



High-level preferences as positive examples in contrastive learning for multi-interest sequential recommendation

Zizhong Zhu¹ · Shuang Li² · Yaokun Liu² · Xiaowang Zhang² · Zhiyong Feng² · Yuexian Hou²

Received: 31 December 2023 / Revised: 7 February 2024 / Accepted: 4 March 2024 /
Published online: 14 March 2024

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

The sequential recommendation task based on the multi-interest framework aims to model multiple interests of users from different aspects to predict their future interactions. However, researchers rarely consider the differences in features between the interests generated by the model. In extreme cases, all interest capsules have the same meaning, leading to the failure of modeling users with multiple interests. To address this issue, we propose the High-level Preferences as positive examples in Contrastive Learning for multi-interest Sequence Recommendation framework (HPCL4SR), which uses contrastive learning to distinguish differences in interests based on user item interaction information. In order to find high-quality comparative examples, this paper introduces the category information to construct a global graph, learning the association between categories for high-level preference interest of users. Then, a multi-layer perceptron is used to adaptively fuse the low-level preference interest features of the user's items and the high-level preference interest features of the categories. Finally, user multi-interest contrastive samples are obtained through item sequence information and corresponding categories, which are fed into contrastive learning to optimize model parameters and generate multi-interest representations that are more in line with the user sequence. In addition, when modeling the user's item sequence information, in order to increase the differentiation between item representations, the category of the item is used to supervise the learning process. Extensive experiments on three real datasets demonstrate that our method outperforms existing multi-interest recommendation models.

Keywords Multi-interest learning · Sequential recommendation · Category information · Contrastive learning

This article belongs to the Topical Collection: *Special Issue on Advancing recommendation systems with foundation models*

Guest Editors: Kai Zheng, Renhe Jiang, and Ryosuke Shibasaki

✉ Xiaowang Zhang
xiaowangzhang@tju.edu.cn

Extended author information available on the last page of the article

1 Introduction

The sequence recommendation (SR) [1, 2] task treats the interaction between users and items as a dynamic sequence, capturing interests of user over time by modeling sequence dependencies, and predicting future user interactions. Sequence recommendation provides services in many aspects of daily life, such as e-commerce [3, 4], news media [5, 6], video music [7], and social networks [8].

Hence, a variety of SR models, including both shallow and deep models, have been proposed to improve the performance of sequential recommendations. Specifically, Recurrent Neural Networks built on Gate Recurrent Units (GRU) have been employed to model the long- and short-term point-wise sequential dependencies over user-item interactions for next-item recommendations [9, 10]. Convolutional Neural Network (CNN) [11], self-attention [12, 13] and Graph Neural Network [14, 15] models have been incorporated into sequential recommendation systems for capturing more complex sequential dependencies for further improving the performance.

When modeling user interests using the above structure, it is common to represent user interests as low dimensional embeddings, which contradicts the fact that each user may have multiple interests in reality [16]. Some studies [3, 17–19] propose capturing multiple user interest vectors from different aspects instead of a single vector. These methods explicitly generate users’ diverse interest representations from their behavior sequences, breaking the representation bottleneck of using a single generic user embedding. Although these solutions have achieved significant performance improvements, they have not taken into account the differences between multiple interests. In the worst-case scenario, all interest capsules have the same meaning and cannot reflect the diversity of user interests. Recent work has further improved the modeling ability of multiple user interests through routing regularization [20]. However, as shown in Figure 1, we calculated the similarity between multiple interest feature vectors generated by REMI [20] and HPCL4SR model on the Amazon-Clothing dataset, and randomly selected 128 users for statistical analysis. We found that although REMI [20] models can also represent multiple interests of users, the representation of user interests has great similarity, and HPCL4SR significantly reduces the similarity between multiple interests of users. In other words, our model can better represent multiple different interests of users, increasing the diversity between interests.

Concretely, we propose a novel multi-interest sequence recommendation framework (named HPCL4SR). HPCL4SR models users’ high-level preference interests by constructing

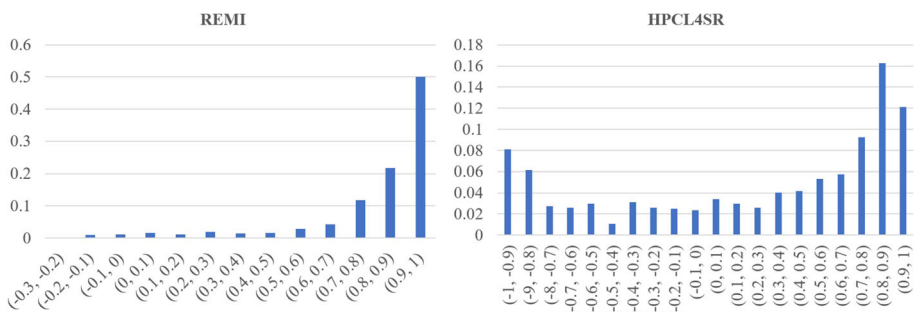


Figure 1 The similarity between multiple interests of users is randomly selected from 128 users for statistical purposes. The horizontal axis represents the range of similarity values, and the vertical axis represents the percentage of similarity values in the corresponding interval

a global graph of categories, and feeding them as positive examples into contrastive learning, optimizing multiple interest representations of users. Specifically, based on the item sequence behavior of all users, the categories of items are constructed as global graph information for high-level preference interest modeling. In this process, in order to alleviate the imbalance problem in class interaction, attention weights are used to reconstruct the adjacency matrix, and multi-hop aggregation is performed on categories that are not directly connected to each other, reducing the sparsity of the interaction matrix and increasing the correlation between classes. For the user's item sequence, the context information of the sequence is obtained by encoding the position of the item and fusing attention weights. Then, capsule networks are used to learn the item sequence information for low-level interest preference modeling. In order to further enhance the vector representation ability of items, the network model parameters are reverse optimized using the category information of items as labels. Naturally, we will integrate high-level and low-level preference interest to generate multiple interest features for users. Contrastive learning can maximize the similarity between related samples and minimize the similarity between unrelated samples. This paper draws inspiration from this idea but distinguishes multiple interests of users through learning while not treating interests as completely unrelated and preserving the hidden correlation information between interests. Therefore, we use low-level preference interest corresponding to the fused features of high-level preference interest as positive examples, and other preference interest as negative examples to learn the differences between user interests.

To summarize, the contributions of this paper are as follows:

- We propose a new novel interest sequence recommendation framework (HPCL4SR), which solves the problem of existing methods not being able to represent the multiple interests of users.
- We construct a global graph based on the category information of items to model the user's high-level preference interest. In addition, using it as a positive example in contrastive learning is a relatively optimal approach.
- We conduct extensive experiments on three real-world datasets to verify the effectiveness of the HPCL4SR. Further analysis has demonstrated that the proposed method can more reasonably model the diversity of multiple interests of users.

2 Related work

This paper mainly uses category information as positive examples to solve the problem of differences among multiple interests in sequence recommendation through contrastive learning. Therefore, this section briefly overviews representative efforts relevant to our work from Multi-Interest Sequence Recommendation, Large Language Models for Recommendation, and Contrastive Learning.

2.1 Multi-interest sequence recommendation

In practical scenarios, users' historical behavior has complex interaction patterns, and modeling interests as a single vector using the above methods is not sufficient to accurately reflect users' true multi preference interest. Therefore, studying sequential recommendation models based on multiple interests has become more important and practical.

MIND [17] proposes a multi-interest extractor layer based on the capsule routing mechanism, which is applicable for clustering historical behaviors and extracting diverse interests.

SDM [21] uses a multi head attention mechanism in encoding behavior sequences to capture multiple interests of the user. SIN [22] adaptively infers user interaction interests from a large number of interest pools and outputs multiple interest embeddings, and then uses the attention weights of items to generate multiple interest embeddings that best match user characteristics. ComiRec [18] is based on multi head attention based multi-interest routing to capture multiple interests of users and introduce controllable factors to achieve diverse recommendations. PIMI [23] models the periodic features of temporal information between user behaviors and the interactive features between sequence items, respectively, and uses their representations to describe users' multiple interests. DuoRec [24] designs contrastive regularization to reshape the distribution of sequence representations and selects sequences with the same target item as hard positive samples, alleviating the problem of representation degradation in multi-interest sequence recommendation tasks. UMI [25] believes that the interests of a user are not only reflected in their historical behavior, but also inherently regulated by the profile information. Therefore, the user profiles are introduced as a source of multi-interest features for users. REMI [20] first mitigates the problem of easy negatives with an ideal interest-aware hard negative sampling distribution and an approximation method to achieve the goal at a negligible computational cost. REMI also incorporates a novel routing regularization to avoid routing collapse and further improve the modeling capacity of multi-interest models.

2.2 Large language models for recommendation

Recent years have witnessed the wide adoption of large language models (LLMs) in different fields, especially natural language processing and computer vision. Such a trend can also be observed in recommendation systems (RS). However, due to the huge number of items in real-world systems, traditional RS usually takes the two-stage filtering paradigm of the matching stage (It aims to extract a small subset of items from the extensive corpus with lightweight models, ensuring low computational costs.) and ranking (It utilizes more sophisticated models to rerank the retrieved items), advanced recommendation algorithms are not applied to all items, but only a few hundred of items. [26]. Therefore, existing large language models (LLMs) (e.g., ChatGPT) methods [27–29] focus on the sorting stage that utilizes more sophisticated models to rerank the retrieved items. We focus on improving the effectiveness of the matching stage, which serves as a crucial foundation for the recommendation systems. To our knowledge, there is currently no research on the application of the LLMs methods in the matching stage. But, in the experimental section, we will attempt to analyze the performance of ChatGPT (ChatGPT 3.5-Turbo-1106 & ChatGPT 4-Turbo) in the matching stage.

2.3 Contrastive learning

Contrastive learning has been widely applied in the field of computer vision. In contrastive learning, methods such as CPC [30, 31] and DIM [32] feed the encoding of the same image at different scales as positive samples, while MoCo [33], SimSiam [34], CaCo [35] and other methods use multiple image enhancements as positive samples for contrastive learning. In the field of text information processing, some studies use different data-transforming methods or strategies, such as dropout and mask, to change the parameters and structure of the encoder to improve the model's ability to perform sentence representation [36–38]. The introduction of contrastive learning in sequence recommendation systems mainly solves the problems of sparse user-item interaction and noise. Scholars improve recommendation

performance by designing auxiliary tasks or loss functions [39]. CBIT [40] combines the cloze task mask and the dropout mask to generate high-quality positive samples and perform multi-pair contrastive learning. ICLRec [39] models user intentions through clustering of item sequences, maximizing the agreement between a view of the sequence and its corresponding intentions to improve recommendation performance. CL4SRec [41] employs contrastive learning to learn consistent perception enhancement representations from sequential pattern encoding and global collaborative relationship modeling.

Although researchers are trying to describe users' various interests in different ways, they rarely consider the issue of diversity in interests. In the worst case, all interest capsules have the same meaning, or all items may activate the same interest capsule, which makes it difficult to express multiple interests. ComiRec [18] uses controllable factors to recommend diverse user interests, but the paper also mentions that increasing diversity can lead to a decrease in recall rates. REMI [20] observe that the interests tend to over-focus on single items in the behavior sequence, which impacts the expressiveness of multi-interest representations. They introduce the variance regularizer on the routing weights to eliminate sparsity and effectively address the problem. MIRACLE [19] forces interest capsules to satisfy orthogonality, which clearly provides each user with K unrelated interests. However, such K interests can cause unnecessary item recommendations for users, which goes against our common sense that there may be implicit correlations between interests. Therefore, it is necessary and meaningful for multi-interest sequence recommendations to preserve implicit correlation information while ensuring the difference between interests. We attempt to solve the above problem through contrastive learning, which distinguishes the differences in interests through self-supervised learning of data features and maintains the correlation information between the representations of interests.

3 Problem formulation

Assume U denotes a set of users, X denotes a set of items, and C is a set of categories. Each item x_i has its corresponding category c_i . Given a user $u \in U$, we have his/her chronological item interaction sequence $S_x^u = \{x_1^u, x_2^u, \dots, x_N^u\}$ and a corresponding category interaction sequence $S_c^u = \{c_1^u, c_2^u, \dots, c_N^u\}$, where $x_t^u \in X$ and $c_t^u \in C$ represents the item and its category that user u interacted with at time step t , respectively. N is the maximum sequence length. The candidate matching stage in RS aims to efficiently retrieve a subset of items the user is likely to interact with from the huge item corpus X .

4 Method

In this section, we propose the High-level Preferences as positive examples in Contrastive Learning for multi-interest Sequence Recommendation framework (HPCL4SR), as shown in Figure 2. There are three parts: high-level preference interest extraction module, low-level preference interest extraction module, and multi-interest contrastive learning module.

4.1 High-level preference interest extraction module

Experiments in numerous fields of contrastive learning applications have shown that selecting good positive and negative samples is the key to the effectiveness of contrastive learning.

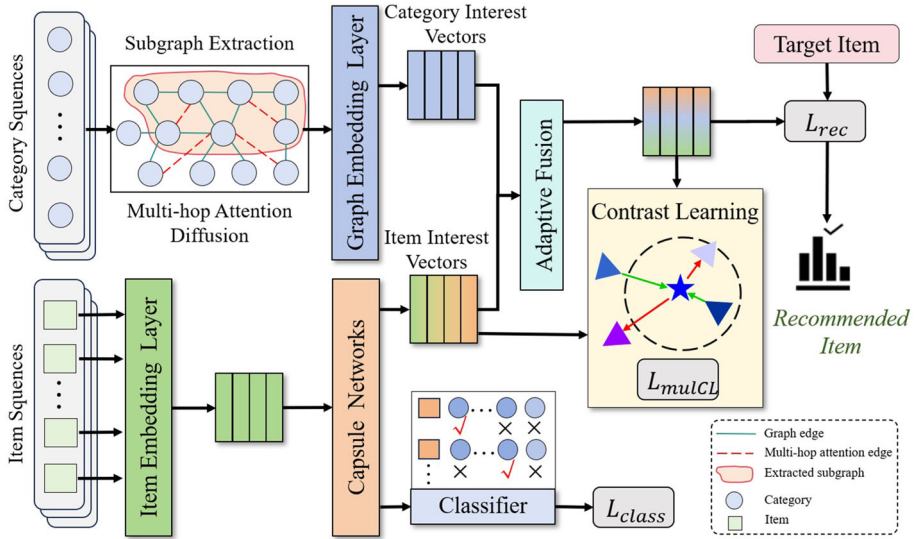


Figure 2 An Overview of Multi-interest Sequential Recommendation (HPCL4SR) framework

In sequence recommendation tasks, most models use methods such as pruning sequences, dropout, and mask to construct contrastive samples. As shown in Figure 3, the user’s historical interaction is extremely sparse, and such operations will not fully represent the user’s interests and may even result in errors. A large amount of excellent work has proven that taking side information (user profile, category, brand, description, price, position, rating, etc.) into recommendation sequences can better capture user preference information [25, 42, 43]. In real scenarios, item category information is the easiest to obtain and is a high-level conceptual representation of the item. Therefore, in this paper, we will use item category information as a contrastive sample. In fact, even though the number of categories is much smaller than the number of items, the interaction between categories is still sparse. So the method

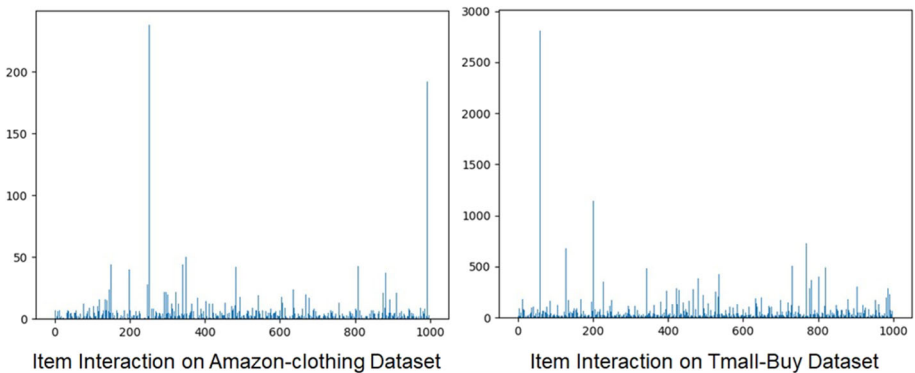


Figure 3 Sparsity analysis on Amazon-Clothing and Tmall-Buy datasets. The horizontal axis represents 1000 randomly selected items in the dataset, and the vertical axis represents the number of times items interact in the dataset

proposed in this paper does not directly take the category sequence corresponding to the item as the user’s high-level preference, but instead models the user’s high-level preference interest by constructing a category global graph, learning more preference interest correlation information from the user.

For user Category sequences $S_c^u = \{c_1^u, c_2^u, \dots, c_N^u\}$, we calculate the number of interactions between c_i^u and c_j^u . And a category global graph (A_1) is constructed based on the historical interaction category sequence $\{S_c^1, S_c^2, \dots, S_c^{|U|}\}$ of all users, where the initial weight of the edges between the two nodes is the total number of interactions between the two categories a_{ij} .

However, such a category global graph (A_1) still has two obvious problems: (1) By analyzing the interaction frequency of categories, it was found that due to the significant difference in the number of items contained in different categories, there is an imbalance in the interaction between categories, which will lead to recommendation results biased towards items in popular categories. (2) The method of constructing a graph through sequential interaction only considers the relationship between adjacent item categories, while ignoring the interaction between non directly adjacent categories. In fact, certain categories that are not adjacent often appear together in the user’s sequence.

In order to alleviate the imbalance of category interaction and avoid the impact of popular items on recommendation results, the adjacency matrix is redefined as follows:

$$A_2(i, j) = \frac{a_{ij}}{\sqrt{|a_i + 1||a_j + 1|}} \tag{1}$$

where a_{ij} denotes the number of interactions with category c_i and category c_j , a_i are the number of interactions category c_i with others, and a_j is similar to a_i .

To learn the correlation between non-adjacent categories, we adapt the Multi-hop Attention Diffusion [44] method to aggregate information further. The attention score of multi-hop neighbors is calculated by:

$$A = \sum_{i=0}^{\infty} \theta_i A_2^i \tag{2}$$

where $\sum_{i=0}^{\infty} \theta_i = 1 (\theta_i > 0)$, θ_i is the attention decay factor, $\theta_i > \theta_{i+1}$, i is the power of the adjacency matrix A_2 which represents the farthest length of the diffusion relation path and also represents the farthest length of the graph diffusion relation path.

Assume $H^{(0)} \in R^{|C| \times d}$ denotes the initial embedding matrix of the category, and d represents the dimension of the node embedding. We use GCN to aggregate the features of neighbors as a new representation of the target node, and introduce residual connections in this process. The message-passing process is as follows:

$$H^{(l)} = W(AH^{(l-1)} + H^{(l-1)}) \tag{3}$$

where l is the number of GCN layers, W is trainable parameter matrices.

The final graph representation $\hat{H} \in R^{|C| \times d}$ is obtained by:

$$\hat{H} = -\frac{1}{L} \sum_{i=0}^L H^{(i)} \tag{4}$$

Based on the category information of user historical interactions, category node embedding representation H_g is selected from the graph:

$$H_g = \text{selecte}(\hat{H})[n, :], n = 1, \dots, N \tag{5}$$

where, $H_g \in \mathbb{R}^{N \times d}$, N is the sequence length of user interaction.

Finally, the user's high-level preference interest vector Q_u calculated as follows:

$$Q_u = WHg \quad (6)$$

where, $Q_u \in \mathbb{R}^{K \times d}$, K is the number of preference interests.

4.2 Low-level preference interest extraction module

In sequence recommendation, positional information can explicitly reflect contextual information between items. Therefore, in this paper, the attention mechanism is first used to encode sequence information:

$$X_i = E_i^{emb} + E_i^{pos} \quad (7)$$

where E_i^{emb} , E_i^{pos} is the embedding of the i -th item, and the positional embedding, respectively, X_i is an item embedding representation with sequence position information.

$$\alpha_{ij} = \frac{\exp(X_i X_j^T)}{\sum_{n=1}^N \exp(X_i X_n^T)} \quad (8)$$

where α_{ij} is the attention weight of item j to item i , We use neural networks to make each item in the sequence perceive the entire contextual information.

$$X_i = W(X_i + \sum_{j=1}^N \alpha_{ij} X_j) \quad (9)$$

In multi-interest recommendation tasks, the effectiveness of Capsule Network [45] has been verified, so we directly draw on the above method to extract user low-level preference interests $P_u \in \mathbb{R}^{K \times d}$.

$$P_u = CapsNet([X_1, X_2, \dots, X_N]) \quad (10)$$

In addition, in order to enable the network model to learn the differences between items, category can be used as a good supervised label. Specifically, we use the first layer information of the Capsule Network as the feature $Z = \{z_1, z_2, \dots, z_N\}$ of the item, and use the fully connected layer as the classifier. The output result $\hat{Z} \in \mathbb{R}^{N \times d}$ can be represented as follows:

$$\hat{Z} = softmax(WZ + b) \quad (11)$$

Category $S_c = \{c_1, c_2, \dots, c_N\}$ is used as a label, and the cross entropy loss function calculates the loss of the classifier:

$$L_{class} = - \sum_{n=1}^N (c_n \log \hat{z}_n + (1 - c_n) \log (1 - \hat{z}_n)) \quad (12)$$

4.3 Multi-interest contrastive learning module

The differentiation processing between interests is the key to achieving multi-interest sequence recommendations. Existing methods rarely consider the differences between interest capsules, resulting in user sequence interest capsules having the same meaning in extreme cases. This paper uses a contrastive learning approach to distinguish between interests while

preserving their implicit correlation information. The sequence information is used to represent the true interests of the user in an adaptive fusion.

We assume that the high-level preference interests of the category sequence (Q_u) are consistent with the low-level preference interests of the item sequence (P_u). We use fully connected layers to adaptively fuse them, and obtain the final multi-interest representation (M_u) of the user. Finally, P_u and M_u are selected as two views for contrastive learning.

$$M_u = W (cat[Q_u, P_u]) \tag{13}$$

where $M_u \in \mathbb{R}^{K \times d}$

Most existing contrastive learning methods are based on InfoNCE:

$$L_{cl} = -\log \frac{e^{(\mathbf{h}_i \cdot \mathbf{h}_{i^*} / \tau)}}{e^{(\mathbf{h}_i \cdot \mathbf{h}_{i^*} / \tau)} + \sum_{j \neq i^*} e^{(\mathbf{h}_i \cdot \mathbf{h}_j / \tau)}} \tag{14}$$

where τ is a temperature hyperparameter, $(\mathbf{h}_i, \mathbf{h}_{i^*})$ is positive pair, $(\mathbf{h}_i, \mathbf{h}_{j \neq i^*})$ is negative pair.

However, due to a lack of decision margin, a small perturbation around the decision boundary may lead to an incorrect decision. To overcome the problem, inspired by ArcFace [46], we propose a new training objective for multi-interest contrastive learning by adding an additive angular margin m between positive pair \mathbf{e}_i and \mathbf{e}_{i^*} . Therefore, (14) can be rewritten as follows:

$$L_{cl} = -\log \frac{e^{\cos(\theta_{i,i^*} + m) / \tau}}{e^{\cos(\theta_{i,i^*} + m) / \tau} + \sum_{j \neq i^*} e^{\cos(\theta_{i,j}) / \tau}} \tag{15}$$

where m is additive angular margin.

$$\theta_{i,j} = \arccos \left(\frac{\mathbf{e}_i^\top \mathbf{e}_j}{\|\mathbf{e}_i\| * \|\mathbf{e}_j\|} \right) \tag{16}$$

To some extent, more negative samples can lead to better performance in contrastive learning. In this paper, we set $\{i \in M_u^i, i^* \in P_u^i\}$ or $\{i \in P_u^i, i^* \in M_u^i\}$, $j \in \{P_u \cup M_u\}$. In this way, for a sequence, any i will have $(2K - 2)$ negative samples, and the contrastive loss of multi-interest sequences recommended function is:

$$L_{mulCL} = -\sum_i \log \frac{e^{\cos(\theta_{i,i^*} + m) / \tau}}{\sum_{j \neq i} e^{\cos(\theta_{i,j}) / \tau}} \tag{17}$$

4.4 Model training

For the given target item embedding y , we use an argmax operator to obtain the interest that is the most related to the target item through (18):

$$m_u = M_u \left[:, \operatorname{argmax} \left(M_u^\top y \right) \right] \tag{18}$$

The loss function between the predicted results of the model and the given target is :

$$L_{rec} = -\log \frac{\exp(m_u y^T)}{\sum_{j \in X'} \exp(m_u y_j^T)} \tag{19}$$

where X' is the item obtained through sampling softmax objective [47].

The joint loss is defined as a linear combination of these three losses :

$$L = L_{rec} + \lambda_1 L_{class} + \lambda_2 L_{multCL} \quad (20)$$

where λ_1 and λ_2 are the hyperparameters to control the impact of different losses.

5 Experiments

5.1 Experimental settings

Dataset We consider three real-world e-commerce datasets. The specific statistics are shown in Table 1.

- **Amazon-Clothing¹**. The Amazon Review Dataset is a classic data set commonly used in recommender systems, which records product reviews. We use the *Clothing Shoes and Jewelry* subset in our experiment.
- **Tmall-Buy²**. The Tmall dataset is collected by Tmall.com, which is an online shopping website. It contains users' shopping history for about six months. We retain users' purchase behaviors as a subset for experiments.
- **Tafeng³**. The Tafeng dataset collects user transaction behavior data from November 2000 to February 2001. The dataset covers everything from food and office supplies to furniture.

Baselines We compare our model with some sequential recommendation methods.

- **GRU4Rec [48]**. GRU4Rec is a representative recommendation model that first introduces recurrent neural networks into sequence recommendation.
- **MIND [16]**. MIND is one of the first frameworks to model users' multiple interests based on dynamic routing algorithms.
- **ComiRec [18]**. ComiRec is a representative baseline for the multi-interest recommendation. It uses two methods to represent user interests: attention mechanism and dynamic routing.
- **PIMI [23]**. Considering the limitations of ComiRec, PIMI introduces the study of periodicity and interactivity of item sequences, capturing both global and local item features.
- **REMI [20]**. REMI consists of an Interest-aware Hard Negative mining strategy and a Routing Regularization method to solve the issues of increased easy negatives and routing collapse during the training process.

Evaluation Metrics We use three common accuracy metrics for performance evaluation: *Recall*, *Normalized Discounted Cumulative Gain (NDCG)*, and *Hit Rate (HR)*. Metrics computation relies on the top 20/50 recommended candidates (e.g., Recall@20). For the three metrics, higher scores demonstrate better recommendation performance.

Implement Details For each dataset, we partition all users into the training, validation, and test sets with a ratio of 8:1:1. The maximum sequence length of the Amazon-Clothing and Tafeng datasets is set to 30, and the maximum sequence length of Tmall-Buy dataset is 20. The user sequence whose length exceeds the maximum value is truncated, and the

¹ <https://nijianmo.github.io/amazon/>

² <https://tianchi.aliyun.com/dataset/dataDetail?dataId=42>

³ <https://www.kaggle.com/chiranjivdas09/ta-feng-grocery-dataset>

Table 1 Statistics of the three datasets

Datasets	Amazon-Clothing	Tmall-Buy	Tafeng
Users	129,827	356,795	17,590
Items	293,446	311,746	10,176
Categories	499	1,442	1,297
Interactions	3,046,500	35,012,248	674,685
Sparsity	99.99%	99.97%	99.62%

user sequence whose length is insufficient is filled with 0. We filter users/items with fewer than 12 interactions to guarantee the length of recent sequences. All parameters are set as follows if not otherwise noted: following [20], the learning rate is 0.001, the mini-batch size is 128, the embedding size is set to 64, the interest number $K = 4$, and Adam is used as a gradient optimizer. We analyze in detail the effects of other hyperparameters in Section 5.4 and ultimately determined their values as $\lambda_1 = 0.1$, $\lambda_2 = 1$, $\tau = 0.05$, $m = 10$, respectively.

5.2 Performance evaluation

To demonstrate the recommendation performance of our model HPCL4SR, we compare it with other multi-interest models. The experimental results of three datasets are presented in Table 2. We have the following observations.

First, although the three datasets have different characteristics, HPCL4SR consistently yields the best performance, indicating the robustness of our model. By modeling users' diverse interests through contrastive learning, even for the Tafeng dataset with limited user interests, we can still fully use the limited information to capture user interests and make optimal recommendations. It proves the effectiveness of finer-grained characterization of user persona contained in a user sequence. Overall, we model the high-level preference interests and low-level preference interests of user sequences, and distinguish the feature representations between interests through contrastive learning, which can more finely characterize users. Considering the meaning behind different user behaviors, that is, the real users of each interaction project, we mine and utilize the information in the user behavior sequence to understand users' interests from the perspective of project usage, which helps to model users more accurately.

Next, judging from the performance of sequential recommendation models, The performance of PIMI, UMI, and HPCL4SR models on three datasets is superior to most models, such as ComiRec and MIND, indicating that adding side information is beneficial for user modeling.

Finally, on the Tafeng dataset with a small number of items (10,176), models that use a single vector to model user interests (GRU4Rec) outperform simple multi-vector models (MIND and ComiRec). However, the model performance is still worse than the PIMI model using time information and the HPCL4SR model the category information we proposed. The multi-interests model is better than the single interests model on the Amazon-Clothing and Tmall-Buy datasets with many users and items. In addition, the results of the REMI model indicate that selecting high-quality negative samples can bring surprises, but this requires a significant time cost in screening negative samples. Overall, modeling users' interests from different aspects are better than using only one vector to model users' overall interests. Because the multi-interest models can provide users with more mixed recommendation results, thereby improving the accuracy of recommendations.

Table 2 Model comparison results on three benchmark datasets (%)

Datasets	Metric	GRU4Rec	MIND	ComiRec	PIMI	UMI	REMI	HPCL4SR	Improv.
Amazon- Clothing	R@20	10.3255	11.3067	11.1502	11.1995	11.2153	<u>11.7424</u>	12.8135	+9.12%
	R@50	12.6394	12.8059	<u>13.1319</u>	12.6933	12.4681	12.7248	13.9153	+5.97%
	NDCG@20	16.9736	13.7823	14.7854	14.6619	14.4982	15.5834	18.4271	+8.56%
	NDCG@50	17.1947	14.2741	15.2312	15.1364	16.0716	18.0857	19.2631	+6.51%
	HR@20	22.1212	20.4652	21.2663	21.8285	21.8462	<u>22.4385</u>	23.4534	+4.52%
	HR@50	25.6489	23.4383	23.9775	24.5860	24.7258	<u>25.8605</u>	29.5961	+6.71%
Tmall-Buy	R@20	6.9243	7.6904	8.4168	9.1229	9.1502	9.2118	9.8637	+7.08%
	R@50	9.3749	10.6973	10.8794	11.9598	11.8167	<u>12.4578</u>	13.6973	+9.95%
	NDCG@20	12.1815	12.3772	13.9452	14.6609	14.5167	<u>15.8735</u>	17.1652	+8.14%
	NDCG@50	13.1349	13.7061	14.9777	15.9583	15.7438	<u>16.2381</u>	18.0591	+11.21%
	HR@20	19.8836	22.0929	23.4026	24.5271	24.8307	<u>25.8650</u>	27.9468	+8.05%
	HR@50	25.5854	29.4087	29.2764	31.2211	32.1578	<u>32.3058</u>	33.6805	+4.26%
Tateng	R@20	5.6437	4.6833	5.5811	9.9490	<u>9.9618</u>	9.9304	10.3804	+4.20%
	R@50	9.9867	8.1007	9.6806	15.4121	15.8671	<u>16.3140</u>	17.1385	+5.05%
	NDCG@20	12.5107	12.1007	12.9474	21.3443	21.4849	<u>21.7319</u>	22.8653	+5.22%
	NDCG@50	15.3459	14.7493	15.5079	23.4896	22.6093	<u>24.6194</u>	26.1562	+6.24%
	HR@20	29.5696	25.0142	29.9034	45.0256	45.7518	<u>46.4991</u>	48.3451	+2.88%
	HR@50	44.6076	38.1467	43.7180	58.4423	58.5316	<u>59.8437</u>	62.2735	+4.06%

The bold and underlined numbers represent the best and second-best results, respectively

In addition, to better demonstrate the effectiveness of our method, we supplemented the experimental results of CL4SRec [41] and DuoRec [24] in the sequence recommendation task. In Table 3, the experiment demonstrates the effectiveness and ingenuity of using contrastive learning methods, and also demonstrates that modeling users with multiple interests in sequence recommendation tasks can improve recommendation performance.

5.3 Ablation study

In this section, we select the Amazon Clothing dataset to analyze the effectiveness of our proposed method (HPCL4SR). Firstly, we refer to the method in HPCL4SR that only uses item information as the base and the method that uses category as the supervised signal as the base ($w L_{class}$). Then, After constructing a global graph based on category to obtain user high-level preference interest information, we further attempted three methods of integrating high-level preference interest (Q_u) and low-level preference interests (P_u): addition, multiplication, and adaptive fusion, represented as HPCL4SR($w '+'$), HPCL4SR($w '*'$), HPCL4SR, respectively. Finally, we analyze and consider the contribution of differences in interests to multi-interest sequence recommendation. We not only attempt to replace the multi-interest contrastive learning module with Capsule Regulation [19], denoted as HPCL4SR($w CR$), but also analyze the impact of the lack of additive angle margin m , denoted as HPCL4SR($w/o m$). The experimental results on three data sets are shown in Table 4. From the table, it can be seen that category, as a supervisory signal, improves certain performance by optimizing the representation of items. When combined with Q_u , the performance improvement is significant, especially when using the MLP method. The experiment confirms that the difference between interests is the main reason for affecting the performance of sequence recommendation, and the contrastive learning method shows greater advantages than the Capsule Regularization form due to its ability to distinguish the differences between interests while preserving the correlation information between them.

5.4 Hyper-parameter study

λ_1 and λ_2 are hyperparameters of the joint loss function during the training process, which directly affect the optimization of model parameters. We selected $\lambda_1 \in \{0.01, 0.05, 0.1, 0.5, 1\}$, $\lambda_2 \in \{0.01, 0.1, 1, 5, 10\}$, and conducted experiments on three datasets using NDCG@50 as the evaluation metric. As shown in the left side of Figure 4, we can see that the best performance is achieved when $\lambda_1 = 0.1$. This matches our intuition since using category as the supervisory signal for items is effective, but excessive weight can cause recommendation loss and reduce model performance. As shown in the right side of Figure 4, an increase in the weight λ_2 of the comparison loss can help distinguish the differences between multiple interests, but if λ_2 is too large, it can also mask the recommendation loss and reduce the model's recommendation task ability. Therefore, a reasonable value $\lambda_2 = 1$ is needed.

Table 3 Model comparison results on Amazon-Clothing datasets with purely contrastive learning-based sequence recommendation method (%)

Metric	CL4SRec	DuoRec	HPCL4SR
R@20	11.06	12.17	12.8135
R@50	12.11	13.43	13.9153
NDCG@20	9.7	10.52	18.4271
NDCG@50	9.91	10.78	19.2631

Table 4 Ablation study on Amazon-Clothing dataset

	Amazon-Clothing					
	@20			@50		
	Recall	NDCG	Hit Rate	Recall	NDCG	Hit Rate
base	11.4263	14.8396	20.4735	12.8437	15.3571	24.6158
base(w L_{class})	11.5174	14.9467	20.9148	12.9166	15.7637	25.1097
HPCL4SR(w '+')	11.8457	15.3918	21.5617	13.2094	16.8463	25.9168
HPCL4SR(w '*')	11.8704	15.3493	21.6869	13.2275	17.1951	26.2942
HPCL4SR(w CR)	12.2387	16.0857	22.7928	13.5469	18.3674	27.0805
HPCL4SR(w/o m)	12.4064	17.6431	23.0348	13.6918	18.8762	27.1506
HPCL4SR	12.8135	18.4271	23.4534	13.9153	19.2631	27.5961

The temperature τ and angular margin m in the multi-interest contrastive learning module affect its effectiveness. For τ , we carry out an experiment with τ varying from 0.01 to 0.1 with an interval of 0.01. The results are shown in the left side of Figure 5. On Amazon-Clothing and Tmall-Buy datasets, the performance is best when $\tau = 0.05$, and on the Tafeng dataset, the result is best when $\tau = 0.03$ (however, the performance difference between it and $\tau = 0.05$ is small). Taking all factors into consideration, we chose $\tau = 0.05$ for all our experiments. For m , as shown in the right side of Figure 5, we selected $m \in \{0, 5, 10, 15, 20\}$. Although the performance is best on the Tmall-Buy dataset when $m = 15$, it is best on the other two datasets when $m = 10$. Therefore, we set $m = 10$ during the experiment.

5.5 Case study

We analyze the proposed model’s effectiveness in solving the multi-interest recommendation problem by showing the model’s recommendation results. Because the Amazon-Clothing dataset contains detailed information such as items and item categories, while the Tmall-Buy only gives the data number, we use the item to represent the recommendation results of the datasets.

Figure 6 shows the recommendation results of the proposed HPCL4SR model for a certain user behavior sequence. It can be seen from the figure that the user is more interested in baby boy suits and boy socks, but the PIMI model only recommends items related to men and women and does not learn the interests of the two demand sides of baby boys and boys. The

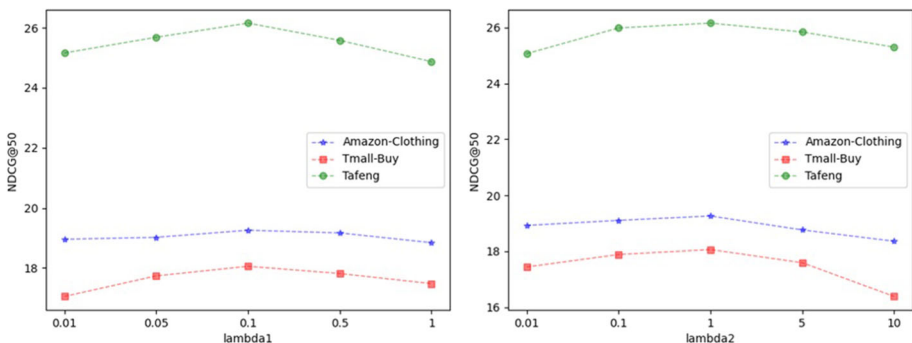


Figure 4 Study on balance parameter λ_1 and λ_2 . We show NDCG@50 on three datasets

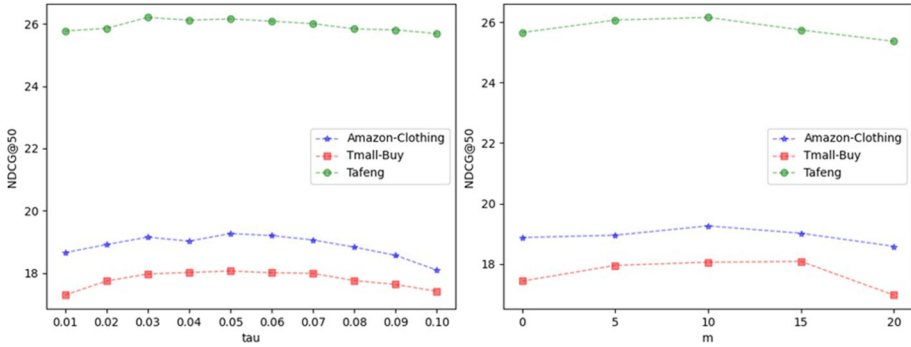


Figure 5 Study on balance parameter τ and m . We show NDCG@50 on three datasets

HPCL4SR models both high-level and low-level preference interests of users, and diversifies the interests to recommend the “baby suit” and “socks” that users want. Moreover, in the list of recommended items given by the HPCL4SR model, the “baby suit” item that the user actually interacts with ranks higher. That is, the ranking quality of the list of recommended items provided by the model is higher. In addition, the HPCL4SR model not only learns the interests of boys but also captures the preferences of boys, and at the same time, learns other categories such as socks and shoes.

Table 5 shows the Top-20 recommendation results of the PIMI and HPCL4SR models on the randomly selected user behavior numbered 68079 in the Tmall-Buy dataset. As can be seen from the table, the HPCL4SR model correctly predicts the items that two users interact with (the number is bold). Compared with the PIMI model, item ID 31744 in the recommendation



Figure 6 A case study on Amazon-Clothing dataset

Table 5 A case study on Tmall-Buy dataset

	Item ID
User Interaction Sequence	96 → 2231 → 7885 → 109 → 1373 → 1373 → 9748 → 9956
Label	674 → 31477 → 4067
Recommended	15104, 9956, 23972, 518, 4039, 14666, 109, 7885, 5421, 4877, 208, 206, 2931,
Items of PIMI	9748, 4820, 10614, 789, 31477 , 7600, 797
Recommended	12097, 4067 , 9956, 134, 15991, 28140, 4877, 19128, 7600, 31477 , 7184,
Items of HPCL4SR	9393, 9748, 34005, 10614, 696, 2071, 4820, 4948, 1021

result list given by the HPCL4SR model ranks higher. Therefore, the HPCL4SR model exceeds the performance of the PIMI model.

5.6 vs. LLMs

In order to compare the recommendation ability of HPCL4SR and large-scale language models in the matching stage, we use two prompt methods and have ChatGPT (ChatGPT 3.5-Turbo-1106 & ChatGPT 4-Turbo) provide recommended items based on user interaction information. The first method is to input user historical interaction information and prompt ChatGP to generate 50 items of interest to the user. The second method is to input user interaction history information and input the user's next real item as well as randomly selected items, prompting ChatGPT to select items that the user may be interested in from the existing 50 items based on interaction history. The experimental details and results are shown in the Table 6. It can be seen that The performance of ChatGPT 4-Turbo is better than that of

Table 6 The result of recommendation compared with LLMs

Prompt	model	Amazon-Clothing
You are a recommender system now: Input: Here is the purchase Item history of a user: {User History item}.	ChatGPT 3.5-Turbo-1106	HR@50 = 4.3
Based on this history, please recommend the user prefer items next. Answer a sequence must contain 50 Item.	ChatGPT 4-Turbo	HR@50 = 7.0
You are a recommender system now: Input: Here is the purchase Item history of a user: {User History item}.	ChatGPT 3.5-Turbo-1106	HR@50 = 17.0
Based on this history, please recommend the user prefer items from Candidate Item. Answer a selected sequence must contain 20 Item.	ChatGPT 4-Turbo	HR@50 = 23.0
Fine-tuning	T5-small	HR@50 = 22.68
/	HPCL4SR	HR@50 = 29.6

ChatGPT 3.5-Turbo-1106, but the results of both far lower than our model and even many existing models. This indicates that the current general LLMs still cannot be well applied to specific task domains [26, 49]. In addition, although T5-small [50] is much lower in model parameters than ChatGPT, we found that the large language model can further improve the performance of recommendation systems through fine-tuning, but it still cannot meet the model we have carefully designed for sequence recommendation. However, the expressive power of the large model will be enlightening for our next work.

6 Conclusion

In this paper, we propose a novel framework named HPCL4SR for multi-interest sequence recommendation. In order to achieve the representation of multiple user interests, HPCL4SR uses contrastive learning methods to differentiate interests, while preserving their correlation information, which is more in line with user behavior in real scenarios. We verify the effectiveness of the proposed method through experiments on three datasets. Additionally, we compare the recommendation ability of our approach in a task-specific domain with LLMs (ChatGPT 3.5-Turbo-1106 & ChatGPT 4-Turbo), further showcasing the superiority of HPCL4SR in multi-interest sequential recommendation. In the future, we consider enhancing the interpretability of recommendation tasks based on multi-interest recommendation models.

Author Contributions Zizhong.Zhu. and Shuang.Li. wrote the main manuscript text, Yaokun.Liu. revised the figure, Xiaowang.Zhang. served as the corresponding author to coordinate and promote the completion of the paper, and Zhiyong.Feng. and Yuexian.Hou. provided reasonable revision opinions. All authors reviewed the manuscript.

Funding National Natural Science Foundation of China (NSFC) (61972455)

Data Availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

References

1. Wang, S., Hu, L., Wang, Y., Cao, L., Sheng, Q.Z., Orgun, M.A.: Sequential recommender systems: Challenges, progress and prospects. In: Proceedings of the twenty-eighth international joint conference on artificial intelligence: IJCAI, pp. 6332–6338 (2019). <https://doi.org/10.24963/ijcai.2019/883>
2. Quadrana, M., Cremonesi, P., Jannach, D.: Sequence-aware recommender systems. *ACM Comput. Surv.* **51**(4), 1–36 (2018). <https://doi.org/10.1145/3190616>
3. Pi, Q., Bian, W., Zhou, G., Zhu, X., Gai, K.: Practice on long sequential user behavior modeling for click-through rate prediction. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining: KDD, pp. 2671–2679 (2019). <https://doi.org/10.1145/3292500.3330666>
4. Zhang, Y., Liu, Y., Xiong, H., Liu, Y., Yu, F., He, W., Xu, Y., Cui, L., Miao, C.: Cross-domain disentangled learning for e-commerce live streaming recommendation. In: 39th IEEE International conference on data engineering: ICDE, pp. 2955–2968 (2023). <https://doi.org/10.1109/ICDE55515.2023.00226>
5. Chu, Q., Liu, G., Sun, H., Zhou, C.: Next news recommendation via knowledge-aware sequential model. In: Chinese computational linguistics - 18th China national conference: CCL. Lecture Notes in Computer Science, vol. 11856, pp. 221–232 (2019). https://doi.org/10.1007/978-3-030-32381-3_18

6. Wu, C., Wu, F., Qi, T., Huang, Y.: User modeling with click preference and reading satisfaction for news recommendation. In: Proceedings of the twenty-ninth international joint conference on artificial intelligence: IJCAI, pp. 3023–3029 (2020). <https://doi.org/10.5555/3491440.3491858>
7. Chaves, P.D.V., Pereira, B.L., Santos, R.L.T.: Efficient online learning to rank for sequential music recommendation. In: WWW '22: The ACM Web Conference 2022, pp. 2442–2450 (2022). <https://doi.org/10.1145/3485447.3512116>
8. Li, Q., Wang, X., Wang, Z., Xu, G.: Be causal: De-biasing social network confounding in recommendation. *ACM Trans. Knowl. Discov. Data* **17**(1), 1–23 (2023). <https://doi.org/10.1145/3533725>
9. Song, W., Wang, S., Wang, Y., Wang, S.: Next-item recommendations in short sessions. In: RecSys '21: fifteenth acm conference on recommender systems, pp. 282–291 (2021). <https://doi.org/10.1145/3460231.3474238>
10. Hidasi, B., Karatzoglou, A.: Recurrent neural networks with top-k gains for session-based recommendations. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management: CIKM, pp. 843–852 (2018). <https://doi.org/10.1145/3269206.3271761>
11. Yuan, F., Karatzoglou, A., Arapakis, I., Jose, J.M., He, X.: A simple convolutional generative network for next item recommendation. In: Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining: WSDM, pp. 582–590 (2019). <https://doi.org/10.1145/3289600.3290975>
12. Kang, W., McAuley, J.J.: Self-attentive sequential recommendation. In: IEEE International Conference on Data Mining: ICDM, pp. 197–206 (2018). <https://doi.org/10.1109/ICDM.2018.00035>
13. Wang, S., Cao, L., Hu, L., Berkovsky, S., Huang, X., Xiao, L., Lu, W.: Hierarchical attentive transaction embedding with intra- and inter-transaction dependencies for next-item recommendation. *IEEE Intell. Syst.* **36**(4), 56–64 (2021). <https://doi.org/10.1109/MIS.2020.2997362>
14. Wang, N., Wang, S., Wang, Y., Sheng, Q.Z., Orgun, M.A.: Modelling local and global dependencies for next-item recommendations. In: Web information systems engineering: WISE. Lecture Notes in Computer Science, vol. 12343, pp. 285–300 (2020). https://doi.org/10.1007/978-3-030-62008-0_20
15. Zheng, X., Su, J., Liu, W., Chen, C.: DDGHM: dual dynamic graph with hybrid metric training for cross-domain sequential recommendation. In: MM '22: The 30th ACM international conference on multimedia: MM, pp. 471–481 (2022). <https://doi.org/10.1145/3503161.3548072>
16. Li, C., Liu, Z., Wu, M., Xu, Y., Zhao, H., Huang, P., Kang, G., Chen, Q., Li, W., Lee, D.L.: Multi-interest network with dynamic routing for recommendation at tmall. In: Proceedings of the 28th ACM international conference on information and knowledge management: CIKM, pp. 2615–2623 (2019). <https://doi.org/10.1145/3357384.3357814>
17. Xiao, Z., Yang, L., Jiang, W., Wei, Y., Hu, Y., Wang, H.: Deep multi-interest network for click-through rate prediction. In: CIKM '20: The 29th ACM international conference on information and knowledge management, pp. 2265–2268 (2020). <https://doi.org/10.1145/3340531.3412092>
18. Cen, Y., Zhang, J., Zou, X., Zhou, C., Yang, H., Tang, J.: Controllable multi-interest framework for recommendation. In: KDD '20: The 26th ACM SIGKDD conference on knowledge discovery and data mining (KDD), pp. 2942–2951 (2020). <https://doi.org/10.1145/3394486.3403344>
19. Tang, Z., Wang, L., Zou, L., Zhang, X., Zhou, J., Li, C.: Towards multi-interest pre-training with sparse capsule network. In: Proceedings of the 46th international ACM SIGIR conference on research and development in information retrieval: SIGIR, pp. 311–320 (2023). <https://doi.org/10.1145/3539618.3591778>
20. Xie, Y., Gao, J., Zhou, P., Ye, Q., Hua, Y., Kim, J.B., Wu, F., Kim, S.: Rethinking multi-interest learning for candidate matching in recommender systems. In: Proceedings of the 17th ACM conference on recommender systems: RecSys, pp. 283–293 (2023). <https://doi.org/10.1145/3604915.3608766>
21. Lv, F., Jin, T., Yu, C., Sun, F., Lin, Q., Yang, K., Ng, W.: SDM: sequential deep matching model for online large-scale recommender system. In: Proceedings of the 28th ACM international conference on information and knowledge management: CIKM, pp. 2635–2643 (2019). <https://doi.org/10.1145/3357384.3357818>
22. Tan, Q., Zhang, J., Yao, J., Liu, N., Zhou, J., Yang, H., Hu, X.: Sparse-interest network for sequential recommendation. In: WSDM '21, The Fourteenth ACM international conference on web search and data mining: WSDM, pp. 598–606 (2021). <https://doi.org/10.1145/3437963.3441811>
23. Chen, G., Zhang, X., Zhao, Y., Xue, C., Xiang, J.: Exploring periodicity and interactivity in multi-interest framework for sequential recommendation. In: Proceedings of the thirtieth international joint conference on artificial intelligence: IJCAI, pp. 1426–1433 (2021). <https://doi.org/10.24963/ijcai.2021/197>
24. Qiu, R., Huang, Z., Yin, H., Wang, Z.: Contrastive learning for representation degeneration problem in sequential recommendation. In: WSDM '22: The fifteenth acm international conference on web search and data mining: WSDM, pp. 813–823 (2022). <https://doi.org/10.1145/3488560.3498433>
25. Chai, Z., Chen, Z., Li, C., Xiao, R., Li, H., Wu, J., Chen, J., Tang, H.: User-aware multi-interest learning for candidate matching in recommenders. In: SIGIR '22: The 45th international acm sigir conference

- on research and development in information retrieval: SIGIR, pp. 1326–1335 (2022). <https://doi.org/10.1145/3477495.3532073>
26. Li, L., Zhang, Y., Liu, D., Chen, L.: Large language models for generative recommendation: A survey and visionary discussions. *CoRR* **abs/2309.01157** (2023) <https://doi.org/10.48550/ARXIV.2309.01157> <https://arxiv.org/abs/2309.01157>
 27. Hou, Y., Zhang, J., Lin, Z., Lu, H., Xie, R., McAuley, J.J., Zhao, W.X.: Large language models are zero-shot rankers for recommender systems. *CoRR* **abs/2305.08845** (2023) <https://doi.org/10.48550/ARXIV.2305.08845> <https://arxiv.org/abs/2305.08845>
 28. Luo, S., He, B., Zhao, H., Huang, Y., Zhou, A., Li, Z., Xiao, Y., Zhan, M., Song, L.: c: Instruction tuning large language model as ranker for top-k recommendation. *arXiv preprint arXiv:2312.16018* (2023) <https://doi.org/10.48550/arXiv.2312.16018>
 29. Bao, K., Zhang, J., Zhang, Y., Wang, W., Feng, F., He, X.: Tallrec: An effective and efficient tuning framework to align large language model with recommendation. In: Zhang, J., Chen, L., Berkovsky, S., Zhang, M., Noia, T.D., Basilio, J., Pizzato, L., Song, Y. (eds.) *Proceedings of the 17th ACM conference on recommender systems, RecSys 2023*, Singapore, Singapore, September 18–22, 2023, pp. 1007–1014. ACM, ??? (2023). <https://doi.org/10.1145/3604915.3608857>
 30. Hénaff, O.J.: Data-efficient image recognition with contrastive predictive coding. In: *Proceedings of the 37th international conference on machine learning: icml. proceedings of machine learning research*, vol. 119, pp. 4182–4192 (2020). <https://proceedings.mlr.press/v119/henaff20a.html>
 31. Oord, A., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. *CoRR* **abs/1807.03748** (2018) <https://doi.org/10.48550/arXiv.1807.03748>
 32. 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: 7th International conference on learning representations: ICLR (2019). <https://openreview.net/forum?id=Bklr3j0cKX>
 33. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.B.: Momentum contrast for unsupervised visual representation learning. In: 2020 IEEE/CVF Conference on computer vision and pattern recognition: CVPR, pp. 9726–9735 (2020). <https://doi.org/10.1109/CVPR42600.2020.00975>
 34. Chen, X., He, K.: Exploring simple siamese representation learning. In: *IEEE Conference on computer vision and pattern recognition: CVPR*, pp. 15750–15758 (2021). <https://doi.org/10.1109/CVPR46437.2021.01549>
 35. Wang, X., Huang, Y., Zeng, D., Qi, G.: Caco: Both positive and negative samples are directly learnable via cooperative-adversarial contrastive learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **45**(9), 10718–10730 (2023). <https://doi.org/10.1109/TPAMI.2023.3262608>
 36. Yan, Y., Li, R., Wang, S., Zhang, F., Wu, W., Xu, W.: Consert: A contrastive framework for self-supervised sentence representation transfer. In: *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing: ACL/IJCNLP*, pp. 5065–5075 (2021). <https://doi.org/10.18653/v1/2021.acl-long.393>
 37. Gao, T., Yao, X., Chen, D.: Simcse: Simple contrastive learning of sentence embeddings. In: *Proceedings of the 2021 conference on empirical methods in natural language processing: EMNLP*, pp. 6894–6910 (2021). <https://doi.org/10.18653/v1/2021.emnlp-main.552>
 38. Zhang, Y., Zhu, H., Wang, Y., Xu, N., Li, X., Zhao, B.: A contrastive framework for learning sentence representations from pairwise and triple-wise perspective in angular space. In: *Proceedings of the 60th annual meeting of the association for computational linguistics: ACL*, pp. 4892–4903 (2022). <https://doi.org/10.18653/v1/2022.acl-long.336>
 39. Chen, Y., Liu, Z., Li, J., McAuley, J.J., Xiong, C.: Intent contrastive learning for sequential recommendation. In: *WWW '22: The ACM web conference 2022: WWW*, pp. 2172–2182 (2022). <https://doi.org/10.1145/3485447.3512090>
 40. Du, H., Shi, H., Zhao, P., Wang, D., Sheng, V.S., Liu, Y., Liu, G., Zhao, L.: Contrastive learning with bidirectional transformers for sequential recommendation. In: *Proceedings of the 31st ACM international conference on information & knowledge management: CIKM*, pp. 396–405 (2022). <https://doi.org/10.1145/3511808.3557266>
 41. Yang, Y., Huang, C., Xia, L., Huang, C., Luo, D., Lin, K.: Debaised contrastive learning for sequential recommendation. In: *Proceedings of the ACM Web Conference 2023: WWW*, pp. 1063–1073 (2023). <https://doi.org/10.1145/3543507.3583361>
 42. Yuan, X., Duan, D., Tong, L., Shi, L., Zhang, C.: ICAI-SR: item categorical attribute integrated sequential recommendation. In: *SIGIR '21: The 44th international ACM SIGIR conference on research and development in information retrieval: SIGIR*, pp. 1687–1691 (2021). <https://doi.org/10.1145/3404835.3463060>

43. Xie, Y., Zhou, P., Kim, S.: Decoupled side information fusion for sequential recommendation. In: SIGIR '22: The 45th international ACM SIGIR conference on research and development in information retrieval :SIGIR, pp. 1611–1621 (2022). <https://doi.org/10.1145/3477495.3531963>
44. Wang, G., Ying, R., Huang, J., Leskovec, J.: Multi-hop attention graph neural networks. In: Proceedings of the thirtieth international joint conference on artificial intelligence: IJCAI, pp. 3089–3096 (2021). <https://doi.org/10.24963/ijcai.2021/425>
45. Sabour, S., Frosst, N., Hinton, G.E.: Dynamic routing between capsules. In: Advances in neural information processing systems 30: annual conference on neural information processing systems 2017: NIPS, pp. 3856–3866 (2017). https://proceedings.neurips.cc/paper_files/paper/2017/file/2cad8fa47bbef282badbb8de5374b894-Paper.pdf
46. Deng, J., Guo, J., Yang, J., Xue, N., Kotsia, I., Zafeiriou, S.: Arcface: Additive angular margin loss for deep face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**(10), 5962–5979 (2022). <https://doi.org/10.1109/TPAMI.2021.3087709>
47. Jean, S., Cho, K., Memisevic, R., Bengio, Y.: On using very large target vocabulary for neural machine translation. In: Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing of the asian federation of natural language processing: ACL, pp. 1–10 (2015). <https://doi.org/10.3115/v1/P15-1001>
48. Hidasi, B., Karatzoglou, A., Baltrunas, L., Tikk, D.: Session-based recommendations with recurrent neural networks. In: 4th International conference on learning representations: ICLR (2016). <https://doi.org/10.48550/arXiv.1511.06939>
49. Fan, S., Wang, Y., Pang, X., Chen, L., Han, P., Shang, S.: Uamc: user-augmented conversation recommendation via multi-modal graph learning and context mining. *World Wide Web*, pp. 1–21 (2023). <https://doi.org/10.1007/s11280-023-01219-2>
50. Xu, S., Hua, W., Zhang, Y.: Openp5: Benchmarking foundation models for recommendation. *CoRR abs/2306.11134* (2023). <https://doi.org/10.48550/ARXIV.2306.11134>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Zizhong Zhu¹ · Shuang Li² · Yaokun Liu² · Xiaowang Zhang² · Zhiyong Feng² · Yuexian Hou²

Zizhong Zhu
zizhong_zhu@tju.edu.cn

Shuang Li
l_s_8565@tju.edu.cn

Yaokun Liu
yaokunl@tju.edu.cn

Zhiyong Feng
zyfeng@tju.edu.cn

Yuexian Hou
yxhou@tju.edu.cn

¹ School of New Media and Communication, Tianjin University, Tianjin, China

² College of Intelligence and Computing, Tianjin University, Tianjin, China