initialization. In *HieRec*, the word and entity self-attention network output 400- and 100-dimensional vectors, respectively. Besides, the unified news representation is 400-dimensional. Attention networks (i.e., $\phi_s(\cdot)$, $\phi_t(\cdot)$, and $\phi_g(\cdot)$) are implemented by single-layer dense networks. Besides, dimensions of topic and subtopic embeddings are 400, both of which are randomly initialized and fine-tuned. The hyper-parameters for combining different interest scores, i.e. λ_t and λ_s , are set to 0.15 and 0.7 respectively. Moreover, we utilize dropout technique (Srivastava et al., 2014) and Adam optimizer (Kingma and Ba, 2015) for training. *HieRec* is trained for 5 epochs with 0.0001

¹We use the small version of MIND for quick experiments. This dataset is at <https://msnews.github.io/index.html>

	<i>MIND</i>				<i>Feeds</i>			
	AUC	MRR	nDCG@5	nDCG@10	AUC	MRR	nDCG@5	nDCG@10
EBNR	61.62±0.15	28.07±0.18	30.55±0.22	37.07±0.21	63.48±0.32	28.01±0.18	32.05±0.23	37.64±0.22
DKN	63.99±0.23	28.95±0.08	31.73±0.14	38.38±0.17	62.94±0.22	28.05±0.26	32.15±0.34	37.68±0.36
DAN	64.68±0.13	29.78±0.13	32.63±0.21	39.27±0.15	62.67±0.49	27.75±0.34	31.74±0.44	37.42±0.43
NAML	64.30±0.30	29.81±0.17	32.64±0.24	39.11±0.20	64.48±0.24	28.99±0.13	33.37±0.16	38.90±0.18
NPA	64.28±0.53	29.64±0.33	32.28±0.37	38.93±0.39	64.02±0.63	28.71±0.39	33.01±0.50	38.55±0.47
LSTUR	65.68±0.35	30.44±0.39	33.49±0.45	39.95±0.39	65.01±0.13	29.28±0.06	33.74±0.09	39.16±0.11
NRMS	65.43±0.15	30.74±0.18	33.13±0.17	39.66±0.15	65.27±0.19	29.40±0.15	33.89±0.16	39.34±0.15
KRED	65.89±0.31	30.80±0.32	33.78±0.27	40.23±0.26	65.51±0.11	29.57±0.06	34.04±0.06	39.60±0.05
GNewsRec	65.91±0.21	30.50±0.21	33.56±0.21	40.13±0.18	65.23±0.16	29.36±0.11	33.87±0.13	39.44±0.12
FIM	64.65±0.14	29.70±0.17	32.51±0.25	39.30±0.16	65.41±0.23	29.57±0.18	34.08±0.25	39.56±0.23
HieRec	67.95±0.14	32.87±0.08	36.36±0.07	42.53±0.10	66.23±0.10	29.82±0.11	34.42±0.13	39.94±0.13

Table 2: Performance of different methods. The improvement of *HieRec* over the best baseline method is significant at level $p < 0.01$ based on t-test.

learning rate. All hyper-parameters of *HieRec* and baseline methods are manually tuned on the validation set.² Following Wu et al. (2019e), we use four ranking metrics, i.e., AUC, MRR, nDCG@5, and nDCG@10, for performance evaluation.

4.2 Main Results

We first introduce the baseline methods we compared in experiments: (1) *EBNR* (Okura et al., 2017): learning user representations from the sequence user’s clicked news via a GRU network. (2) *DKN* (Wang et al., 2018): using a candidate-aware attention network to learn user representations. (3) *DAN* (Zhu et al., 2019): using an attentive LSTM network to learn user representations. (4) *NAML* (Wu et al., 2019a): learning user representations by attentively aggregating user’s clicked news. (5) *NPA* (Wu et al., 2019b): learning news and user representations via personalized attention networks. (6) *LSTUR* (An et al., 2019): modeling short-term user interests from user’s clicked news via a GRU network and long-term user interests from user-news interactions via user ID embeddings. (7) *NRMS* (Wu et al., 2019e): applying multi-head self-attention networks to learn news representations and user representations. (8) *KRED* (Liu et al., 2020): proposing a knowledge graph attention network to learn news representations from texts and entities of news titles. (9) *GNewsRec* (Hu et al., 2020): modeling short-term user interests from clicked news sequences via an attentive GRU network and long-term user interests from user-news click graph via a graph neural network. (10) *FIM* (Wang et al., 2020): modeling user interests in candidate news from semantic relevance of user’s clicked news and candidate news

via a 3-D CNN network.

Each experiment is repeated 5 times. The average results and standard deviations are listed in Table 2, from which we have several observations. First, *HieRec* significantly outperforms other baseline methods which learn a single user embedding to model overall user interests, such as *NRMS*, *NPA*, and *NAML*. This is because user interests are usually diverse and multi-grained. However, it is difficult for a single representation vector to model user interests in different aspects and granularities, which may be suboptimal for personalized news recommendation. Different from these methods, we propose a hierarchical user interest modeling framework, which can represent diverse and multi-grained user interests via a three-level hierarchy. Besides, we also propose a hierarchical user interest matching framework to match user interest with candidate news from different granularities, which can better target user interests. Second, *HieRec* can significantly outperform *FIM*, which directly model user interests in candidate news from the semantic relevance of candidate news and user’s clicked news. This may be because *FIM* did not consider user interests from different granularities for matching candidate news.

4.3 Effectiveness in User Modeling

To fairly compare different methods with *HieRec* on the performance of interest modeling, we compare them based on the same news modeling method (the news modeling method introduced in Section 3.4). Experimental results are summarized in Table 3 and we only show experimental results on *MIND* in the following sections. Table 3 shows that *HieRec* significantly outperforms existing interest modeling methods. This is because user in-

²<https://github.com/JulySinceAndrew/HieRec>

	AUC	MRR	nDCG@5	nDCG@10
NAML	65.81±0.27	30.89±0.21	34.16±0.30	40.55±0.24
DKN	66.03±0.27	31.17±0.25	34.47±0.33	40.85±0.29
EBNR	65.90±0.27	30.86±0.21	34.14±0.30	40.58±0.24
LSTUR	66.02±0.14	31.16±0.15	34.37±0.15	40.83±0.12
GNewsRec	66.16±0.14	31.19±0.05	34.40±0.09	40.82±0.10
NRMS	66.04±0.21	31.20±0.19	34.53±0.22	40.89±0.18
HieRec	67.95±0.14	32.87±0.08	36.36±0.07	42.53±0.10

Table 3: Effect of *HieRec* in user interest modeling.

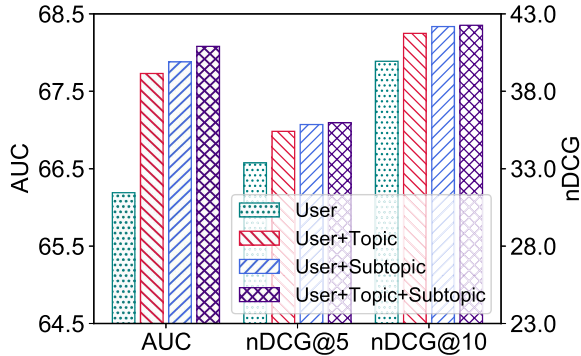


Figure 5: Effect of hierarchical user interest modeling.

terests are usually diverse and multi-grained. It is difficult for existing methods with single user embedding to capture user interests in different aspects and granularities. Different from these methods, *HieRec* learns a three-level hierarchy to represent diverse and multi-grained user interests.

4.4 Ablation Study

We evaluate the effectiveness of user interest representations of different levels by removing the corresponding interest matching scores from Eq. 4. Results are shown in Fig. 5 and we have several findings. First, *HieRec* with user- and topic- or subtopic-level interest representation significantly outperforms *HieRec* with only user-level interest representation. This is because matching candidate news with fine-grained user interests has the potential to improve the accuracy of news recommendation. Topic- and subtopic-level interest representation can model finer-grained user interests than the user-level interest representation. Thus, they can provide additional information to match candidate news than user-level interest representation. Second, *HieRec* with interest representations of three levels also outperforms *HieRec* with user- and topic- or subtopic-level interest representation. This may be because matching candidate news with user interests of different granularities can help perform more accurate interest matching. Since topic- and subtopic-level interest representa-

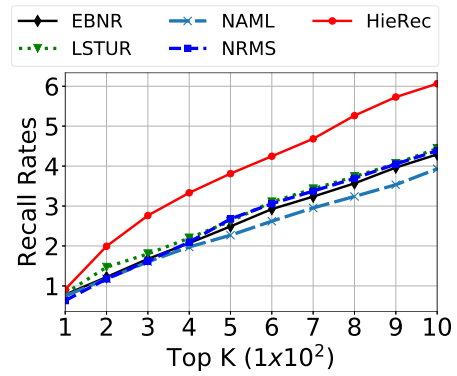


Figure 6: Recall rates of different methods.

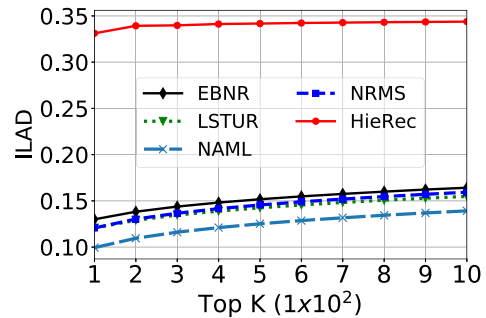


Figure 7: Performance of different methods on recommendation diversity.

tion capture user interests at different granularities, incorporating both of them can further improve the recommendation performance.

4.5 Performance on Recall and Diversity

Next, we compare different user interest modeling methods on the news recall task.³ Since methods that model user interests with candidate news information, e.g., *DKN* and *GNewsRec*, cannot be applied in the news recall task due to efficiency issues (Pal et al., 2020), we do not compare them in experiments. We evaluate the accuracy and diversity of top K recalled candidate news. Following existing works (Pal et al., 2020; Chen et al., 2018), the former is measured by recall rates, and the latter is measured by intra-list average distance (ILAD). For *HieRec*, we employ subtopic-level interest representations to perform multi-channel news recall and equally integrate news recalled by different interest channels. Experimental results are summarized in Fig. 6 and Fig. 7, which show that *HieRec* significantly outperforms other methods in terms of both recall rates and diversity. This is because user interests are usually very diverse and multi-

³News recall task aims to recall a small number of candidate news from a large news pool according to user interests.

User's clicked news		Top 5 recommend news by HieRec		Top 5 recommend news by GNewsRec	
Topic	Title	Topic	Title	Topic	Title
Sports	MLB free-agent predictions: Rankings, rumors for the top 10 players on the market.	Politics	Tense TV: CNN's Blitzer asks Kellyanne Conway about husband	Politics	Tense TV: CNN's Blitzer asks Kellyanne Conway about husband
Sports	MLB free-agent rumors: Braves making Madison Bumgarner a priority, Report Says	Sports	Sign-stealing penalties could be "unlike anything seen in the sport's recent history"	Crime	A 16-second spasm of violence leaves 2 dead at Saugus High School
Politics	Ocasio-Cortez apologizes for blocking critic on twitter	Lifestyle	Family told to take down Christmas display because it's too soon to decorate	Politics	Judge agrees Alabama Islamic state recruit is not us citizen
Lifestyle	Woman finds a giant wild snake just chilling in her bed	Crime	A 16-second spasm of violence leaves 2 dead at Saugus High School	Politics	Student president faces impeachment for Trump JR.'s \$ 50k campus talk
Crime	Six deaths in one family over 14 years. Police say they have found their killer	Politics	Judge agrees Alabama Islamic state recruit is not us citizen	Politics	Trump admin preparing to take over private land for border wall

Figure 8: Recommendation results of *HieRec* and *GNewsRec* for the same impression. The news clicked by the user in this impression is in blue and bold.

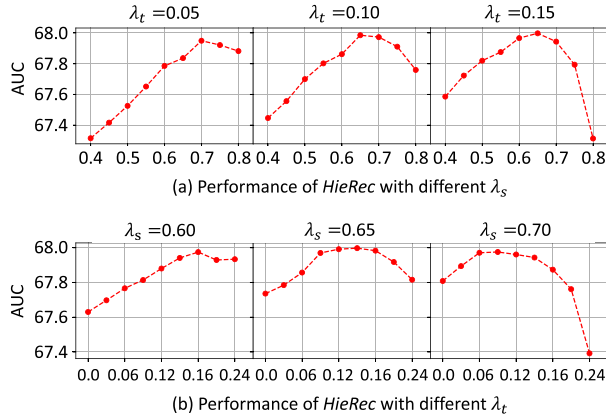


Figure 9: Influence of hyper-parameters.

grained, which are difficult to be comprehensively modeled by a single representation vector. Different from these methods, *HieRec* hierarchically represents user interests and can better model user interests in different aspects and granularities. Besides, this also implies that compared to existing personalized methods, *HieRec* can help users explore more diverse information and alleviate filter bubble issues (Nguyen et al., 2014) to some extent.

4.6 Hyper-parameters Analysis

As shown in Fig. 9, we analyze the influence of two important hyper-parameters of *HieRec* (i.e., λ_t , λ_s) used for combining different levels of interest scores. First, when λ_t is fixed, performance of *HieRec* first gets better with the increase of λ_s . This is because λ_s controls the importance of o_s . Besides, o_s measures the relevance of candidate news and fine-grained user interests, which can provide accurate information to understand user interests in the candidate news. When λ_s is too small, *HieRec* cannot effectively exploit information in o_s . Second, large value of λ_s also hurts the performance of *HieRec*. This is because when λ_s is too large, *HieRec* cannot effectively exploit user- and topic-level matching scores to recommend can-

didate news. However, matching candidate news with both overall and coarse-grained user interests is important for personalized news recommendation. Thus, a moderate λ_s , i.e., 0.65 or 0.7, is suitable for *HieRec*. Third, when λ_s is fixed, the performance of *HieRec* also first gets better with the increase of λ_t and gets worse when λ_t is too large. This is because *HieRec* cannot effectively utilize information of o_t when λ_t is too small. Besides, *HieRec* cannot effectively utilize information of o_g and o_s when λ_t is too large. Thus, a moderate λ_t , i.e., 0.12 or 0.15, is suitable for *HieRec*.

4.7 Case Study

We conduct a case study to show the superior performance of *HieRec*. We compare *HieRec* with *GNewsRec* since *GNewsRec* achieves best AUC score in Table 2 among baseline methods. In Fig. 8, we show the top 5 news recommended by *HieRec* and *GNewsRec* in a randomly sampled impression. Besides, we also show the historical clicks of the target user in this impression. We can find that the top 5 news recommended by *GNewsRec* is dominated by news on politics, which cannot comprehensively cover different user interests. This is because user interests are usually diverse and multi-grained. However, it is difficult for *GNewsRec*, which learns a single representation to model overall user interests, to effectively capture user interests in different aspects and granularities. Different from *GNewsRec*, the top 5 news recommended by *HieRec* are diverse and can cover topics that the user may be interested in. Besides, the user clicked a news recommended by *HieRec*. This is because *HieRec* learns a hierarchical user interest representation which can effectively model user interests in different aspects and granularities. With the help of the hierarchical user interest representation, *HieRec* can match candidate news with user interests in different aspects and granularities.

5 Conclusion

In this paper, we propose a personalized news recommendation method named *HieRec* for hierarchical user interest modeling, which can effectively model diverse and multi-grained user interests. *HieRec* learns a three-level hierarchy to represent user interest in different aspects and granularity. First, we learn multiple subtopic-level interest representations to model fine-grained user interests in different news subtopics. Second, we learn multiple topic-level interest representations to model coarse-grained user interests in several major news topics. Third, we learn a user-level interest representation to model overall user interests. Besides, we propose a hierarchical user interest matching framework to match candidate news with user interest from different granularity for more accurate user interest targeting. Extensive experiments on two real-world datasets show the effectiveness of *HieRec* in user interest modeling.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant numbers U1936208, U1705261, U1936216, and U1836204. We thank Tao Di and Wei He for their great comments and suggestions.

Ethics and Impact Statement

In this paper, we present *HieRec* to model diverse and multi-grained user interest. *HieRec* can be applied to online news platforms for personalized news recommendation, which can help platforms improve user experience and help users find interested news information. Although *HieRec* can bring many benefits, it may also have several potential risks, which we will discuss in detail.

Accuracy Although *HieRec* outperforms baseline methods in term of recommendation accuracy (Table 2), it may also have some inaccurate recommendation results that users are not interested in. Users usually just ignore them and will not click them to read. The user experience may be harmed and users may use the online news service less in the future, or turn to other online news platforms.

Privacy In *HieRec*, we rely on user behavior data centrally stored on the news platform for model training and online services. User behavior data is usually privacy-sensitive, and its centralized storage may lead to privacy concerns and risks. In

the future, we will explore to train and deploy *HieRec* in a more privacy-preserving way based on some effective privacy protection techniques like Federated Learning (Qi et al., 2020).

Diversity Filter bubbles and echo chambers are the common problem for many recommender systems (Nguyen et al., 2014), which harms user experience. Improving recommendation diversity has the potential to alleviate the problem of filter bubbles and echo chambers. Through experiments in Fig. 7, we find that *HieRec* can outperform many news recommendation methods in term of recommendation diversity. Thus, compared with existing methods, *HieRec* has the potential to alleviate filter bubble problem to some extent. Besides, in order to further improve recommendation diversity, *HieRec* can be combined with some existing methods in this field like DPP (Chen et al., 2018).

Fake News and Clickbait There may be some fake news and clickbait in some online platforms. In order to handle the negative social impact and the user experience harm brought by these fake news and clickbait, online news platforms can use some existing fake news detection and clickbait detection techniques such as (Kumar et al., 2018; Shu et al., 2019) to filter these kinds of news before applying *HieRec* for personalized recommendation.

Fairness Like many other recommender systems, *HieRec* relies on user behavior data for model training and online service. The bias in user behavior data may lead to some specific groups of users not be able to receive news information with sufficient accuracy and diversity, and the recommendation results may be more suitable for some major populations. Recently, some fairness-aware recommendation methods like FairRec (Wu et al., 2021) have been proposed to eliminate bias and unfairness in recommender systems. We can combine *HieRec* with these methods to improve the fairness of the recommendation results and mitigate the harms for marginalized populations.

Misuse The proposed *HieRec* method works in a data-driven way. It trains the model from the user logs and makes personalized recommendations to users based on their interest inferred from their clicked news. However, in some extreme cases, the recommendation results may be maliciously manipulated to influence users. To avoid the potential misuse, the usage of *HieRec* should comply with the regulations and laws, and intentional manipulation should be prohibited.

References

- 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*, pages 336–345.
- Trapit Bansal, Mrinal Das, and Chiranjib Bhattacharyya. 2015. Content driven user profiling for comment-worthy recommendations of news and blog articles. In *RecSys.*, pages 195–202.
- Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *NIPS*, pages 2787–2795.
- Laming Chen, Guoxin Zhang, and Eric Zhou. 2018. Fast greedy map inference for determinantal point process to improve recommendation diversity. In *NIPS*, pages 5622–5633.
- Abhinandan S Das, Mayur Datar, Ashutosh Garg, and Shyam Rajaram. 2007. Google news personalization: scalable online collaborative filtering. In *WWW*, pages 271–280.
- Suyu Ge, Chuhan Wu, Fangzhao Wu, Tao Qi, and Yongfeng Huang. 2020. Graph enhanced representation learning for news recommendation. In *WWW*, pages 2863–2869.
- 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*, page 102142.
- 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*, pages 1855–1858.
- Diederik P Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *ICLR*.
- Vaibhav Kumar, Dhruv Khattar, Siddhartha Gairola, Yash Kumar Lal, and Vasudeva Varma. 2018. Identifying clickbait: A multi-strategy approach using neural networks. In *SIGIR*, pages 1225–1228.
- Chen Lin, Runquan Xie, Xinjun Guan, Lei Li, and Tao Li. 2014. Personalized news recommendation via implicit social experts. *Information Sciences*, pages 1–18.
- 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.*, pages 200–209.
- Jiahui Liu, Peter Dolan, and Elin Rønby Pedersen. 2010. Personalized news recommendation based on click behavior. In *IUI*, pages 31–40.
- Tien T Nguyen, Pik-Mai Hui, F Maxwell Harper, Loren Terveen, and Joseph A Konstan. 2014. Exploring the filter bubble: the effect of using recommender systems on content diversity. In *WWW*, pages 677–686.
- Shumpei Okura, Yukihiro Tagami, Shingo Ono, and Akira Tajima. 2017. Embedding-based news recommendation for millions of users. In *KDD*, pages 1933–1942.
- Aditya Pal, Chantat Eksombatchai, Yitong Zhou, Bo Zhao, Charles Rosenberg, and Jure Leskovec. 2020. Pinnersage: Multi-modal user embedding framework for recommendations at pinterest. In *KDD*, pages 2311–2320.
- Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *EMNLP*, pages 1532–1543.
- Tao Qi, Fangzhao Wu, Chuhan Wu, Yongfeng Huang, and Xing Xie. 2020. Privacy-preserving news recommendation model learning. In *EMNLP: Findings*, pages 1423–1432.
- Kai Shu, Limeng Cui, Suhang Wang, Dongwon Lee, and Huan Liu. 2019. defend: Explainable fake news detection. In *KDD*, pages 395–405.
- Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: A simple way to prevent neural networks from overfitting. *JMLR*, pages 1929–1958.
- 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*, pages 6000–6010.
- Heyuan Wang, Fangzhao Wu, Zheng Liu, and Xing Xie. 2020. Fine-grained interest matching for neural news recommendation. In *ACL*, pages 836–845.
- Hongwei Wang, Fuzheng Zhang, Xing Xie, and Minyi Guo. 2018. Dkn: Deep knowledge-aware network for news recommendation. In *WWW*, pages 1835–1844.
- Chuhan Wu, Fangzhao Wu, Mingxiao An, Jianqiang Huang, Yongfeng Huang, and Xing Xie. 2019a. Neural news recommendation with attentive multi-view learning. *IJCAI*, pages 3863–3869.
- Chuhan Wu, Fangzhao Wu, Mingxiao An, Jianqiang Huang, Yongfeng Huang, and Xing Xie. 2019b. Npa: Neural news recommendation with personalized attention. In *KDD*, pages 2576–2584.
- Chuhan Wu, Fangzhao Wu, Mingxiao An, Yongfeng Huang, and Xing Xie. 2019c. Neural news recommendation with topic-aware news representation. In *ACL*, pages 1154–1159.

- Chuhan Wu, Fangzhao Wu, Mingxiao An, Tao Qi, Jianqiang Huang, Yongfeng Huang, and Xing Xie. 2019d. Neural news recommendation with heterogeneous user behavior. In *EMNLP*, pages 4876–4885.
- Chuhan Wu, Fangzhao Wu, Suyu Ge, Tao Qi, Yongfeng Huang, and Xing Xie. 2019e. Neural news recommendation with multi-head self-attention. In *EMNLP*, pages 6390–6395.
- Chuhan Wu, Fangzhao Wu, Tao Qi, and Yongfeng Huang. 2020a. Sentirec: Sentiment diversity-aware neural news recommendation. In *AAACL*, pages 44–53.
- Chuhan Wu, Fangzhao Wu, Tao Qi, and Yongfeng Huang. 2020b. User modeling with click preference and reading satisfaction for news recommendation. In *IJCAI*, pages 3023–3029.
- Chuhan Wu, Fangzhao Wu, Tao Qi, Jianxun Lian, Yongfeng Huang, and Xing Xie. 2020c. Ptum: Pre-training user model from unlabeled user behaviors via self-supervision. In *EMNLP: Findings*, pages 1939–1944.
- Chuhan Wu, Fangzhao Wu, Xiting Wang, Yongfeng Huang, and Xing Xie. 2021. Fairrec: fairness-aware news recommendation with decomposed adversarial learning. In *AAAI*.
- Fangzhao Wu, Ying Qiao, Jiun-Hung Chen, Chuhan Wu, Tao Qi, Jianxun Lian, Danyang Liu, Xing Xie, Jianfeng Gao, Winnie Wu, et al. 2020d. Mind: A large-scale dataset for news recommendation. In *ACL*, pages 3597–3606.
- 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*, pages 167–176.
- Qiannan Zhu, Xiaofei Zhou, Zeliang Song, Jianlong Tan, and Guo Li. 2019. Dan: Deep attention neural network for news recommendation. In *AAAI*, pages 5973–5980.