

# Preference Aware Dual Contrastive Learning for Item Cold-Start Recommendation

Wenbo Wang<sup>1</sup>, Bingquan Liu<sup>1\*</sup>, Lili Shan<sup>1</sup>, Chengjie Sun<sup>1</sup>, Ben Chen<sup>2</sup>, Jian Guan<sup>3</sup>

<sup>1</sup> Faculty of Computing, Harbin Institute of Technology, Harbin, 150001, China

<sup>2</sup> Alibaba Group, Hangzhou, 310000, China

<sup>3</sup> Group of Intelligent Signal Processing, College of Computer Science and Technology, Harbin Engineering University, Harbin, 150001, China

wwb1325864697@outlook.com, {liubq, shanlili, sunchengjie}@hit.edu.cn, chenben.cb@alibaba-inc.com, j.guan@hrbeu.edu.cn

## Abstract

Existing cold-start recommendation methods often adopt item-level alignment strategies to align the content feature and the collaborative feature of warm items for model training, however, cold items in the test stage have no historical interactions with users to obtain the collaborative feature. These existing models ignore the aforementioned condition of cold items in the training stage, resulting in the performance limitation. In this paper, we propose a preference aware dual contrastive learning based recommendation model (PAD-CLRec), where the user preference is explored to take into account the condition of cold items for feature alignment. Here, the user preference is obtained by aggregating a group of collaborative feature of the warm items in the user's purchase records. Then, a group-level alignment between the user preference and the item's content feature can be realized via a proposed preference aware contrastive function for enhancing cold-item recommendation. In addition, a joint objective function is introduced to achieve a better trade-off between the recommendation performance of warm items and cold items from both item-level and group-level perspectives, yielding better overall recommendation performance. Extensive experiments are conducted to demonstrate the effectiveness of the proposed method, and the results show the superiority of our method, as compared with the state-of-the-arts.

## Introduction

Collaborative filtering (CF) has achieved remarkable success in personalized recommendation. The CF based models aim at learning quality collaborative feature (i.e., ID-embedding) from the historical interactions (i.e., clicks, rates, and views) to represent users and items (Yu et al. 2021; Wu et al. 2021; He et al. 2020, 2017; Rendle et al. 2012; Wei et al. 2019). However, these methods cannot generate collaborative feature for cold items that are newly coming without any historical interaction, and suffer from item cold-start issue. A common solution is to introduce items' multimedia information (i.e., image, video, and attributes) as content feature to represent the cold items. From the perspective of

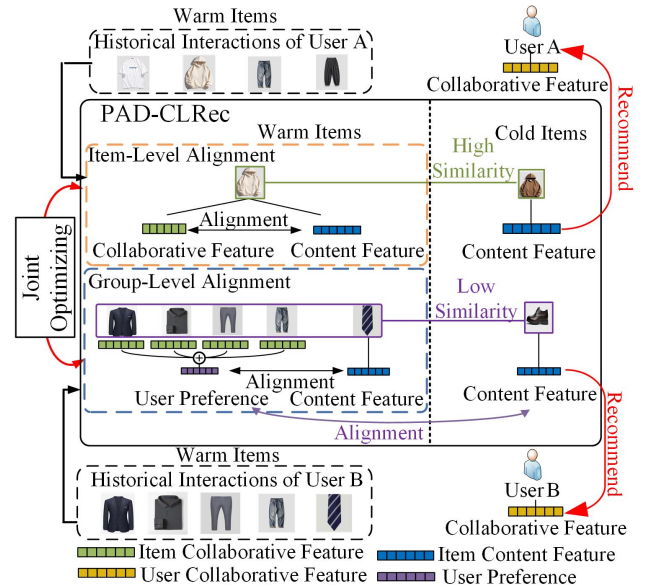


Figure 1: Illustration of the limitation of the item-level alignment strategy and the advantage of the proposed alignment strategy in our PAD-CLRec from both item-level and group-level perspectives.

robustness, some methods (i.e., DropoutNet (Volkovs, Yu, and Poutanen 2017) and MTPR (Du et al. 2020)) incorporate the item's content feature with the collaborative feature as the item's final representation, and treat a subset warm items as cold items via dropping their collaborative feature part in the training stage. Although these methods consider the condition of cold items with the dropout operation in the training stage, they often neglect the relation between item's content feature and collaborative feature, resulting in the limited overall performance.

To this end, recent studies often adopt item-level alignment strategy to explore the relation between item's content feature and collaborative feature from an item-level perspective in the training stage, and achieve improved performance. For example, GAR (Chen et al. 2022) and CLCRec

\*Corresponding author

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

(Wei et al. 2021) employ generative adversarial network (GAN) (Goodfellow et al. 2020) and contrastive learning, respectively, to align the collaborative feature and the content feature of the same warm item. Whereas CFCCRec (Zhou, Zhang, and Yang 2023) samples two warm items with co-occurrence users, and aligns the content feature of one item to the collaborative feature of the other item. However, since cold items have no collaborative feature and co-occurrence users with warm items, these item-level alignment strategy based methods did not take this condition into consideration in the training stage, leading to limited performance for the cold items that firstly appear in the test stage, especially for the low similarity cold items, i.e., the cold items whose multimedia information is less similar to the warm items. As illustrated in Figure 1, because of the highly consistent image information between the brown hoodie and the white hoodie, the brown hoodie’s content feature will be aligned to the white hoodie’s collaborative feature via the item-level alignment of the white hoodie’s content feature and its content feature. Therefore, the brown hoodie can be recommended to user A, who bought a white hoodie before. However, as for the leather shoes, item-level alignment is less effective because neither user A nor user B has ever interacted with an item whose content is highly consistent with the leather shoes. Whereas by analyzing user B’s preference for formal style clothes based on his purchase record, the leather shoes can be recommended to user B, which illustrates the necessity of exploring the relation between a selected item and a group of warm items from a group-level perspective.

To this end, we propose a preference aware dual contrastive learning based recommendation model (PAD-CLRec) for cold-start recommendation from both item-level and group-level perspectives, where a joint objective function is presented for model training, which includes a dual recommendation contrastive loss, a dual item-level aligning contrastive loss and a preference aware contrastive loss. From the group-level perspective, the user’s preference is firstly explored by aggregating the collaborative feature of a group of warm items that interacted with the user. Then, a preference aware contrastive loss is proposed to consider the condition of cold items in the training stage by exploring the relation between a randomly selected warm item’s content feature and the obtained user preference, where the randomly selected warm item can be treated as a cold item for model training. In addition, with the combination of the dual item-level aligning contrastive loss and the preference aware contrastive loss in the joint objective function, we can achieve both item-level and group-level feature alignment to explore the relation between items’ content feature and collaborative feature. Moreover, the relation between users and items can be also captured by the dual recommendation contrastive loss. Therefore, the proposed method can not only improve low similarity cold items’ recommendation performance, but also achieve a better trade-off between the recommendation of cold items and warm items. The contributions of our study can be summarized as follows,

- A preference aware contrastive loss is proposed to fully consider the cold items in the training stage by treating a randomly selected warm item as the cold item, and ex-

ploring the relation between its content feature and the user preference aggregated via a group of warm items’ collaborative feature from a group-level perspective.

- A joint objective function is proposed to achieve a better trade-off between the performance of warm items and cold items via exploring the relation between items’ collaborative feature and content feature from both item-level and group-level perspectives.
- Extensive experiments are conducted on two benchmark datasets to demonstrate the effectiveness of our proposed method, and results show that our method significantly outperforms the state-of-the-art methods for the item cold-start issue.

## Related Work

Item cold-start issue can be further divided into two categories. The first category is the few-shot item cold-start issue, whose cold items have just a few historical interactions with users. To fully utilize these extremely sparse interactions, existing works are mainly based on meta-learning methods (Lee et al. 2019; Lu, Fang, and Shi 2020; Zhu et al. 2021; Sankar et al. 2021). For example, MeLU (Lee et al. 2019) uses support set and query set to calculate training loss and test loss, respectively. Then, the training loss is minimized in local update phase, while test loss is minimized in global update phase. MetaHIN (Lu, Fang, and Shi 2020) introduces heterogeneous information networks to the meta-learning method for cold-start recommendation from both model-level and data-level simultaneously.

In this paper, we focus on the second category, i.e., zero-shot item cold-start problem, where cold items are newly coming with no historical interactions. To address this problem, existing methods introduce the item’s multimedia information to obtain the content feature as the representation of the cold item (Volkovs, Yu, and Poutanen 2017; Du et al. 2020; Wang et al. 2021; Wei et al. 2021; Zhou, Zhang, and Yang 2023). Some of these methods integrate the item’s collaborative feature into its content feature, and attempt to infer the integrated feature via randomly corrupting the collaborative feature of a subset of warm items in the training stage from the perspective of robust learning (Volkovs, Yu, and Poutanen 2017; Du et al. 2020). Other methods, which consider the relation between the content feature and the collaborative feature from an item-level perspective, can represent cold items and warm items with content feature and collaborative feature, respectively. For example, PGD (Wang et al. 2021) learns a distilled output with the privileged embeddings from the teacher embeddings as the representation of cold items. GAR (Chen et al. 2022) generates an item’s content feature with a similar distribution as its collaborative feature, by training the generator and the recommender, adversarially. CLCRec (Wei et al. 2021) and CFCCRec (Zhou, Zhang, and Yang 2023) obtain the item’s content feature by maximizing the mutual information with its own collaborative feature, or another item’s collaborative feature with co-occurrence signal. However, these item-level alignment methods can not fully consider the condition of cold items in the training stage, resulting in limited performance.

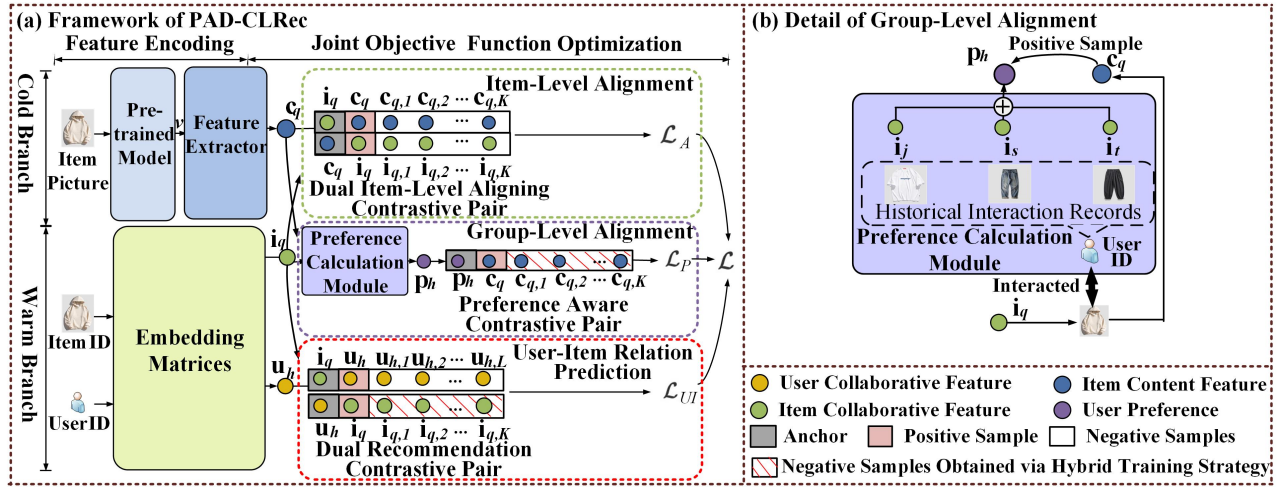


Figure 2: Overall framework of the proposed PAD-CLRec method. (a) The proposed PAD-CLRec includes two stages, i.e., feature encoding and joint objective function optimization, where feature encoding aims to obtain user’s collaborative feature, item’s collaborative feature and item’s content feature; and joint objective function optimization explores the relation among these three different kinds of feature from both item-level and group-level perspectives. (b) Details of Group-Level Alignment. Here,  $\mathbf{i}_q \in \mathbb{R}^{d \times 1}$  and  $\oplus$  denote item’s collaborative feature and aggregating calculation, respectively.

Different from these existing works, our PAD-CLRec not only fully considers the condition of cold items in the training stage, but also achieves a better trade-off between the performance of warm items and cold items from both item-level and group-level perspectives.

## Proposed Method

In this section, we introduce our proposed PAD-CLRec, which contains a feature encoding stage and a joint objective function optimization stage. The overall framework of the proposed method is given in Figure 2.

### Feature Encoding

In the feature encoding stage, our PAD-CLRec adopts a parallel two-branch structure (i.e., including a warm branch and a cold branch). The warm branch aims to obtain the user’s and the item’s collaborative feature via their historical interactions, and the cold branch is employed to extract the item’s content feature from its image input.

**Warm Branch:** To encode historical interactions, two trainable embedding matrices  $\mathbf{U} \in \mathbb{R}^{d \times M}$  and  $\mathbf{I} \in \mathbb{R}^{d \times N}$  are established as follows

$$\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_M], \mathbf{I} = [\mathbf{i}_1, \dots, \mathbf{i}_N] \quad (1)$$

where the  $h$ -th column of  $\mathbf{U}$  represents the user whose ID is  $h$ , and the  $q$ -th column of  $\mathbf{I}$  represents the item whose ID is  $q$ .  $M$  and  $N$  are the total number of users and items, respectively. When we input the user’s and item’s ID into the warm branch, it can output their corresponding ID-embedding as their collaborative feature from the embedding matrices via a lookup operation.

**Cold Branch :** To obtain the content feature  $\mathbf{c}_q \in \mathbb{R}^{d \times 1}$  of the item  $q$ , the cold branch contains a frozen pre-trained model and a trainable feature extractor. Firstly, the  $q$ -th

item’s picture is fed into the pre-trained model to obtain a  $f$  dimensional feature vector  $\mathbf{v}_q \in \mathbb{R}^{f \times 1}$ . Then, the feature extractor consisting of two trainable fully connected layers is used to further encode  $\mathbf{v}_q$  as follows

$$\mathbf{c}_q = \mathbf{W}_2 \cdot \delta(\mathbf{W}_1 \cdot \mathbf{v}_q + \mathbf{b}_1) + \mathbf{b}_2 \quad (2)$$

where  $\mathbf{W}_1 \in \mathbb{R}^{d \times f}$ ,  $\mathbf{W}_2 \in \mathbb{R}^{d \times d}$  and  $\mathbf{b}_1, \mathbf{b}_2 \in \mathbb{R}^{d \times 1}$  are linear mapping matrices and biases of the fully connected layers, respectively.  $\delta(\cdot)$  denotes the LeakyRelu function.

### Joint Objective Function Optimization

For model training, a joint objective function  $\mathcal{L}$  is proposed to achieve a better trade-off between the performance of warm items and cold items by exploring the relation between item’s content feature and collaborative feature, from both item-level and group-level perspectives, which can be formulated as follows

$$\mathcal{L} = (1 - \lambda)\mathcal{L}_{UI} + \lambda\mathcal{L}_A + \eta\mathcal{L}_P \quad (3)$$

where  $\mathcal{L}_{UI}$  is the dual recommendation contrastive loss that is used for user-item relation prediction.  $\mathcal{L}_A$  is the dual item-level aligning contrastive loss, which can maximize the mutual information between the same item’s collaborative feature and its content feature to achieve an item-level feature alignment.  $\mathcal{L}_P$  is the proposed preference aware contrastive loss, which is used to capture the relation between the content feature of a randomly selected item and the user’s preference for group-level feature alignment. Thereby, with the joint optimization, we can achieve a better trade-off between the recommendation performance of cold items and warm items. Here,  $\lambda$  and  $\eta$  are hyper-parameters.

Note that, these functions (i.e.,  $\mathcal{L}_{UI}$ ,  $\mathcal{L}_A$  and  $\mathcal{L}_P$ ) are based on contrastive learning, whose core is using a selected anchor vector to identify the positive sample from negative

samples. To achieve this, contrastive pair, containing an anchor vector, a positive sample, and negative samples, needs to be constructed for the contrastive loss calculation.

**User-Item Relation Prediction:** To predict the relation between the user and the item, we argue that if there is an interaction between the  $h$ -th user and the  $q$ -th item, the  $q$ -th item's collaborative feature  $\mathbf{i}_q \in \mathbb{R}^{d \times 1}$  should be more similar with the  $h$ -th user's collaborative feature  $\mathbf{u}_h \in \mathbb{R}^{d \times 1}$  than that of other items.

Let  $\mathcal{N}_q^{(U)}$  and  $\mathcal{N}_h^{(I)}$  denote the set of users that interacted with item  $q$  and the set of items that have been purchased by user  $h$ , respectively. According to this, a dual recommendation contrastive pair can be constructed, which contains two parts (i.e.,  $\mathcal{P}(u, i)$  and  $\mathcal{P}(i, u)$ ) for the dual recommendation contrastive loss calculation.  $\mathcal{P}(u, i)$  as the first part leverages  $\mathbf{u}_h$  as the anchor,  $\mathbf{i}_q$  as the positive sample, and the collaborative feature of items in set  $\hat{\mathcal{I}}_q = \{\mathbf{i}_{q,j} | \mathcal{N}_q^{(U)} \cap \mathcal{N}_j^{(U)} = \phi, j = 1, 2, \dots, K\}$  as the negative samples.  $\mathcal{P}(u, i)$  can be defined as follows

$$\mathcal{P}(u, i) = \{(\mathbf{u}_h, \mathbf{i}_q), (\mathbf{u}_h, \mathbf{i}_{q,1}), (\mathbf{u}_h, \mathbf{i}_{q,2}), \dots, (\mathbf{u}_h, \mathbf{i}_{q,K})\} \quad (4)$$

Meanwhile,  $\mathcal{P}(i, u)$  as the second part leverages  $\mathbf{i}_q$  as the anchor,  $\mathbf{u}_h$  as the positive sample, and the collaborative feature of users in set  $\hat{\mathcal{I}}_h = \{\mathbf{u}_{h,j} | \mathcal{N}_h^{(I)} \cap \mathcal{N}_j^{(I)} = \phi, j = 1, 2, \dots, L\}$  as the negative samples.  $\mathcal{P}(i, u)$  can be formulated as

$$\mathcal{P}(i, u) = \{(\mathbf{i}_q, \mathbf{u}_h), (\mathbf{i}_q, \mathbf{u}_{h,1}), (\mathbf{i}_q, \mathbf{u}_{h,2}), \dots, (\mathbf{i}_q, \mathbf{u}_{h,L})\} \quad (5)$$

Based on the dual recommendation contrastive pair,  $\mathcal{L}_{UI}$  is formulated as

$$\mathcal{L}_{UI} = \frac{1}{|\mathbb{O}|} \sum_{o(h,q) \in \mathbb{O}} \left( \beta \cdot \frac{e^{\mathbf{u}_h^\top \mathbf{i}_q}}{e^{\mathbf{u}_h^\top \mathbf{i}_q} + \sum_{\mathbf{i}_{q,j} \in \hat{\mathcal{I}}_q} e^{\mathbf{u}_h^\top \mathbf{i}_{q,j}}} + (1 - \beta) \cdot \frac{e^{\mathbf{i}_q^\top \mathbf{u}_h}}{e^{\mathbf{i}_q^\top \mathbf{u}_h} + \sum_{\mathbf{u}_{h,c} \in \hat{\mathcal{I}}_h} e^{\mathbf{i}_q^\top \mathbf{u}_{h,c}}} \right) \quad (6)$$

where  $o(h,q) \in \mathbb{O}$  denotes an interaction between user  $h$  and item  $q$  in the interaction set  $\mathbb{O}$ .

**Item-Level Feature Alignment:** To maximize the mutual information of the content feature and the collaborative feature of the same item, the item-level aligning contrastive pair is constructed, which also includes two parts (i.e.,  $\mathcal{P}(i, c)$  and  $\mathcal{P}(c, i)$ ).  $\mathcal{P}(i, c)$  applies the  $q$ -th item's collaborative feature  $\mathbf{i}_q$  as the anchor, its corresponding content feature  $\mathbf{c}_q \in \mathbb{R}^{d \times 1}$  as the positive sample, and the content feature of items in set  $\hat{\mathcal{C}}_q = \{\mathbf{c}_{q,j} | \mathcal{N}_q^{(U)} \cap \mathcal{N}_j^{(U)} = \phi, j = 1, 2, \dots, K\}$  as negative samples, thus  $\mathcal{P}(i, c)$  can be denoted as

$$\mathcal{P}(i, c) = \{(\mathbf{i}_q, \mathbf{c}_q), (\mathbf{i}_q, \mathbf{c}_{q,1}), (\mathbf{i}_q, \mathbf{c}_{q,2}), \dots, (\mathbf{i}_q, \mathbf{c}_{q,K})\} \quad (7)$$

Whereas  $\mathcal{P}(c, i)$  leverages the  $q$ -th item's content feature  $\mathbf{c}_q$  as the anchor, its corresponding collaborative feature  $\mathbf{i}_q$  as the positive sample, and collaborative features of items in set  $\hat{\mathcal{I}}_q$  as the negative samples, it can be defined as

$$\mathcal{P}(c, i) = \{(\mathbf{c}_q, \mathbf{i}_q), (\mathbf{c}_q, \mathbf{i}_{q,1}), (\mathbf{c}_q, \mathbf{i}_{q,2}), \dots, (\mathbf{c}_q, \mathbf{i}_{q,K})\} \quad (8)$$

Then based on the dual item-level aligning contrastive pair,  $\mathcal{L}_A$  can be calculated as

$$\mathcal{L}_A = \frac{1}{|\mathbb{O}|} \sum_{o(h,q) \in \mathbb{O}} \left( \alpha \cdot \frac{e^{\mathbf{i}_q^\top \mathbf{c}_q}}{e^{\mathbf{i}_q^\top \mathbf{c}_q} + \sum_{\mathbf{c}_{q,j} \in \hat{\mathcal{C}}_q} e^{\mathbf{i}_q^\top \mathbf{c}_{q,j}}} + (1 - \alpha) \cdot \frac{e^{\mathbf{c}_q^\top \mathbf{i}_q}}{e^{\mathbf{c}_q^\top \mathbf{i}_q} + \sum_{\mathbf{i}_{q,c} \in \hat{\mathcal{I}}_q} e^{\mathbf{c}_q^\top \mathbf{i}_{q,c}}} \right) \quad (9)$$

**Group-Level Feature Alignment:** To consider the condition of cold items, a warm item is randomly selected and treated as a cold item for model training. To exploit the relation between the selected item's content feature and a group of warm items' collaborative feature from a group-level perspective, the user's preference is introduced and used to construct the preference aware contrastive pair.

The preference  $\mathbf{p}_h \in \mathbb{R}^{d \times 1}$  of user  $h$  can be obtained by aggregating warm items' collaborative feature in set  $\mathcal{I}_{h-q} = \{\mathbf{i}_j | \forall \mathbf{i}_j \in \mathcal{N}_h^{(I)} \wedge \mathbf{i}_j \neq \mathbf{i}_q\}$ . For calculating  $\mathbf{p}_h$ , we give the following measures, including *Weighted Sum* and *Average*.

**Weighted Sum:** As different items often have different effects on a specific user's preference, the weighted sum can be adopted to calculate the user's preference as follows

$$\mathbf{p}_h = \frac{1}{|\mathcal{I}_{h-q}|} \sum_{\mathbf{i}_j \in \mathcal{I}_{h-q}} \text{softmax}\left(\frac{\mathbf{u}_h^\top \cdot \mathbf{i}_j}{\|\mathbf{u}_h\|_2 \cdot \|\mathbf{i}_j\|_2}\right) \cdot \mathbf{i}_j \quad (10)$$

where  $\text{softmax}(\cdot)$  is the softmax function to measure the importance of the item to the preference.

**Average:** To avoid the effects of the quality of user representation, we directly use the average of collaborative feature in set  $\mathcal{I}_{h-q}$  as the  $h$ -th user's preference  $\mathbf{p}_h$ , denoted as follows

$$\mathbf{p}_h = \frac{1}{|\mathcal{I}_{h-q}|} \sum_{\mathbf{i}_j \in \mathcal{I}_{h-q}} \mathbf{i}_j \quad (11)$$

After that, we assume that when an interaction between the  $h$ -th user and the  $q$ -th item occurs, the  $q$ -th item's content feature  $\mathbf{c}_q$  will be more similar to the  $h$ -th user's preference  $\mathbf{p}_h$  than other items' content feature in set  $\hat{\mathcal{C}}_q$ . Therefore, the preference aware contrastive pair can be built as follows

$$\{(\mathbf{p}_h, \mathbf{c}_q), (\mathbf{p}_h, \mathbf{c}_{q,1}), (\mathbf{p}_h, \mathbf{c}_{q,2}), \dots, (\mathbf{p}_h, \mathbf{c}_{q,K})\} \quad (12)$$

Then based on the dual preference aware contrastive pair constructed,  $\mathcal{L}_P$  can be calculated as

$$\mathcal{L}_P = \frac{1}{|\mathbb{O}|} \sum_{o(h,q) \in \mathbb{O}} \frac{e^{\mathbf{p}_h^\top \mathbf{c}_q}}{e^{\mathbf{p}_h^\top \mathbf{c}_q} + \sum_{\mathbf{c}_{q,j} \in \hat{\mathcal{C}}_q} e^{\mathbf{p}_h^\top \mathbf{c}_{q,j}}} \quad (13)$$

Furthermore, we adopt a hybrid training strategy following the study in (Wei et al. 2021), which can maximize the correlation among user's collaborative feature, item's collaborative feature, and item's content feature.

## Experiments and Results

### Dataset

We evaluate the proposed model on two real-world datasets including Amazon Rec dataset<sup>1</sup> and Amazon Fashion

<sup>1</sup><https://github.com/weiyinwei/CLCRec>

Task	Model	Amazon Rec						Amazon Fashion					
		@10		@20		@50		@10		@20		@50	
		Recall	NDCG	Recall	NDCG	Recall	NDCG	Recall	NDCG	Recall	NDCG	Recall	NDCG
<i>All</i>	BPR	0.88	0.43	1.40	0.56	2.60	0.80	7.76	7.33	8.12	7.42	8.78	7.56
	DropoutNET	0.91	0.37	1.40	0.45	2.54	0.60	22.67	6.92	23.12	5.47	23.98	4.39
	MTPR	0.96	0.49	1.56	0.64	3.15	0.96	23.56	17.30	24.86	18.76	26.72	19.02
	GAR	0.21	0.11	0.32	0.14	0.55	0.18	23.46	19.98	23.74	19.89	24.33	20.07
	CFCCRec	0.38	0.20	0.65	0.27	1.49	0.44	22.13	15.92	24.09	18.96	26.08	17.20
	CLCRec	1.95	1.10	2.95	1.35	4.88	1.74	26.83	24.52	27.64	20.88	28.75	24.89
	<b>PAD-CLRec</b>	<b>2.20</b>	<b>1.25</b>	<b>3.22</b>	<b>1.51</b>	<b>5.26</b>	<b>1.91</b>	<b>27.24</b>	<b>24.70</b>	<b>28.11</b>	<b>21.01</b>	<b>29.73</b>	<b>25.17</b>
	%Improv.	12.8%	13.6%	9.2%	11.9%	7.8%	9.8%	1.5%	0.7%	1.7%	0.6%	3.4%	1.1%
<i>Warm</i>	BPR	1.30	0.64	2.06	0.83	3.83	1.18	23.31	21.74	24.43	22.02	26.49	22.43
	DropoutNET	1.35	0.56	2.06	0.66	3.74	0.89	22.93	7.07	24.27	5.93	26.13	4.90
	MTPR	1.39	0.71	2.29	0.94	4.54	1.38	21.09	16.89	22.56	17.26	24.52	17.63
	GAR	0.38	0.20	0.57	0.25	1.17	0.36	22.06	20.89	22.71	21.05	23.79	21.27
	CFCCRec	1.22	0.69	1.88	0.86	3.33	1.15	23.21	21.57	24.24	21.84	26.12	22.22
	CLCRec	2.67	1.55	3.81	1.84	6.13	2.30	24.53	22.52	25.69	22.82	27.73	23.23
	<b>PAD-CLRec</b>	<b>2.85</b>	<b>1.63</b>	<b>4.18</b>	<b>1.97</b>	<b>7.12</b>	<b>2.55</b>	<b>25.85</b>	<b>23.55</b>	<b>27.39</b>	<b>23.95</b>	<b>29.72</b>	<b>24.42</b>
	%Improv.	6.7%	5.2%	9.7%	7.1%	16.2%	10.9%	5.4%	4.6%	6.6%	5.0%	7.2%	5.1%
<i>Cold</i>	BPR	0.10	0.04	0.12	0.05	0.36	0.09	0.12	0.05	0.18	0.06	0.40	0.11
	DropoutNET	0.06	0.03	0.08	0.03	0.29	0.07	23.75	7.18	26.43	6.43	27.70	5.10
	MTPR	0.92	0.49	1.62	0.67	3.26	0.99	26.50	21.68	27.68	21.44	29.42	22.31
	GAR	0.10	0.04	0.21	0.07	0.47	0.12	24.88	21.57	25.25	20.99	26.00	21.02
	CFCCRec	1.14	0.60	1.98	0.81	4.51	1.31	25.54	18.96	27.19	22.07	29.29	21.84
	CLCRec	2.99	1.62	4.45	1.99	7.36	2.58	28.94	26.72	29.74	26.84	31.11	27.13
	<b>PAD-CLRec</b>	<b>3.30</b>	<b>1.83</b>	<b>4.89</b>	<b>2.24</b>	<b>7.75</b>	<b>2.80</b>	<b>29.30</b>	<b>27.43</b>	<b>30.41</b>	<b>27.12</b>	<b>32.17</b>	<b>27.98</b>
	%Improv.	10.4%	13.0%	9.9%	12.6%	5.3%	8.5%	1.2%	2.7%	2.3%	1.0%	3.4%	3.1%

Table 1: Performance comparisons ( $\times 10^{-2}$ ) with different cold-start recommendation models on Amazon Rec and Fashion dataset. %Improv. is the percentage by which our PAD-CLRec improves the performance of the best baseline methods.

dataset<sup>2</sup>. Both datasets not only provide user-item interaction records, but also give images of items. In order to show the recommendation performance in more detail in the case of cold items, warm items, and a mix of cold and warm items, we set up three different tasks (i.e., *Cold*, *Warm* and *All*), respectively. For the cold-start task, we randomly select 20% items as cold items. In which, 50% interactions of these cold items are randomly selected as the *Cold* validation set, with the remainder interactions as *Cold* test set. Whereas, the rest 80% items are used as warm items. These warm item’s interactions are divided into three parts, with 80% as the training set, 10% as the *Warm* validation set and the rest 10% as the *Warm* test set. In addition, an extra *All* validation (test) set is built by combining the *Warm* and *Cold* validation (test) sets.

### Baseline Methods

In order to illustrate the effectiveness of our PAD-CLRec, we compare it with following state-of-the-art models on both Amazon dataset and Amazon Fashion dataset:

- **BPR** (Rendle et al. 2012): BPR learns collaborative feature (i.e. ID-embedding) to represent the user and the item according to their interactions.
- **DropoutNet** (Volkovs, Yu, and Poutanen 2017): DropoutNet considers the condition of cold items in the training phase via randomly discarding partial warm items’ collaborative feature.

- **MTPR** (Du et al. 2020): MTPR replaces the item’s collaborative feature with the all-zero vector to simulate cold-start scenario via counterfactual thinking.
- **CLCRec** (Wei et al. 2021): CLCRec based on contrastive learning method can generate the item’s content feature by maximizing its mutual information with its corresponding collaborative feature.
- **GAR** (Chen et al. 2022): GAR generates an item’s content feature with a similar distribution to its corresponding collaborative feature, via adversarially training the generator and the recommender.
- **CFCCRec** (Zhou, Zhang, and Yang 2023): CFCCRec obtains an item’s content feature by contrasting the content feature with the collaborative feature of another item with co-occurrence signal.

### Performance Metrics

We adopt recall@K (Wang et al. 2019; He et al. 2017) and Normalized Discounted Cumulative Gain (NDCG@K) (i.e., K=10, 20, 50) (Lee et al. 2019) as the metrics to evaluate the performance of our model, which are all widely used in recommender systems.

### Implementation Details

In our experiments, the Xavier algorithm (Glorot and Bengio 2010) is utilized for parameters initialization. Adam optimizer (Kingma and Ba 2014) is adopted for model optimization with the learning rate of  $1e - 2$ , and the batch size

<sup>2</sup><http://jmcauley.ucsd.edu/data/amazon/>







## Acknowledgments

This work was partly supported by the National Key Research and Development Project (No. 2021YFF0901602), the National Natural Science Foundation of China (No. 62176074), the Interdisciplinary Development Program of Harbin Institute of Technology (No. SYL-JC-202203), and the Fundamental Research Funds for the Central Universities (No.2022FRFK0600XX).

## References

- Chen, H.; Wang, Z.; Huang, F.; Huang, X.; Xu, Y.; Lin, Y.; He, P.; and Li, Z. 2022. Generative adversarial framework for cold-start item recommendation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2565–2571.
- Du, X.; Wang, X.; He, X.; Li, Z.; Tang, J.; and Chua, T.-S. 2020. How to learn item representation for cold-start multimedia recommendation? In *Proceedings of the 28th ACM International Conference on Multimedia*, 3469–3477.
- Glorot, X.; and Bengio, Y. 2010. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*, 249–256.
- Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2020. Generative adversarial networks. *Communications of the ACM*, 63(11): 139–144.
- He, X.; Deng, K.; Wang, X.; Li, Y.; Zhang, Y.; and Wang, M. 2020. LightGCN: Simplifying and powering graph convolution network for recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 639–648.
- He, X.; Liao, L.; Zhang, H.; Nie, L.; Hu, X.; and Chua, T.-S. 2017. Neural collaborative filtering. In *Proceedings of the 26th International Conference on World Wide Web*, 173–182.
- Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Lee, H.; Im, J.; Jang, S.; Cho, H.; and Chung, S. 2019. MeLU: Meta-learned user preference estimator for cold-start recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1073–1082.
- Lu, Y.; Fang, Y.; and Shi, C. 2020. Meta-learning on heterogeneous information networks for cold-start recommendation. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1563–1573.
- Rendle, S.; Freudenthaler, C.; Gantner, Z.; and Schmidt-Thieme, L. 2012. BPR: Bayesian personalized ranking from implicit feedback. *arXiv preprint arXiv:1205.2618*.
- Sankar, A.; Wang, J.; Krishnan, A.; and Sundaram, H. 2021. ProtoCF: Prototypical collaborative filtering for few-shot recommendation. In *Proceedings of the 15th ACM Conference on Recommender Systems*, 166–175.
- Volkovs, M.; Yu, G.; and Poutanen, T. 2017. DropoutNet: Addressing cold start in recommender systems. *Advances in Neural Information Processing Systems*, 30.
- Wang, S.; Zhang, K.; Wu, L.; Ma, H.; Hong, R.; and Wang, M. 2021. Privileged graph distillation for cold start recommendation. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1187–1196.
- Wang, X.; He, X.; Wang, M.; Feng, F.; and Chua, T.-S. 2019. Neural graph collaborative filtering. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 165–174.
- Wei, Y.; Wang, X.; Li, Q.; Nie, L.; Li, Y.; Li, X.; and Chua, T.-S. 2021. Contrastive learning for cold-start recommendation. In *Proceedings of the 29th ACM International Conference on Multimedia*, 5382–5390.
- Wei, Y.; Wang, X.; Nie, L.; He, X.; Hong, R.; and Chua, T.-S. 2019. MMGCN: Multi-modal graph convolution network for personalized recommendation of micro-video. In *Proceedings of the 27th ACM International Conference on Multimedia*, 1437–1445.
- Wu, J.; Wang, X.; Feng, F.; He, X.; Chen, L.; Lian, J.; and Xie, X. 2021. Self-supervised graph learning for recommendation. In *Proceedings of the 44th international ACM SIGIR Conference on Research and Development in Information Retrieval*, 726–735.
- Yu, J.; Yin, H.; Li, J.; Wang, Q.; Hung, N. Q. V.; and Zhang, X. 2021. Self-supervised multi-channel hypergraph convolutional network for social recommendation. In *Proceedings of the Web Conference 2021*, 413–424.
- Zhou, Z.; Zhang, L.; and Yang, N. 2023. Contrastive Collaborative Filtering for Cold-Start Item Recommendation. *arXiv preprint arXiv:2302.02151*.
- Zhu, Y.; Xie, R.; Zhuang, F.; Ge, K.; Sun, Y.; Zhang, X.; Lin, L.; and Cao, J. 2021. Learning to warm up cold item embeddings for cold-start recommendation with meta scaling and shifting networks. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1167–1176.