Multi-Interest Extraction Joint with Contrastive Learning for News Recommendation

Shicheng Wang[1,2], Shu Guo[3] ⋆(✉), Lihong Wang[3], Tingwen Liu[1,2], and Hongbo Xu[1,2]

[1] Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
[2] School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China
wangshicheng@iie.ac.cn, liutingwen@iie.ac.cn, hbxu@iie.ac.cn
[3] National Computer Network Emergency Response Technical Team/Coordination Center of China, Beijing, China
guoshu@cert.org.cn, wlh@isc.org.cn

Abstract. News recommendation techniques aim to recommend candidate news to target user that he may be interested in, according to his browsed historical news. At present, existing works usually tend to represent user reading interest using a single vector during the modeling procedure. Actually, it is obviously that users usually have multiple and diverse interest in reality, such as sports, entertainment and so on. Thus it is irrational to represent user sophisticated semantic interest simply utilizing a single vector, which may conceal fine-grained information. In this work, we propose a novel method combining multi-interest extraction with contrastive learning, named MIECL, to tackle the above problem. Specifically, first, we construct several interest prototypes and design a multi-interest user encoder to learn multiple user representations under different interest conditions simultaneously. Then we adopt a graph-enhanced user encoder to enrich user corresponding semantic representation under each interest background through aggregating relevant information from neighbors. Finally, we contrast user multi-interest representations and interest prototype vectors to optimize the user representations themselves, in order to promote dissimilar semantic interest away from each other. We conduct experiments on two real-world news recommendation datasets MIND-Large, MIND-Small and empirical results demonstrate the effectiveness of our approach from multiple perspectives.

Keywords: news recommendation · multi-interest extraction · contrastive learning.

1 Introduction

Recently, with the rapid development of Internet and online news services [3, 8], massive news are released on various news platforms all the time and can make

⋆ Corresponding Author

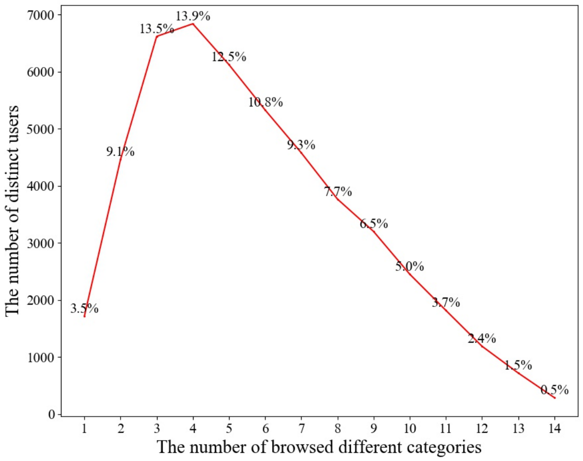

Fig. 1. The number of distinct users with specific number of browsed categories.

users overwhelmed. Thus, personalized news recommendation is necessary for news platforms to help users alleviate information overload as well as improve their reading experience. Briefly, news recommendation aims to recommend candidate news to target user that he may be interested in.

Traditional collaborative filtering methods [3, 17] reconstruct interactive matrix to learn user and news representations. However, they suffer from severe cold-start problem due to the short life cycles characteristic of news. Then content-based methods are designed to represent news semantic information and mine user reading interests accurately. Therefore, they share a general framework, including news encoder, user encoder and click predictor. In the light of this framework, existing methods usually adopt BERT [4] or Transformer [14] to learn news representations based on text content such as news titles [1, 21]. User interest modeling is another significant procedure in such framework. As deep learning methods are widely concerned, Recurrent Neural Network and Transformer, regarded as effective methods for dealing with sequence modeling, are adopted to model user interests [12, 19, 20]. In addition, with the huge influence of graph neural networks techniques, some works introduce GAT [15] into user modeling process to enhance user representations by leveraging neighborhood information [5, 7]. However, above methods usually represent overall user interests by a single vector based on his/her browsing news history.

Actually, users usually have multiple and diverse interests. For ease of explanation, we utilize category information to represent different interests. We conduct statistic analysis on a real-world dataset MIND-Small [23]. We count the number of distinct users respectively grouped by the number of browsed different categories. As shown in Figure 1, it is obviously that massive users have browsed multiple categories of news. For example, users may be interested in sports, movies, and finance according to historical news. Nevertheless, previous works usually tend to represent user reading interests using a single vector, resulting in the integration of semantic information related to different interests. As a consequence, such approaches may conceal fine-grained information in user modeling and reduce the variousness of recommending.

In this work we propose a novel method combining Multi-Interest Extraction with Contrastive Learning, named MIECL, to tackle the above problem. Our method mainly innovates in the process of user modeling. Specifically, given a target user with his browsed history, we first construct several interest prototypes and design a multi-interest user encoder to simultaneously learn multiple user representations under each prototype. Through applying attention mechanisms between user historical news and interest prototypes, we are able to model diverse user interests in a fine-grained way. Then we adopt a graph-enhanced user encoder to enrich user corresponding semantic representations and capture their potential interests under each prototype, through aggregating relevant information from neighbors. Next, inspired by the ideology of contrastive learning [2], we optimize the user multi-interest representations through contrasting representations themselves and interest prototypes, in order to promote the multiple representations to be differentiated. Finally we aggregate user multi-interest representations adaptively and calculate click probability between target user and candidate news.

The major contributions of this paper include:

- We design a multi-interest user encoder to explore diverse and multiple user interests in a fine-grained way for more accurate user interest modeling.
- We utilize the superiority of contrastive learning to optimize the above user multi-interest representations, making them more differentiated.
- We conduct experiments on real-world datasets MIND-Large and MIND-Small. The empirical results demonstrate the effectiveness of our approach.

2 Related Work

2.1 Personalized News Recommendation

News recommendation has attracted more and more attention recently with the growth of individual and social needs. Therefore, a variety of methods have been proposed, including collaborative filtering based methods and content based methods. Most traditional methods achieved news recommendation based on collaborative filtering framework [3, 17]. They parameterized users and items in a latent space and aimed at reconstructing historical interactions. However, due to the short life cycles characteristic of news articles, CF methods based on IDs always suffered from severe cold start problem, which required us to understand news contents and user interests.

To address this issue, content-based or hybrid methods have been proposed. For example, Okura [10] utilized an auto-encoder to learn news representations from news bodies. Then they applied a GRU network to model user interests from clicked historical news sequences. Finally, they calculated click probability between user interests and news representations based on dot product. NAML [19] leveraged a CNN network to model news semantic representations from news titles and categories. Then they learnt user representations through attentively aggregating clicked news. Similarly, LSTUR [1] also learnt news representations based on CNN network. However, unlike NAML method, they applied GRU neural network to model user short-term interests from clicked history and further model user long-term interests via user ID embeddings. Then they adopted attention mechanisms to integrate the above user representations. NRMS [21] utilized multi-head self-attention networks with similar structures to learn news representations and user representations separately, in order to capture interaction information in word sequences and news sequences respectively. With the development of graph neural network technology, some work proposed to model news contents and user interests based on GCN or GAT. KRED [9] first introduced news titles and entities to construct a news recommendation knowledge graph. Then they applied graph attention network to learn news representations. GNewsRec [7] learnt user short-term interests by applying attentive GRU neural network to clicked history and user long-term interests via graph neural networks. However, there is few work recognizing the importance of modeling diverse user interests explicitly. The above-mentioned methods usually tend to represent user interests using a single vector. Different from these methods, in this paper, we propose to model diverse user interests explicitly in a fine-grained way.

2.2 Contrastive Learning

Contrastive learning, as a branch of self-supervised learning, is devised to learn by comparing among different input samples. The objective of contrastive learning is to map the representations of similar samples close together, while that of dissimilar samples should be further away in the embedding space. According to the scale of the samples involved in the comparison, recent contrastive learning methods can be formulated as global-local contrast and local-local contrast. The global-local contrastive learning focuses on modeling the belonging relationship between the local feature of a sample and its global context representation. For example, Deep Infomax [6] proposed to maximize the mutual information between a local patch and its global context, which provided us with a new paradigm. Deep Graph Infomax [16] introduced the ideology of DIM into graph representation learning, which regarded target node representation as local feature and its high-level summary of graph as global feature.

However, the global-local contrastive learning may generate ill-conditioned representations. Recently local-local contrastive learning discards mutual information and directly studies the relationships between different samples instance-level local representations. For example, CMC [13] proposed to adopt multiple different views of an image as positive samples and sampled another irrelevant image as the negative. SimCLR [2] illustrated the importance of introducing

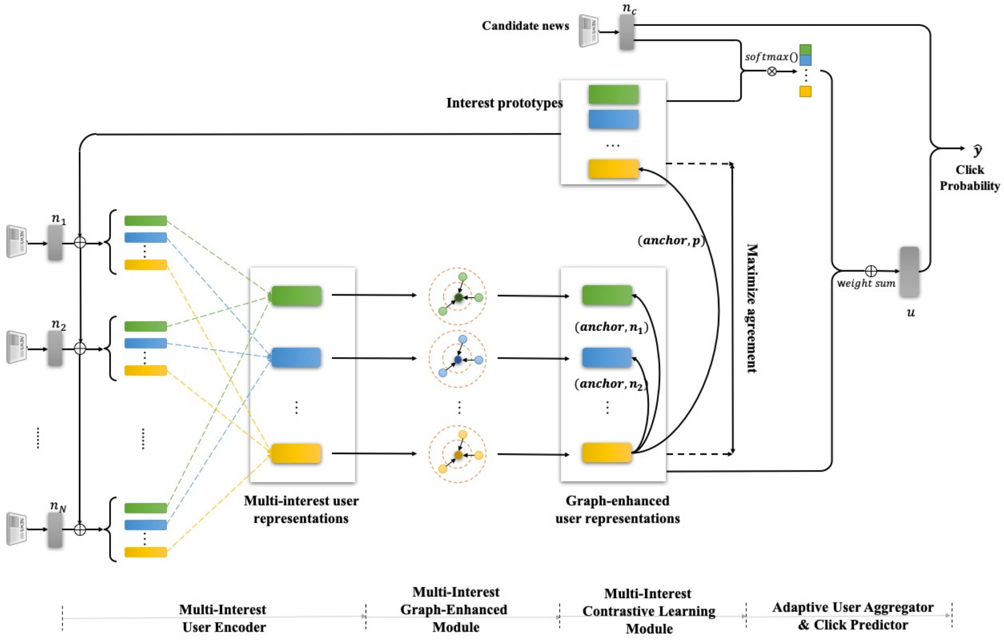
Candidate news
n_c
$softmax()$
Interest prototypes
$\hat{y}$
Click Probability
n_1
n_2
n_N
$(anchor, p)$
Maximize agreement
$(anchor, n_1)$
$(anchor, n_2)$
weight sum
u
Multi-interest user representations
Graph-enhanced user representations
Multi-Interest User Encoder
Multi-Interest Graph-Enhanced Module
Multi-Interest Contrastive Learning Module
Adaptive User Aggregator & Click Predictor

Fig. 2. Overall framework of our MIECL method.

data augmentation operations for contrastive representation learning based on a simple framework. Inspired by the recent prominent advances in local-local contrast methods, in this work we explore contrastive learning to assist us in differentiating diverse user interests.

3 Methodology

In this section, we first present the problem formulation of personalized news recommendation. Specifically, given a candidate news n_c and a target user u with his clicked news history $[n_1, \cdots, n_N]$, we aim to learn candidate news representation as well as user representation respectively, followed by calculating the relevance score between their representations. Finally we decide whether to recommend n_c or not according to the score. Then we propose a method combining multi-interest extraction with contrastive learning, named MIECL, to model diverse user interests in a fine-gained way. As shown in Figure 2, our model MIECL is mainly innovative in user modeling, including three modules: multi-interest user encoder, multi-interest graph-enhanced module and multi-interest contrastive learning module. We will elaborate our method in the subsequent sections in detail [4].

[4] Our source code is available at https://github.com/wangsc2113/MIECL.

3.1 News Encoder

In this section, we introduce how to learn news semantic representations from news titles. We indicate the title word sequence as $[w_1, \cdots, w_M]$, where w_i is denoted as the i-th word in title. Here we encode titles based on the traditional Transformer framework.

First, at the bottom of news encoder module, it applies word embedding layer to convert each word w_i into corresponding vector e_i. Then, it adopts a multi-head self-attention network to capture semantic interactions between title words $[e_1, \cdots, e_M]$. The representation of the word w_i learned by the s-th attention head h_i^s is calculated as:

$$h_i^s = V_s^w \sum_{j=1}^{M} \alpha_{i,j}^s \cdot e_j, \quad \alpha_{i,j}^s = \frac{\exp(e_i^T Q_s^w e_j)}{\sum_{t=1}^{M} \exp(e_i^T Q_s^w e_t)} \tag{1}$$

where $V_s^w \in \mathbb{R}^{d/S \times d}$ and $Q_s^w \in \mathbb{R}^{d \times d}$ are the projection parameters in the s-th self-attention head, and $\alpha_{i,j}$ indicates the interaction score between the word w_i and w_j. Then the multi-head representation of word w_i is concatenated as $h_i \in \mathbb{R}^d$, i.e., $h_i = [h_i^1; h_i^2; \cdots; h_i^S]$, where S denotes the number of separate self-attention heads.

Finally, it applies an additive attention network to aggregate contextual word representations into a news representation n, formulated as:

$$n = \sum_{i=1}^{M} \alpha_i^w \cdot h_i, \quad \alpha_i^w = \frac{\exp(q_\alpha^T \cdot h_i)}{\sum_{j=1}^{M} \exp(q_\alpha^T \cdot h_j)} \tag{2}$$

where $q_\alpha \in \mathbb{R}^d$ is a projection parameter vector.

3.2 Multi-Interest User Encoder

We argue that one representation vector has a low potential to reflect user diverse and complex reading interests. Obviously, an intuitive solution is to learn multiple representations to model diverse user interests.

At the beginning, since there is semantic relatedness between news articles browsed by the same user, it applies a multi-head self-attention to enhance the news representations by capturing their relatedness. Given clicked news history $[n_1, \cdots, n_N]$, the news representation learned by the s-th attention head n_i^k is calculated as:

$$n_i^s = V_s^n \sum_{j=1}^{N} \beta_{i,j}^s \cdot n_j, \quad \beta_{i,j}^s = \frac{\exp(n_i^T Q_s^n n_j)}{\sum_{t=1}^{N} \exp(n_i^T Q_s^n n_t)} \tag{3}$$

where $V_s^n \in \mathbb{R}^{d/S \times d}$ and $Q_s^n \in \mathbb{R}^{d \times d}$ are the news-level projection parameters in the s-th self-attention head. Then the multi-head representation of news n_i is defined as n_i, i.e., $n_i = [n_i^1; n_i^2; \cdots; n_i^S]$.

We assume that each user possesses K different interests and each interest has a corresponding prototype vector $I_k \in \mathbb{R}^d$. Interest prototypes are introduced instead of the "hard" category information, since we deem that news is semantic related to the category, even if there is no subordinate relationship between them. Though similar to the previous content-based user modeling methods but in a fine-grained way, we design to aggregate specific interest-relevant information from historical news to acquire corresponding interest-level user representation.

First, it extracts semantic information n_i^k from history news n_i oriented to specific interest I_k. n_i^k is expected to contain specific interest-relevant information. It simply adopts concatenation here, which is formulated as:

$$n_i^k = W(n_i || I_k), \tag{4}$$

where $W \in \mathbb{R}^{d \times 2d}$ is a learnable parametric matrix for semantic transforming. In this way, for each news in clicked history, we can obtain relevant semantic information under given interest condition.

Next, under different interest conditions, it adopts additive attention networks to generate interest-level user representations $u_k \in \mathbb{R}^d, k \in [1, K]$ through aggregating interest-relevant information respectively, according to user historical news. The formula is defined as:

$$u_k = \sum_{i=1}^{N} \beta_{i,k}^n \cdot n_i^k, \quad \beta_{i,k}^n = \frac{\exp(q_\beta^T \cdot n_i^k)}{\sum_{j=1}^{N} \exp(q_\beta^T \cdot n_j^k)} \tag{5}$$

where $q_\beta \in \mathbb{R}^d$ is a projection vector and $\beta_{i,k}^n$ denotes the attention weight of the i-th clicked news contributing to user representation in condition of k-th interest prototype. Now we obtain K interest-level user representations $[u_1, \cdots, u_K]$ to reveal his/her diverse interests.

3.3 Multi-Interest Graph-Enhanced Module

In this section, we introduce graph attention neural networks to enrich user interest-level semantic representations through utilizing neighbor users information. For the given target user u, we search his/her second-order neighbor users based on user-news-user co-occurrence relation in advance. For the convenience of calculation, we randomly choose T neighbors for each user. For each neighbor user u^i, we utilize the Equation (7) to obtain their semantic representations under different prototypes, noted as $[u_1^1, \cdots, u_K^1]$, $\cdots$, $[u_1^T, \cdots, u_K^T]$ respectively. Then under each interest condition, it adopts a separate graph attention network to aggregate interest-specific neighbor information:

$$u_k = \sum_{i=1}^{T \cup \{u\}} \gamma_k^i \cdot u_k^i, \quad \gamma_k^i = \frac{\exp(q_k^\gamma \cdot [u_k || u_k^i])}{\sum_{j=1}^{T \cup \{u\}} \exp(q_k^\gamma \cdot [u_k || u_k^j])} \tag{6}$$

where $q_k^\gamma \in \mathbb{R}^{1 \times 2d}$ is weight matrix and γ_k^i denotes the attention weight between the target user and his i-th neighbor user in condition of k-th interest prototype. Now we obtain the graph-enhanced user multi-interest representations.

3.4 Multi-Interest Contrastive Learning Module

In the previous sections we already obtain user multi-interest representations. However, there might be information redundancy among these vectors since we do not impose any constraints on them. This may have a harmful effect on model performance since similar interest-level user representations are difficult to reflect diverse interests.

Inspired by the local-local contrast methods recently [2, 13], we design a joint learning paradigm and construct contrastive learning objective to optimize above-mentioned user multi-interest representations, in order to promote dissimilar semantic interests away from each other. According to the universal methods of contrastive learning, we have to construct positive pairs as well as corresponding negative pairs first. Considering the purpose of optimizing target user u multi-interest representations, we randomly choose an interest prototype vector I_k, then we treat (u_k, I_k) as positive pair, and (u_k, u_j) as corresponding negative pairs for all $j \neq k$. Afterwards, we define a interest-level contrastive learning objective function as follows:

$$\mathcal{L}_{ssl} = -\sum_{i \in TS} \log \frac{\exp(f(u_k, I_k))}{\sum_{j \neq k} \exp(f(u_k, u_j))} \tag{7}$$

where TS denotes the training set and $f(\cdot)$ is a scoring function for sample pairs, such as cosine similarity. Through optimizing this objective function, we are able to make interest-specific representation close to the corresponding interest prototype. In the meantime, the interest-level user representations under different conditions become more differentiated. In this way, our method is able to model diverse user interests more accurately.

3.5 Adaptive User Aggregator & Click Predictor

After obtaining the desired fine-grained user multi-interest representations, we learn target user representation u in an adaptive way considering given candidate news n_c. First we calculate δ_k, the normalization of similarity score between candidate news and interest prototype vector I_k, for $k \in [1, K]$. This indicates the probability that the candidate news related to the specific interest.

$$\delta_k = \frac{\exp(f(n_c, I_k))}{\sum_{i=1}^{K} \exp(f(n_c, I_i))} \tag{8}$$

The probability δ_k can also be regarded as the weight of corresponding interest-level user representation when aggregated into final summary representation. Next it adopts a weight summation operation to generate adaptive user representation considering candidate news, i.e., $u = \sum_{k=1}^{K} \delta_k \cdot u_k$.

It can help fine-grained matching between candidate news and target user reading interests. Finally, the click probability score y is computed by the dot product between the target user representation and the candidate news representation, i.e., $y = u^T \cdot n_c$.

# News	161,013	# Users	1,000,000
# News category	20	# Impression	15,777,377
# Entity	3,299,687	# Click behavior	24,155,470
Avg. title len.	11.52	Avg. abstract len.	43.00
Avg. body len.	585.05		

Table 1. Statistic information of MIND-Large dataset.

3.6 Model Training

Following [21], we use negative sampling techniques for model training. Given a positive sample n_i (clicked news, labeled as 1) in the training dataset, we then randomly select P negative samples $[n_{i,1}, \cdots, n_{i,P}]$ (non-clicked news, labeled as 0) from the same impression displayed to target user. Denote the click probability score of the positive and the P negative news as y_i^+ and $[y_{i,1}^-, y_{i,2}^-, \cdots, y_{i,P}^-]$ respectively. The supervised classification loss is formulated as follows:

$$\mathcal{L}_{ce} = -\sum_{i \in TS} \log \frac{\exp(y_i^+)}{\exp(y_i^+) + \sum_{j=1}^{P} \exp(y_{i,j}^-)} \tag{9}$$

Since we have already obtained the main classification loss function and auxiliary contrastive loss function, we define the final loss function in a joint paradigm as:

$$\mathcal{L} = \mathcal{L}_{ce} + \alpha \cdot \mathcal{L}_{ssl} \tag{10}$$

where α is a hyper-parameter that makes a trade-off between classification loss and contrastive loss. Minimizing the joint loss $\mathcal{L}$ helps to obtain fine-grained and differentiated user multi-interest representations.

4 Experiment

4.1 Dataset and Experimental Settings

We conduct extensive experiments on two large-scale real-world datasets, MIND-Large [5] and MIND-Small [6], to evaluate the effectiveness of our method. MIND-Large dataset collected from Microsoft News platform contains two record documents. One document describes text content of news, including titles and abstracts. The other document describes interaction behaviors that each user clicked news and these click behaviors are gathered from October 12 to November 22, 2019 (six weeks). The click behaviors in the first four weeks are regarded as user reading history, the behaviors in the penultimate week is applied for training, and the data in last week is used for performance evaluation. Detailed statistic information about MIND-Large dataset is summarized in Table 1.

[5] https://msnews.github.io/

[6] A small version of the MIND-Large dataset by randomly sampling 50,000 users and their behavior logs.

Next, we introduce experimental and hyper-parameters settings of our method. For news text content modeling, we utilize the first 30 words of news titles to learn news representations. In addition, a special character [PAD] is used for filling when the length of word sequence does not meet the condition. Besides, we adopt pre-trained Glove embeddings [11] for word initialization. For user interest modeling, we treat the recent 50 clicked news as user reading history. Moreover, news representations and user representations, including user multi-interest representations and adaptive user representations are both 400-dimensional vectors. For hyper-parameters, the number of interest prototypes K is set to 5. The joint learning weight α is set to 1.0. And the number of user neighbors T is merely set to 2 to promote the time efficiency. In addition, we utilize dropout technique and Adam optimizer for training. The dropout rate and learning rate are 0.1 and 0.001 respectively. Following [21], we use four metrics, i.e., AUC, MRR, nDCG@5, and nDCG@10, for performance evaluation. Notably, AUC is the most important one among them.

4.2 Performance Evaluation

We first introduce the baseline methods we compared in experiments, including six sequence-based and three GNN-based methods [7]: (1) EBMR [10] learns user representations from clicked news history via a GRU network. (2) DKN [18] utilizes an adaptive attention network to learn user representations considering relatedness between candidate news and historical news. (3) NPA [20] employs personalized attention networks to learn individual representation for each user. (4) NAML [19] leverages CNN networks to model news semantic representations and learns user representations through attentively aggregating clicked news. (5) LSTUR [1] models short-term user interests via a GRU network and long-term user interests via user ID embeddings. (6) NRMS [21] learns news representations and user representations through utilizing multi-head self-attention networks respectively. (7) GNewsRec [7] models user short-term interests by applying attentive GRU neural network and user long-term interests via graph neural networks based on user-news-topic heterogeneous graph. (8) GERL [5] uses the neighbors of news and users on the user-news graph to enhance their representations. (9) User-as-Graph [22] proposes a heterogeneous graph pooling method to learn user interest representations from the personalized heterogeneous graph.

The purpose of this section is to verify the effectiveness of our method. Thus we first conduct experiments to compare our model with several baseline models on MIND-Large dataset and then apply them to MIND-Small dataset for supplement. The overall performance results are displayed in Table 2 and Table 3 respectively, from which we have several observations: First, in terms of AUC, our proposed MIECL outperforms all baselines on both two datasets. We achieve 1.55% and 2.28% improvement comparing to state-of-the-art result respectively

[7] Due to the limitation of computer resources, we did not use the pretrained language models to encode the news titles and compare with baselines based on pretrained models.

Method	AUC	MRR	nDCG@5	nDCG@10
EBNR	65.42	31.24	33.76	39.47
DKN	64.60	31.32	33.84	39.48
NPA	66.69	32.24	34.98	40.68
NAML	66.86	32.49	35.24	40.91
LSTUR	67.73	32.77	35.59	41.34
NRMS	67.76	33.05	35.94	41.63
GNewsRec	67.53	32.68	35.46	41.17
GERL	68.24	33.46	36.38	42.11
User-as-Graph	69.23	**34.14**	37.21	43.04
MIECL*	**70.30**	34.13	**37.87**	**44.31**

Table 2. Performance of different methods on MIND-Large Dataset (%). *The improvement is significant at the level p < 0.001.

Method	AUC	MRR	nDCG@5	nDCG@10
EBNR	61.62	28.07	30.55	37.07
DKN	63.99	28.95	31.73	37.07
NPA	64.28	29.64	32.28	38.93
NAML	64.30	29.81	32.64	39.11
LSTUR	65.68	30.44	33.49	39.95
NRMS	65.43	30.74	33.13	39.66
GNewsRec	65.91	30.50	33.56	40.13
GERL	66.22	30.89	34.28	40.50
User-as-Graph	66.71	31.13	34.51	40.95
MIECL*	**68.23**	**32.46**	**36.17**	**42.32**

Table 3. Performance of different methods on MIND-Small Dataset (%). *The improvement is significant at the level p < 0.001.

on each dataset. Besides, our model achieves excellent performance in terms of other evaluation metrics, which shows the effectiveness and adaptability to data scale of our model. Since neither sequence-based methods nor GNN-based methods recognize the significance of modeling diverse user interests, they merely use one vector to represent user reading interests. This result illustrates the necessity of modeling diverse user interests explicitly. Second, GNN-based methods usually perform better than sequence-based methods, since they utilize GNN to obtain high-order neighbor information. However, as we elaborate in the ablation study below, although GNN module does not produce particularly significant improvement to our model, we still acquire even better performance than them. This result further demonstrates the effectiveness and importance of modeling diverse user interests, which deserves our further exploration.

4.3 Ablation Study

In order to further evaluate the effectiveness of each module in our method, we conduct the ablation experiments on MIND-Small dataset and report the results

	AUC	MRR	nDCG@5	nDCG@10
MIECL	68.23	32.46	36.17	42.32
w/o multi-interest user encoder	66.56	31.51	34.55	41.00
w/o multi-interest graph-enhanced module	67.93	32.83	36.27	42.43
w/o multi-interest contrastive learning module	67.36	31.83	35.13	41.42

Table 4. Effect of each module perform in our model.

in Table 4. According to the performance of different variants, we then discuss the effect of each module in our method.

w/o multi-interest user encoder. The module is the concrete implementation portion of motivation. This variant indicates each user only has a single representation vector, in addition contrastive learning here is not applicable. Compared with the results of the complete model, this variant showed a significant decline of 2.45% in terms of AUC. Therefore we verify the significance of mining fine-grained and diverse user interests and conclude that such modeling method is beneficial to personalized news recommendation.

w/o multi-interest graph-enhanced module. In this variant, we learn user representations from their clicked historical news without aggregating neighbors information. Excluding the module decreases the AUC by 0.44%. Obviously, utilizing neighbors information can enrich user semantic representations to a certain extent.

w/o multi-interest contrastive learning module. This variant describes the situation that the original joint learning framework is turned into a separate recommendation task. It is noticed that the performance of this variant drops by 1.28% in terms of AUC. Without the contrastive learning module, the discrimination between user multiple representations is reduced, which leads to user multi-interest representations become similar. To some extent, this is equivalent to model user interests using a single vector. As the first variant discussed, the performance of model will degrade. The result demonstrates the importance of the contrastive learning module.

4.4 Hyper-Parameters Analysis

In order to further evaluate the sensitivity of our method to hyper-parameters, we conduct experiments on MIND-Small dataset with different values of hyper-parameters, including the number of interest prototypes K and the joint learning weight α.

Figure 3 shows the trend variation of our method with various number of interest prototypes. It can be seen that with the increases of K, the performance of MIECL will also improve first in terms of all metrics. This is because user interests is usually diverse and rarely relatively single. Hence, utilizing the fine-grained modeling method can better explore the diversity of user interests. However, when the value of K becomes larger, the performance of our model gradually decreases. We speculate that although user interests is diverse, the

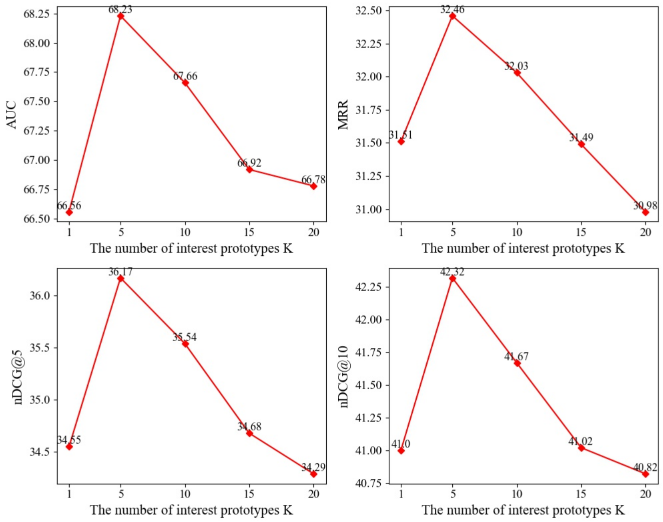

Fig. 3. Performance of MIECL with different values of hypermeter K.

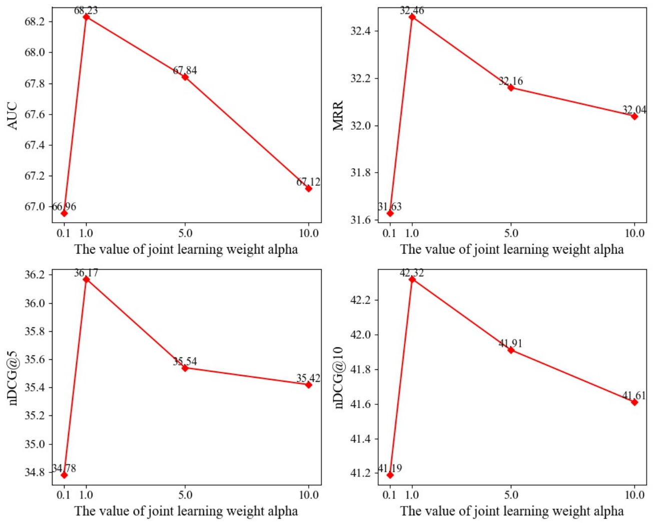

Fig. 4. Performance of MIECL with different values of hypermeter α.

number of them can not be particularly numerous. Once the value of K is too large, a lot of interest-level noise information will be introduced while aggregated into the final user representations. This is likely to be detrimental to the recommendation performance. The deduction is also conform to the reality and intuition. Furthermore, we discover that the result echoes with Figure 1, because most users browse about 5 categories of news.

Figure 4 shows the trend variation of our method with different value of joint learning weight. The abscissa represents the value of α and the ordinate represents the results of evaluation metrics. We can discover that the trend variation is similar to that in Figure 3. When α is relatively small, the impact of contrastive learning decreases as well. This is because we are hardly to model fine-grained interests once the learnt multi-interest representations are not sufficiently differentiated. Then when α becomes larger, the influence of contrastive learning is extremely exaggerated at this time. The consequence is that the affect of recommendation classification loss will be correspondingly reduced, which is not conducive to the main recommendation task.

4.5 Statistic Analysis

In this section, we conduct statistical experiments on distances between interest prototypes on MIND-Small dataset. Specifically, we record the representations of interest prototypes after each epoch of training procedure. Then we calculate the euclidean distances between each pair of interest prototype vectors and conduct statistical analysis, at different training epoch. The statistical results are displayed in the Figure 5.

Quite evidently, with the increases of training epoch, the average (max, min) distance between interest prototype vectors is also gradually increasing. This phenomenon further implies that contrastive learning can distinguish the interest prototypes effectively. In addition, this will be beneficial to learning more differentiated user multi-interest representations.

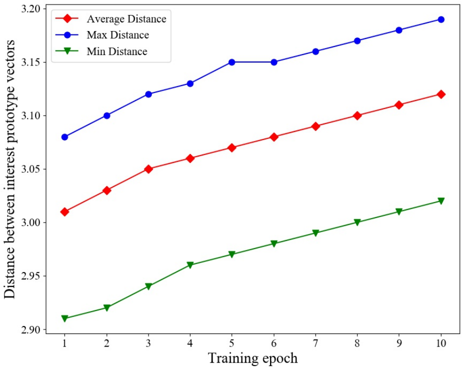

Fig. 5. Statistics of the distances between interest prototype vectors, at different training epoch.

5 Conclusion

In this paper, we propose a novel method combining multi-interest extraction with contrastive learning, named MIECL, to model diverse user interests effectively. Specifically, first, we construct several interest prototypes and design a multi-interest user encoder to simultaneously learn multiple user representations under each prototype. Then we adopt a graph-enhanced user encoder to enrich user corresponding semantic representation under each interest background. Finally, we contrast user multi-interest representations and interest prototypes to optimize the user representations themselves, in order to promote dissimilar semantic interest away from each other. Extensive experiments on real-world datasets validate the effectiveness of our approach.

Acknowledgements We sincerely thank all the anonymous reviewers for their comments and suggestions. This work is supported by the National Key Research and Development Program of China (grant No.2021YFB3100600), the National Natural Science Foundation of China (No.62106059), the Strategic Priority Research Program of Chinese Academy of Sciences (grant No.XDC02040400), , and the Youth Innovation Promotion Association of CAS (Grant No. 2021153).

References

1. An, M., Wu, F., Wu, C., Zhang, K., Liu, Z., Xie, X.: Neural news recommendation with long-and short-term user representations. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. pp. 336–345 (2019)
2. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International conference on machine learning. pp. 1597–1607. PMLR (2020)
3. Das, A.S., Datar, M., Garg, A., Rajaram, S.: Google news personalization: scalable online collaborative filtering. In: Proceedings of the 16th international conference on World Wide Web. pp. 271–280 (2007)
4. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
5. Ge, S., Wu, C., Wu, F., Qi, T., Huang, Y.: Graph enhanced representation learning for news recommendation. In: Proceedings of The Web Conference 2020. pp. 2863–2869 (2020)
6. Hjelm, R.D., Fedorov, A., Lavoie-Marchildon, S., Grewal, K., Bachman, P., Trischler, A., Bengio, Y.: Learning deep representations by mutual information estimation and maximization. In: International Conference on Learning Representations (2018)
7. Hu, L., Li, C., Shi, C., Yang, C., Shao, C.: Graph neural news recommendation with long-term and short-term interest modeling. Information Processing & Management **57**(2), 102142 (2020)
8. Khattar, D., Kumar, V., Varma, V., Gupta, M.: Weave&rec: A word embedding based 3-d convolutional network for news recommendation. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management. pp. 1855–1858 (2018)

9. Liu, D., Lian, J., Wang, S., Qiao, Y., Chen, J.H., Sun, G., Xie, X.: Kred: Knowledge-aware document representation for news recommendations. In: Fourteenth ACM Conference on Recommender Systems. pp. 200–209 (2020)
10. Okura, S., Tagami, Y., Ono, S., Tajima, A.: Embedding-based news recommendation for millions of users. In: Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining. pp. 1933–1942 (2017)
11. Pennington, J., Socher, R., Manning, C.D.: Glove: Global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). pp. 1532–1543 (2014)
12. Qi, T., Wu, F., Wu, C., Yang, P., Yu, Y., Xie, X., Huang, Y.: Hierec: Hierarchical user interest modeling for personalized news recommendation. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). pp. 5446–5456 (2021)
13. Tian, Y., Krishnan, D., Isola, P.: Contrastive multiview coding. In: European conference on computer vision. pp. 776–794. Springer (2020)
14. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
15. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y.: Graph attention networks. arXiv preprint arXiv:1710.10903 (2017)
16. Veličković, P., Fedus, W., Hamilton, W.L., Liò, P., Bengio, Y., Hjelm, R.D.: Deep graph infomax. In: International Conference on Learning Representations (2018)
17. Wang, C., Blei, D.M.: Collaborative topic modeling for recommending scientific articles. In: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 448–456 (2011)
18. Wang, H., Zhang, F., Xie, X., Guo, M.: Dkn: Deep knowledge-aware network for news recommendation. In: Proceedings of the 2018 world wide web conference. pp. 1835–1844 (2018)
19. Wu, C., Wu, F., An, M., Huang, J., Huang, Y., Xie, X.: Neural news recommendation with attentive multi-view learning. In: Proceedings of the 28th International Joint Conference on Artificial Intelligence. pp. 3863–3869 (2019)
20. Wu, C., Wu, F., An, M., Huang, J., Huang, Y., Xie, X.: Npa: neural news recommendation with personalized attention. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 2576–2584 (2019)
21. Wu, C., Wu, F., Ge, S., Qi, T., Huang, Y., Xie, X.: Neural news recommendation with multi-head self-attention. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). pp. 6389–6394 (2019)
22. Wu, C., Wu, F., Huang, Y., Xie, X.: User-as-graph: User modeling with heterogeneous graph pooling for news recommendation. In: Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence. pp. 1624–1630 (2021)
23. Wu, F., Qiao, Y., Chen, J.H., Wu, C., Qi, T., Lian, J., Liu, D., Xie, X., Gao, J., Wu, W., et al.: Mind: A large-scale dataset for news recommendation. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 3597–3606 (2020)