Intent Disentanglement and Feature Self-Supervision for Novel Recommendation

Tieyun Qian, *Member, IEEE*, Yile Liang, Qing Li, *Fellow, IEEE*, Xuan Ma, Ke Sun, and Zhiyong Peng, *Member, IEEE*

Abstract—One key property in recommender systems is the long-tail distribution in user-item interactions where most items only have few user feedback. Improving the recommendation of tail items can promote novelty and bring positive effects to both users and providers, and thus is a desirable property of recommender systems. Current novel recommendation methods over-emphasize the importance of tail items without differentiating the degree of users' intent on popularity and often incur a sharp decline of accuracy. Moreover, none of existing studies has ever taken the extreme case of tail items, i.e., cold-start items without any interaction, into consideration. In this work, we first disclose the mechanism that drives a user's interaction towards popular or niche items by disentangling her intent into conformity influence (popularity) and personal interests (preference). We then present a unified end-to-end framework to simultaneously optimize accuracy and novelty targets based on the disentangled intent of popularity and that of preference. We further develop a new paradigm for novel recommendation of cold-start items which exploits the self-supervised learning technique to model the correlation between collaborative features and content features. We conduct extensive experiments on three real-world datasets. The results demonstrate that our proposed model yields significant improvements over the state-of-the-art baselines in terms of the trade-off between accuracy and novelty.

Index Terms—Disentangled representation, novel recommendation, self-supervised learning, recommender systems

1 Introduction

With the rapid growth of the Web, the users are overwhelmed with the choice of "finding the right thing" from a vast number of products. The recommender systems, which use historical data to infer the users' preference on particular items like movies, commodities, and places, are a crucial component of many e-commerce platforms. One key property in recommender systems is the long-tail distribution in user-item interactions, where a tiny number of popular items receive most of the user attention and a high proportion of tail items only have few user feedback. Improving the recommendation of tail items can enrich users' experience by providing them with more chances to find interesting yet unpopular items. It can also bring positive effects to the providers since the niche products may increase the companies' marginal profits. As a result, novel recommendation becomes a desirable property of modern recommender systems; this, however, is a challenging task due to the conflict between accuracy and novelty.

Most of existing studies on novel recommendation [1], [2], [3], [4], [5], also known as the long-tail recommendation in the literature ,[1] adopt a two stage re-ranking approach. At the first stage, the recommender systems aim to achieve a high accuracy by using a base model. At the second stage, the results from the first stage are re-ranked towards the novelty target by introducing more tail items into the candidate top-N item list. Such approaches have an inherent limitation, i.e., the popularity bias that the recommenders incline to popular items much more than tail ones still exists at the first stage.

Several recent studies [6], [7], [8] propose the end-to-end framework where the recommendation is optimized for both the accuracy and novelty at the same time. While making progress, the PPNW and TailNet methods [6], [7] tend to recommend tail items and result in a relatively low accuracy. More importantly, none of the re-ranking and the end-to-end based novel recommendation methods differentiate the degree of users' intent to popularity. Indeed, though the consumers often refer to others for product choice and will comply with the group norm to some extent, an individual's conformity is affected by many factors like intelligence, personality, and status [9], [10]. That is to say, facing the same popular item, different consumers may react in different directions. Some users tend to conform the others' action and purchase the popular item, while others may or may not undertake the same action as they are more likely to meet their own personal interests.

In view of the different levels of conformity among users, we propose to disentangle a user's intent into conformity influence and personal interests, which drives her interaction

- *Tieyun Qian, Yile Liang, Xuan Ma, Ke Sun, and Zhiyong Peng are with the School of Computer Science, Wuhan University, Hubei 430072, China. E-mail: {qty, liangyile, yijunma0721, sunke1995, peng}@whu.edu.cn.*
- *Qing Li is with Hong Kong Polytechnic University, Hong Kong. E-mail: qing-prof.li@polyu.edu.hk.*

Manuscript received 30 June 2021; revised 19 Apr. 2022; accepted 12 May 2022. Date of publication 19 May 2022; date of current version 15 Sept. 2023.
This work was supported in part by NSFC Projects under Grants 61572376, U1811263, 62032016, and 61972291 and in part by Hong Kong Research Grants Council through a Collaborative Research Fund under Grant C1031-18 G. (Corresponding author: Tieyun Qian.)
Recommended for acceptance by L. Chen, X. Zhou, X. Yang, and T. Sellis.
Digital Object Identifier no. 10.1109/TKDE.2022.3175536

1. We will use the terms "novel recommendation" and "long-tail recommendation" to describe the existing work interchangeably.

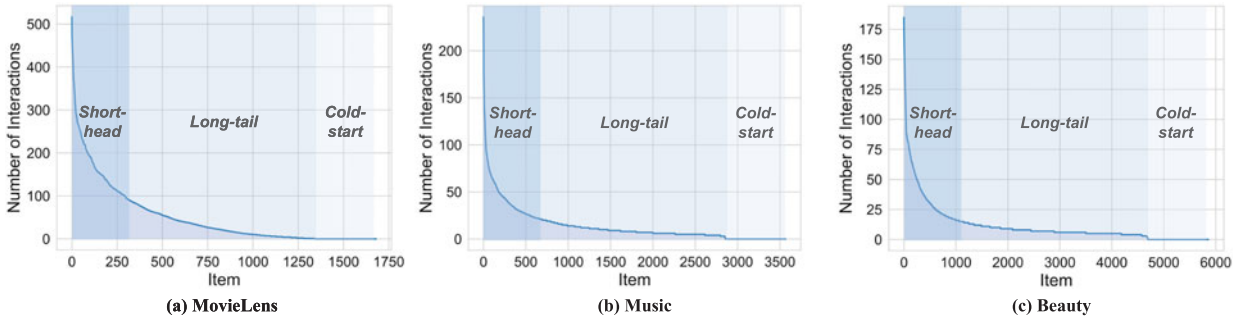


Fig. 1. Long-tail distribution of the items' interaction frequencies on three datasets.

towards popular and novel niche items, respectively. By doing this, the users' great diversity of intent can be uncovered to enhance the expressiveness of users' representations. The same operation is performed on the item side by disentangling items' representation into popularity and characteristic factors. We then present a unified end-to-end framework to simultaneously optimize the accuracy and novelty targets based on the learned user and item representations. In this way, the intrinsic popularity and preference factors are introduced into the novel recommendation process, and the system can naturally achieve a balance between the recommendation of tail items and that of popular items.

We further make in-depth analyses on the relationship between the normal long-tail items and the cold-start ones. As shown in Fig. 1, the cold-start items are actually the extreme case of long-tail items which do not have any interactions. Kapoor *et al.* [11] point out in their pioneering work that an item can be novel in three ways: (1) it is new to the system and thus for every user (cold-start), (2) it is known to the system but new to the single user, (3) it is known to the user long before but forgotten at the moment. Unfortunately, existing studies [1], [2], [3], [4], [5], [6], [7], [8] focus on novel recommendation tasks under the above definitions (2) (3), and none of them has ever taken the cold-start items under def. (1) into consideration. Meanwhile, previous research on cold-start recommendation [12], [13], [14], [15], [16], [17], [18] does not include long-tail items, either.

In this paper, we develop a new paradigm for novel recommendation of cold-start items. To address the problem of missing collaborative features for cold-start items, we exploit two types of self-supervised learning technique, including the variational autoencoder and mutual information maximization, to model the correlation between collaborative features and content features. To the best of our knowledge, this is the first attempt to exploring both the known long-tail and the unknown cold-start items in novel recommendation.

Extensive experimental results on three datasets demonstrate that our proposed model yields significant improvements over the state-of-the-art baselines in terms of the overall trade-off among accuracy, coverage, and novelty metrics on novel recommendation task for both the standard long-tail and the extended cold-start items.

2 Related Work

In this section, we review the related work from three perspectives, namely novel recommendation, disentangled representation learning, and self-supervised learning.

2.1 Novel Recommendation

Standard Long-Tail Recommendation. Current research on long-tail recommendation can be categorized into two groups. The first group designs different re-ranking methods, by post-processing the ranking list of a standard model to account for additional objectives like coverage rather than devising a new model. Some studies [1], [4], [5] improve novelty by countering the effects of item popularity, and others [2], [3] propose clustering approaches and leverage tail items directly in a recommendation list. The second group adopts the end-to-end models. For example, Lo *et al.* [6] propose a personalized pairwise novelty weighting for BPR loss function as an end-to-end method. Liu and Zheng [7] present a network architecture for long-tail session-based recommendation by introducing an adjustable preference mechanism. Zhang *et al.* [8] transfer knowledge from head items to tail items for leveraging the rich user feedback in head items and the semantic connections between head and tail items.

Overall, both groups of long-tail recommendation models ignore the users' different levels of conformity. In contrast, we disentangle users' intent into popularity and preference embeddings, so as to capture the inherent factors that determine users' choice on popular or niche items.

Backbone Recommendation Methods. Our proposed method is a general framework and can be added upon various backbone methods to enhance the performance of novel recommendation. In particular, we choose three classical learning methods with either point-wise loss including LFM [19] and NCF [20] or pair-wise loss like CML [21] as the backbone. LFM is a matrix factorization method which learns the latent factors by alternating least squares. NCF leverages deep neural networks to model latent features of users and items and utilizes a multi-layer perceptron to endow modelling with a high level of non-linearities. CML introduces the distance metrics to help encode not only users preferences but also the user-user and item-item similarity.

Cold-Start Recommendation. Early cold-start recommendation methods [22], [23], [24] mainly exploit side information as regularization in MF objective function [22] or adopt similarity based or feature mapping technique for integrating side information. More recent studies design various types of neural models to incorporate side information, including graph-based neural methods [12], [13], [18], the model based meta-learning methods [25] and the gradient based meta-learning method [26], and the active learning scheme [14], [15].

While these methods achieve promising performance for cold-start recommendation, they are not designed for recommending long-tail items. For example, DropoutNet [27]

and STAR-GCN [16] simulate the cold-start scenario by masking and reconstructing a part of input features in training without changing the item representations. Such methods cannot be adapted to the novel recommendation. One of our contributions is to view the cold-start items as the extreme case of tail items, and present a new learning paradigm for cold-start novel recommendation.

2.2 Disentangled Representation Learning

Disentangled representation learning originates from the field of computer vision [28], [29]. It aims to learn representations that separate explanatory factors of variations behind the data to improve the robustness and interpretability.

With the superior performance, disentangled representation learning sheds new light on recommender systems. MacridVAE [30] is the first work that introduces the disentangled representation learning into user behavior data at both a macro and a micro level. Inspired by it, DICER [31] combines the content information into the procedure of disentanglement. Besides, a few studies [32], [33] apply disentangled graph convolutional networks to reflect the fine-grained latent intents. Ma *et al.* [34] propose to disentangle the intents behind any given sequence of behaviors for sequential recommendation.

All the aforementioned methods need to define K latent intents/channels before performing disentangling, and the comprehension of these latent representations needs post-hoc explanation. A seminal work DICE [35] also presents to disentangle user interest and conformity for recommendation. However, the aims, as well as the idea and the technique for addressing the problems in DICE and our paper, are rather different. DICE is proposed for eliminating bias in causal recommendation while our model is for novel recommendation. Moreover, it adopts the causal model to learn corresponding embeddings which only involves the user-item interaction behaviors. In contrast, we propose a unified framework of feature self-supervision and intent disentanglement for utilizing the attributes besides the interaction behaviors, which tallies well with consumer conformity theory [9], [10], [36] since the behaviors are driven by the users'/items' inherent attributes.

2.3 Self-Supervised Learning

Self-supervised learning models can leverage input data itself as supervision and benefit almost all types of downstream tasks. The objectives in self-supervised learning can be categorized into generative, contrastive, and adversarial types [37] such as auto-encoding [38], [39], VAE variants [40], [41], discriminative models [42], and adversarial self-supervised learning [43].

As the research of self-supervised learning is still in its infancy, there are only several studies incorporating it into recommender systems [34], [44], [45], [46], [47], [48]. Some of these efforts mine the self-supervision signals from sequential data [34], [46], [48] and others capture the structure properties from user-item bipartite graphs [44], [47] or user-user social graphs [45]. Different from the above approaches, our work is the first to consider the correlations between the collaborative features in interaction behaviors and the user's/item's inherent content features, and we perform the generative and contrastive self-supervised learning jointly under these two types of features to tackle the cold-start problem in recommendation.

3 Proposed Model

In this section, we introduce the proposed model by first formulating the problem and then presenting the details.

3.1 Problem Formulation

Let $U = \{u_1, u_2, \ldots, u_M\}$ be a set of users and $V = \{v_1, v_2, \ldots, v_N\}$ be a set of items, where M and N denote the corresponding cardinalities. Let $\boldsymbol{R} \in \mathbb{R}^{M \times N}$ be the user-item interaction matrix, indicating whether the user purchased or clicked on the item. We focus on the recommendation task with implicit feedbacks [49], where the interaction matrix is defined as

$$R_{uv} = \begin{cases} 1, \text{ if an interaction (user } u, \text{ item } v) \text{ is observed,} \\ 0, \text{ otherwise.} \end{cases}$$

The observed entries reflect users' interest in the item, while the unobserved entries are mixed with unknown data and negative views of the item. Given the above interaction information, our goal is to estimate users' interest for unobserved entries and rank the candidate items according to the predicted scores, such that we can recommend the top-N items for the traditional recommendation task. Beyond that, we are particularly interested in novel recommendation for both long-tail and cold-start items.

Recent studies show that the item space can be divided into two parts, namely, popular short-head items and niche long-tail items. The long-existing but unpopular tail items have high novelty, and they can provide more surprises for users and more profits for providers. Besides, the cold-start items without any interaction can be regarded as the extreme case of tail items. The cold-start scenario is commonly found in the real world, e.g., the newly released movies have no audience. These new items can stimulate users' interest and avoid over-specialization in recommendation. Thus we extend the definition of standard long-tail recommendation to cover cold-start items.[2]

The *main target for novel recommendation* is to optimize the trade-off between accuracy and novelty [5], [6], [7]. On one hand, the recommendation accuracy is evaluated by a matching score between the recommendation list and the ground-truth list. On the other hand, the recommendation quality is measured by the proportion of novel items and the coverage of item space of the list.

3.2 Overview

Due to the different popularity/preference degrees of users and those of items, the users have specific intents to purchase or watch different items, and vice versa. Our goal is to find out users' intrinsic intents for different items and eliminate the dominance of popular items, such that we can select the relevant long-tail items and cold-start items for target users to improve the novelty of recommendation. To

2. We add attributes as part of the input to solve the problem of missing collaborative features for cold-start items.

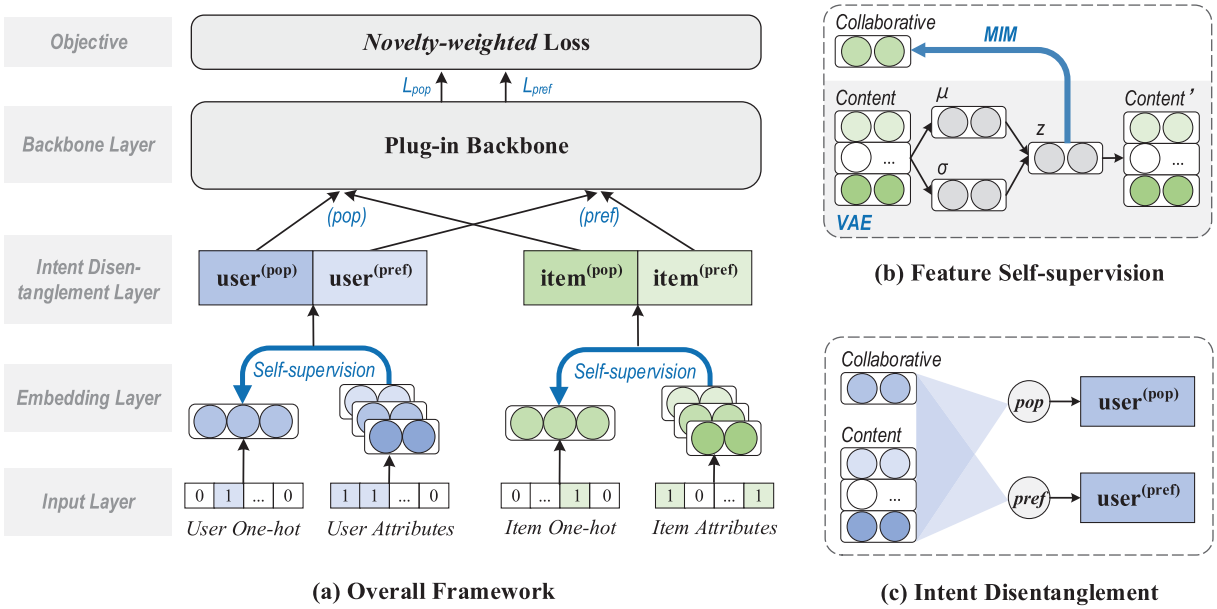


Fig. 2. An overview of our IDS4NR model. (a) Overall framework. (b) Feature self-supervision module for alleviating the cold-start problem. (c) Intent disentanglement module for extracting the specific factors.

this end, we propose an intent disentanglement and feature self-supervision (IDS4NR) model.

The architecture of our model is shown in Fig. 2a. It consists of four layers. We first present an input and embedding layer, which takes users'/items' unique one-hot ID encoding and their corresponding attributes as input and transform them into low-dimensional latent vectors. We then develop a self-supervision module to correlate the features from different views to tackle the cold-start problem. Next, we design an intent disentanglement layer to disentangle the latent features into popularity and preference factors. Finally, we introduce a plug-in backbone layer to model the interactions between the user and the item based on the disentangled representations, which can adopt different base models to make the framework generic.

3.3 Model Architecture

3.3.1 Input and Embedding Layer

Initial collaborative features. We set up a lookup table to transform the one-hot representation of each user and item into low-dimensional vector. After transformation, $\boldsymbol{p}_u \in \mathbb{R}^D$ and $\boldsymbol{q}_v \in \mathbb{R}^D$ are the latent factor representations of the user u and the item v, respectively. We denote them as collaborative features because they are learned from user-item interaction behaviors.

Initial content features. Besides the interactions, each user/item is equipped with side information that describes her/its own characteristics, such as attribute information, text description, and image. To be general, we term them as content features. Specifically, in our scenario, each user or item is associated with a set of attributes from different fields. Below is an example of user attributes

$$\underbrace{[0,1]}_{gender}\ \underbrace{[1,0,0,\ldots,0]}_{age}\ \underbrace{[0,1,0,\ldots,0]}_{occupation}.$$

Similar to the transformation of collaborative features, we set up an attribute encoding matrix of users and items. The attribute encoding of a user u and an item v is denoted as $[\boldsymbol{a}_u^1, \boldsymbol{a}_u^2, \ldots, \boldsymbol{a}_u^k]$ and $[\boldsymbol{a}_v^1, \boldsymbol{a}_v^2, \ldots, \boldsymbol{a}_v^k]$, respectively, where $\boldsymbol{a}^i \in \mathbb{R}^D$.

3.3.2 Feature Self-Supervision Module

Newly released items can be a good way to surprise users. However, these cold-start items do not contain any historical interaction data, so the valuable collaborative features cannot be learned by the traditional methods. In our work, since we extend the standard long-tail recommendation to cold-start items, we have to face such *a missing collaborative feature problem*. Below we elaborate our new paradigm on solving this challenge.

The content information of the new item has been proven to be important in solving the cold-start problem. Different from simply fusing the auxiliary content information into the item representation, we believe that it is critical to establish the correlation between the content features and the collaborative features for cold-start items. We term this as *the feature self-supervision* between two views of the target item. For example, animation movies are the mainstream entertainment among teenage children. The animation movies (content features) and the behaviors of their audience (collaborative features) have a strong correlation, while the love movies (other content features) and the behaviors of teenage children (collaborative features) have a weak correlation.

In light of this, we introduce an auxiliary self-supervision task to tackle the missing collaborative feature problem. In particular, we first set up the mapping between collaborative and content features, and then generate the pseudo collaborative features for cold-start items when making recommendations.

As shown in Fig. 2b, our self-supervision module has two key components: *variational autoencoder (VAE)* and *mutual information maximization (MIM)*. Both of them are used to establish the correlation between the behaviors of items and their related contents. Specifically, we adopt VAE to encode content features into the latent space, so that the latent

representation has the same distribution as the original one. We further maximize the mutual information between the generated collaborative features and the inherent content features in order to enhance the representation's discriminative ability.

VAE assumes that the data are generated from underlying latent representation. Given the data, the posterior distribution over a set of unobserved variables is approximated by a variational distribution. A typical VAE structure consists of a generation part and an inference part. Take the item v as an example, we denote its content features as the concatenation of all of its attribute encoding, i.e., $\boldsymbol{x}_v = \boldsymbol{a}_v^1 \oplus \ldots \oplus \boldsymbol{a}_v^k$. In the generation part, the reconstructed content features $\boldsymbol{x}'_v$ is generated from its latent variables z_v through a generation network like MLP parameterized by θ

$$\boldsymbol{x}'_v \sim p_\theta(\boldsymbol{x}'_v|\boldsymbol{z}_v). \tag{1}$$

In the inference part, the variational inference approximates the true intractable posterior of the latent variable $\boldsymbol{z}_v$ by introducing an inference network parameterized by ϕ [39]

$$q_\phi(\boldsymbol{z}_v) = \mathcal{N}(\boldsymbol{\mu}_v, diag(\boldsymbol{\sigma}_v^2)), \tag{2}$$

The objective of variational inference is to optimize the free variational parameters so that the KL-divergence $D_{KL}(q(z_v) \| p(\boldsymbol{z}_v \| \boldsymbol{x}_v))$ is minimized. With the reparameterization trick, we sample $\boldsymbol{\epsilon} \sim N(0, \boldsymbol{I})$ and reparameterize $\boldsymbol{z}_v = \boldsymbol{\epsilon}_\phi(\boldsymbol{x}_v) + \boldsymbol{\epsilon} \odot \sigma_\phi(\boldsymbol{x}_v)$. In this case, the gradient towards ϕ can be back-propagated through the sampled z_v. In summary, the loss function in VAE is defined as follows:

$$\mathcal{L}_{VAE} = -D_{KL}(q_\phi(\boldsymbol{z}_v|\boldsymbol{x}_v) \| p(\boldsymbol{z}_v)) + \mathbb{E}_{q_\phi(\boldsymbol{z}_v|\boldsymbol{x}_v)}[\log p_\theta(\boldsymbol{x}'_v|\boldsymbol{z}_v)]. \tag{3}$$

Mutual information (MI) is a basic concept in statistics, which measures the dependency between random variables. When encoding the content features into latent space, to increase the discriminative ability, we further strength the correlation between the latent content features and its corresponding collaborative features by using the MIM strategy.

Intuitively, the item v's latent content features $\boldsymbol{z}_v$ is more relevant to its own collaborative features $\boldsymbol{q}_v$ than those of other items. Therefore, we devise the MIM loss function as

$$\mathcal{L}_{MIM} = -\sum_{(a,v,v')\in\mathcal{O}} \log \sigma(f_D(\boldsymbol{z}_v, \boldsymbol{q}_v) - f_D(\boldsymbol{z}_v, \boldsymbol{q}'_v)), \tag{4}$$

where a, v, and v' are the anchor item's attributes, the positive item, and the negative item, respectively. We uniformly sample the item v' as the negative one, which does not contain any attributes of a from the entire item space. $f_D(\cdot)$ is the discriminator function that takes two vectors as the input and then scores the agreement between them. We simply implement it as the dot product between two representations to exclude additional parameters [45].

Finally, the whole loss function of the auxiliary feature self-supervision task can be defined as

$$\mathcal{L}_{SS} = \mathcal{L}_{VAE} + \mathcal{L}_{MIM}. \tag{5}$$

3.3.3 Intent Disentanglement Module

Recall that in the introduction part we analyze the users' different levels of conformity, which are ignored by existing novel recommendation methods. In this subsection, we present our intent disentanglement module to tackle this problem.

To distinguish the reasons why users interact popular or niche items, we set the intents as popularity and preference factors and separately model their contributions to the user-item interactions. In this way, we alleviate the dominance of popularity and increase the novelty naturally. Our proposed intent disentanglement module is shown in Fig. 2c, which has three unique properties:

- We present an intent prototype strategy which explicitly disentangles the intents into two factors (popularity and preference), rather than the parameterized K channels in other disentanglements methods [30], [31], [32], [33], [34].
- Both collaborative and content features take effects when disentangling intents, which tallies better with the conformity theory than previous casual model and negative sampling strategy [35].
- Our disentangled factors are naturally integrated into the end-to-end novel recommendation framework, where the popularity and preference factors account for the accuracy and novelty, respectively.

Intent prototype strategy. We take the user u as an example for illustration. Formally, the user u's overall feature list is denoted as $[\boldsymbol{a}_u^0, \boldsymbol{a}_u^1, \boldsymbol{a}_u^2, \ldots, \boldsymbol{a}_u^k]$, where $\boldsymbol{a}_u^0$ is equal to her collaborative feature $\boldsymbol{p}_u$ and the rests are k content features.

We start by defining the prototypical intent representations, i.e., the popularity prototype $\boldsymbol{c}^{(pop)} \in \mathbb{R}^D$ and the preference prototype $\boldsymbol{c}^{(pref)} \in \mathbb{R}^D$, which are part of model parameters and reflect the position of specific intent in the latent space. We then cluster the initial features according to their distances to two intent prototypes

$$p_u^{(pop)|i} = \frac{\exp(\boldsymbol{c}^{(pop)}\boldsymbol{a}_u^i)}{\exp(\boldsymbol{c}^{(pop)}\boldsymbol{a}_u^i) + \exp(\boldsymbol{c}^{(pref)}\boldsymbol{a}_u^i)}, \tag{6}$$

$$p_u^{(pref)|i} = \frac{\exp(\boldsymbol{c}^{(pref)}\boldsymbol{a}_u^i)}{\exp(\boldsymbol{c}^{(pop)}\boldsymbol{a}_u^i) + \exp(\boldsymbol{c}^{(pref)}\boldsymbol{a}_u^i)}, \tag{7}$$

where $i = 0, 1, 2, \ldots, k$. We adopt dot product to measure the similarity between a given input feature vector and an intent prototype.

Intent aggregation based on collaborative and content features. The clustering weight p describes the correlation between a specific feature vector and the intent prototype. We now aggregate all features and obtain the integrated representation of the user u under each disentangled intent

$$\boldsymbol{p}_u^{(pop)} = \mathrm{FFN}\left(\sum_{i=0}^{k} p_u^{(pop)|i} \cdot \boldsymbol{a}_u^i\right), \tag{8}$$

$$\boldsymbol{p}_u^{(pref)} = \mathrm{FFN}\left(\sum_{i=0}^{k} p_u^{(pref)|i} \cdot \boldsymbol{a}_u^i\right), \tag{9}$$

where $\mathrm{FFN}(\boldsymbol{x}) = \boldsymbol{W}_f\boldsymbol{x} + \boldsymbol{b}_f$ is a feed-forward network. The item v's popularity and characteristic representation $\boldsymbol{q}_v^{(pop)}, \boldsymbol{q}_v^{(pref)}$ can be obtained in the same way. So far, we have disentangled the user/item representation under specific intents. We will later explain how the disentangled intents contribute to the whole framework.

3.3.4 Backbone Layer

After obtaining the user and item representations, we take traditional recommendation methods as a plug-in backbone layer to model the relationship between the user and the item. Such a plug-in structure ensures the generality of our framework in enhancing recommendation novelty.

Traditional recommendation methods can be roughly grouped into two types: point-wise and pair-wise [50]. Note that some methods use the list-wise objective, but they are not widely adopted due to the expensive computational cost. *Point-wise objective.* The basic idea in point-wise methods is that they consider the individual user-item pair for ranking prediction. This kind of methods takes a triplet (u, v, y_{uv}) as the input, where $y_{uv} = 1$ indicates that the user u has interacted with the item v (implicit preference), otherwise $y_{uv} = 0$. The general form of the point-wise loss function can be defined as

$$\sum_{(u,v,y_{uv})\in\mathcal{O}} \mathcal{L}_{point}^{(intent)}(y_{ui}, \hat{y}_{ui}) + \lambda\Omega(\theta), \tag{10}$$

where $intent \in \{pop, pref\}$ and $\Omega(\theta)$ is a regularization term. A series of traditional methods adopt this kind of optimization strategy and we employ LFM [19] and NCF [20] as the representative base models. During this process, we feed the user/item representation under specific intents into the backbone layer, then obtain the predicted score $\hat{y}_{ui}$ and finally calculate the ranking loss according to these base models.

Pair-Wise Objective. The pair-wise approaches try to construct the set of item pairs by concerning their relative ordering. It is usually considered to be more suitable for optimizing the top-N recommendation since it compares the relevance of samples instead of getting one ranking score. This kind of methods inputs a triplet (u, v^+, v^-), where v^+, v^- are the implicitly liked and not observed item by the user u. The general form of the pair-wise loss function can be defined as

$$\sum_{(u,v^+,v^-)\in\mathcal{O}} \mathcal{L}_{pair}^{(intent)}(y_{uv^+v^-}, \hat{y}_{uv^+v^-}) + \lambda\Omega(\theta), \tag{11}$$

We choose the classic CML [21] as a representative base model of this kind of approaches.

3.4 Optimization and Learning

3.4.1 Novelty-Weighted Recommendation Loss

After feeding the disentangled user and item representations into the backbone layer, we can obtain the specific loss $\mathcal{L}^{(intent)}$ ($intent \in \{pop, pref\}$). However, if simply summing the losses without distinction, the learned representation cannot well reflect the impacts of the popularity and preference factors. Thus we design a novelty-weighted loss to control the impacts of different intents.

First, inspired by [6], the item novelty can be measured by its popularity degree, i.e., an item is more likely to be novel when more users have never interacted with it. The novelty score of the item v can be defined as

$$\alpha_v = \frac{\alpha'_v - \min(\alpha_{V'})}{\max(\alpha_{V'}) - \min(\alpha_{V'})} \left(\alpha'_v = \log\frac{|U|}{|U_v|}\right) \tag{12}$$

After that, the recommendation loss takes a weighted sum of two intents through the novelty score, where $\sigma(.)$ is sigmoid function for normalization

$$\mathcal{L}_{Rec} = \begin{cases} \alpha_v\mathcal{L}^{(pref)} + (1-\alpha_v)\mathcal{L}^{(pop)}, & \text{if point-wise,} \\ \sigma(\alpha_v^+ - \alpha_v^-)\mathcal{L}^{(pref)} + [1-\sigma(\alpha_v^+ - \alpha_v^-)]\mathcal{L}^{(pop)}, & \text{otherwise.} \end{cases} \tag{13}$$

Even if it is a common practice to regard the novelty score as weight, we argue that this is an effective way to align the intents to the popularity and preference factors.

3.4.2 Training on Disentangled Representations

Below we discuss the training strategy for point-wise and pair-wise objectives based on the disentangled user and item representations.

(1) For the point-wise learning, the model takes a triplet (u, v, y_{uv}) as input. Assuming that the novelty of the item v is high, e.g., $\alpha_v > 0.8$, the users' interaction to v is most probably due to her preference intent. Therefore, the loss $\mathcal{L}^{(pref)}$ has a larger weight and it would have more impacts on $\boldsymbol{p}_u^{(pref)}, \boldsymbol{q}_v^{(pref)}$. Conversely, the popularity intent becomes the leading factor when the user interacts with hot items.

(2) For the pair-wise learning, the model takes a triplet (u, v^+, v^-) as input. Assuming that the novelty of the item v^+ is higher than v^-, i.e., $\alpha_{v^+} > \alpha_{v^-}$, the user u is more likely to interact with the item v^+ than v^- for personal preference. Hence $\mathcal{L}^{(pref)}$ has a larger weight, and vice versa.

Finally, the overall loss of our IDS4NR framework contains the loss for the main task of novel recommendation and that for the auxiliary task of feature self-supervision

$$\mathcal{L} = \mathcal{L}_{Rec} + \gamma\mathcal{L}_{SS}, \tag{14}$$

where γ is a constant weighting factor.

4 Experiments

In this section, we conduct extensive experiments on three real-world datasets to validate our proposed IDS4NR model. We aim to answer the following research questions:

- *RQ1:* Does the proposed model outperform the state-of-the-art novelty promoting and cold-start oriented methods in our scenario?
- *RQ2:* How do different components such as the intent disentanglement module affect the results of our model?
- *RQ3:* How to comprehend the novelty factor in recommender systems, and what is the difference between our model and existing methods in recommendation novelty?
- *RQ4:* How about the computational cost of IDS4NR, i.e., is it easy to train and is it expensive?

4.1 Experimental Setup

4.1.1 Datasets

We use three publicly available datasets from different domains. *MovieLens*[3] is a widely adopted dataset in the application domain of movie recommendation, and we

3. https://grouplens.org/datasets/movielens/

TABLE 1
Statistics of the Datasets

Dataset	#Users	#Items	#Interactions	Sparsity
MovieLens	943	1682	100000	93.70%
Music	5541	3568	64706	99.67%
Beauty	8159	5863	98566	99.79%

employ the MovieLens-100 K version. *Music* and *Beauty* are chosen from Amazon.[4] Following the evaluation settings in [51], [52], we take the 5-core version for experiments, where each user or item has at least five interactions.

We use the users' and items' content information for addressing the missing collaborative feature problem. Specifically, for Movielens, the users' feature includes IDs, gender, age, and occupation, and the items' feature includes IDs, genres, directors, writers, stars, and countries. For Music and Beauty, the users' features include IDs and neighbors (we use the 10-nearest neighbors as the relation since there are no attribute information for users on these two datasets), and the items' features include IDs and categories.

In order to be consistent with the implicit feedback setting [20], [53], we transform the detailed rating into a value of 0 or 1, indicating whether a user has rated an item. The statistics of the datasets are shown in Table 1.

4.1.2 Evaluation Metrics

We evaluate the performance of novel recommendation in terms of accuracy, coverage, novelty, and the trade-off.

We adopt Recall@N (Rec@N) [5], [7], [21] as the accuracy metric, which considers whether the ground-truth is ranked amongst the top-N items (normally @5 or @10)

$$Rec@N = \frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}}\frac{|\mathcal{I}_u^{Te}\cap\mathcal{P}_u|}{|\mathcal{I}_u^{Te}|}, \tag{15}$$

where $\mathcal{I}_u^{Te}$ is the ground-truth list of the user u and $\mathcal{P}_u$ is u's top-N recommendation list.

Coverage@N (Cov@N) [2], [5], [7] is the ratio of the total number of distinct recommended items to the total items

$$Cov@N = \frac{|\cup_{u\in\mathcal{U}}\mathcal{P}_u|}{|\mathcal{I}|}. \tag{16}$$

Following the Pareto principle [54], the head items are the 20% most popular items, and the rest are long-tail novel items. Note that a recent study [55] presents a new definition about popularity which is against time. This is an interesting problem especially when the items are time-sensitive. We adopt the mainstream frequency based popularity in our experiment since we aim to develop a general approach that can be applied to any types of items regardless of their time sensitivity. We extend the NovAccuracy@N (Nov@N) [5], [6], [7] to measure how many novel items are in each top-N recommendation list

$$Nov@N = \frac{1}{N|\mathcal{U}|}\sum_{u\in\mathcal{U}}|\mathcal{I}_{Nov}\cap\mathcal{P}_u|, \tag{17}$$

where $\mathcal{I}_{Nov}$ is the novel item set including both long-tail and cold-start items.

Lastly, we employ F-score as the trade-off metric for the conflicting accuracy (recall), novelty, and coverage

$$F1@N = \frac{3 * Rec@N * Nov@N * Cov@N}{Rec@N + Nov@N + Cov@N}. \tag{18}$$

4.1.3 Baselines

To demonstrate the effectiveness of our proposed IDS4NR model, we compare it with the following state-of-the-art baseline methods. In addition, we adopt different classical base models as the backbone layer to show the generality of our framework.

Long-tail recommendation baselines:

- *GANC* [5] presents a re-ranking framework to integrate the learned user long-tail preference for accuracy, coverage, and novelty oriented recommendation.
- *PPNW* [6] proposes a loss weighting approach by leveraging users' and items' novelty information for end-to-end novel recommendation.
- *TailNet* [7] designs a preference mechanism to determine users' preference on popular or niche items in session-based long-tail recommendation. We remove its GRU encoder as it is irrelevant to our experiments.
- *MIRec* [8] is a dual transfer learning framework which collaboratively transfers knowledge from both model-level and item-level and from head items to tail items.

Cold-start recommendation baselines:

- *DropoutNet* [27] incorporates content and collaborative information with a neural model. It presents the dropout technique to tackle the cold-start issues.
- *HERS* [17] employs users social relations as user-user graph and builds item-item graph based on the common tags between two items. It aggregates user and item relations for addressing cold-start problem.
- *MetaEmb* [26] is a meta-learning approach to cold-start problem. It trains an embedding generator for new IDs through gradient-based meta-learning technique.
- *AGNN* [18] develops a graph neural network variant for attribute graph and designs an extended VAE structure for strict cold-start recommendation.

Representation Disentanglement baseline:

- *DICE* [35] learns disentangled representations for users' interest and conformity with the structural causal model and causal graph. Its goal is to maximize recommendation accuracy rather than the trade-off in novel recommendation.

Classical base models:

- *LFM* [19] is a classical matrix factorization method for collaborative filtering, which learns the latent factors by alternating least squares.
- *NCF* [20] presents a neural collaborative filtering method that combines multi-layer perceptron with generalized matrix factorization to encode non-linearities.

4. http://jmcauley.ucsd.edu/data/amazon/links.html

- *CML* [21] is a metric learning method, which encodes the user and item into a joint metric space and measures the user-item pair by the Euclidean distance.

4.1.4 Settings

To simulate the strict item cold-start scenario, we choose 20% items which have the latest average interactions chronologically as cold-start items. Note all their interactions are put into the testing set such that they are unseen during training. In addition, to be consistent with normal training scenario, we randomly select 10% of historical interactions for each user and add them into the testing set, and the remaining data are treated as the training set. During the test phrase, our evaluation protocol ranks all unobserved items in the training set for each user [5], [56].

For the baselines, we follow the same hyper-parameter settings if they are reported by the authors, and we fine-tune the hyper-parameters if they are not reported. For our IDS4NR, we set the batch size = 128, the initial learning rate = 0.001, and the number of max epoch = 40, 40, 30 for the base model LFM, NCF, and CML, respectively. We use Adam [57] as optimizer to self-adapt the learning rate. We sample 4 unobserved items as negative samples for each positive user-item pair [20]. The embedding dimension is set to 50 for our IDS4NR and all the baselines, to ensure the trainable parameters comparable.

4.2 Performance Comparison (RQ1)

The performance comparison between our proposed IDS4NR model and the baselines on three datasets is reported in Table 2. From the results, we have the following important observations.

First, it is clear that our IDS4NR significantly outperforms all the baselines in terms of the trade-off F1 scores of accuracy, coverage, and novelty on three datasets. The relative improvement of our model over the best baselines on three datasets are 32.46%, 45.35%, 32.32% on F1@5, and 24.21%, 31.46%, 26.69% on F1@10, respectively. It verifies the superiority of our proposed framework with the novelty-regulated intent disentanglement and collaborative-content feature self-supervision module. Also note that IDS4NR gets better accuracy than the baselines in many cases. This is because these baselines only involve the interaction features, and IDS4NR(+baselines) is a model which uses the additional content features. The comparison results between the models with the same extra information can be found in the ablation study below.

Second, compared with four long-tail recommendation baselines, our IDS4NR achieves superior performance in terms of accuracy and coverage. Among these four baselines, GANC, PPNW, and TailNet emphasize the impact of niche items through the re-ranking strategy, or the modified loss function and adjusted user representation, and they do not consider the inherent mechanism that drives users to interact with the items. Consequently, they reach high novelty but sacrifice the coverage of entire item space. Meanwhile, MIRec transfers the knowledge from head items to tail ones and is better than three other baselines, but it is still inferior to our model.

Third, compared with four cold-start recommendation baselines, our IDS4NR gets remarkable improvements on coverage and novelty while keeping a comparable accuracy performance. These baselines methods either learn a better representation for cold-start items by incorporating the neighbor information in the form of graph neural network (HERS and AGNN), or simulate and adapt to the cold-start scenario by dropout technique (DropoutNet) or meta-learning (MetaEmb), and thus accurately capture the user preference for cold-start items. However, they ignore the long-tail problem under normal circumstances and do not perform well in many cases of novel recommendation.

Fourth, the performance of the representation disentanglement method DICE is poor. We believe the main reason is that it is designed for causal recommendation which aims to maximize the accuracy metric like recall and hit ratio. Instead, the goal of novel recommendation is to balance the trade-off between the accuracy and novelty. Another reason might be due to the different disentanglement approaches. We adopt the intent prototype strategy and the intent is aggregated based on collaborative and content features, which conforms to the consumer conformity theory better than the causal model in DICE.

Finally, for the base models LFM, NCF, and CML, these traditional methods do not take into account the novelty factor, resulting in high accuracy but low coverage and novelty. However, when we integrate them as the backbone layer in our model, we can get dramatic increases on coverage, novelty, and trade-off F1 scores. For example, the relative improvement of F1@5 are 60.73%, 50.61%, and 78.80% on MovieLens, 31.44%, 68.10%, and 150.15% on Music, and 86.27%, 109.09%, and 156.86% on Beauty. These results prove that our IDS4NR can enhance the overall performance for various types of classical methods.

We also perform an additional analysis from the perspective of the dataset property. We can find that Beauty and Music are much sparser and they have more users and items than MovieLens from Table 1. AGNN exploits the attribute graph as side information. However, when the data is sparse, the relationship of attributes is not sufficient. HERS with additional social relationship information can make up for this deficiency and its accuracy is the best on Beauty and Music. In terms of novelty, PPNW proposes a personalized pairwise novelty weighting for BPR loss function while GANC groups the users according to their preference for long-tail novelty. When there are more users and items, GANC will be more targeted for novelty, and thus it outperforms PPNW on Beauty and Music. In terms of coverage, our method obtains the true intents of users through disentanglement, and thus we can promote long-tail items for the right group of users and significantly increase the system's coverage while maintaining accuracy and novelty.

4.3 Detailed Analysis on IDS4NR (RQ2)

4.3.1 Ablation Study

Given the appealing performance of IDS4NR, we further examine how the individual components in our model contribute to the improved performance on novel recommendation. Concretely, we conduct a series of ablation studies

TABLE 2
The Overall Performance Comparison on Three Datasets in Terms of Accuracy, Coverage, Novelty, and Trade-Off Evaluation, Respectively

Dataset	Model	Accuracy		Coverage		Novelty		Trade-off	
		Rec@5	Rec@10	Cov@5	Cov@10	Nov@5	Nov@10	F1@5	F1@10
MovieLens	GANC	0.0686	0.1123	0.3565	0.4194	0.4445	0.4506	0.0375	0.0648
	PPNW	0.0328	0.0562	0.4349	0.5062	**0.6937**	**0.6896**	0.0256	0.0470
	TailNet	0.0761	0.1247	0.3565	0.4497	0.2483	0.2386	0.0297	0.0493
	MIRec	0.0713	0.1260	0.3357	0.4212	0.4627	0.4598	0.0382	0.0727
	DropoutNet	0.0811	0.1387	0.3689	0.4385	0.2517	0.2614	0.0322	0.0569
	HERS	0.0855	0.1397	0.3701	0.4569	0.1862	0.2170	0.0275	0.0510
	MetaEmb	0.0826	0.1296	0.4159	0.4979	0.2430	0.2619	0.0337	0.0570
	AGNN	**0.0996**	**0.1598**	0.3184	0.4147	0.1756	0.2045	0.0281	0.0522
	DICE	0.0756	0.1271	0.4349	0.5187	0.3397	0.3418	0.0394	0.0684
	LFM	0.0888	0.1455	0.3891	0.4699	0.1974	0.2283	0.0303	0.0555
	IDS4NR(+LFM)	0.0870	0.1439	0.4414	0.5721	0.3870	0.3901	0.0487	0.0871
	NCF	0.0806	0.1274	0.3951	0.4842	0.2492	0.2745	0.0328	0.0573
	IDS4NR(+NCF)	0.0852	0.1386	0.4670	0.6155	0.3906	0.3853	0.0494	0.0865
	CML	0.0886	0.1376	0.3755	0.4533	0.1836	0.2121	0.0283	0.0494
	IDS4NR(+CML)	0.0752	0.1302	**0.5519**	**0.6613**	0.4296	0.4253	**0.0506**	**0.0903**
Music	GANC	0.0428	0.0581	0.6304	0.6657	**0.7515**	**0.7773**	0.0427	0.0601
	PPNW	0.0573	0.0911	0.4956	0.6010	0.5426	0.5356	0.0422	0.0717
	TailNet	0.0246	0.0383	0.5850	0.6789	0.7324	0.7004	0.0236	0.0385
	MIRec	0.0779	0.1297	0.4836	0.5942	0.5547	0.5312	0.0562	0.0979
	DropoutNet	0.1010	0.1411	0.2723	0.3600	0.1181	0.1853	0.0198	0.0411
	HERS	**0.1284**	**0.1864**	0.4934	0.6351	0.2733	0.2937	0.0580	0.0935
	MetaEmb	0.0894	0.1355	0.6990	0.7699	0.3633	0.3701	0.0591	0.0908
	AGNN	0.1121	0.1660	0.5707	0.6867	0.3169	0.3439	0.0608	0.0983
	DICE	0.0645	0.1007	0.4768	0.5934	0.3001	0.3467	0.0329	0.0597
	LFM	0.1147	0.1620	0.6203	0.7223	0.2534	0.2755	0.0547	0.0834
	IDS4NR(+LFM)	0.0891	0.1366	0.6853	0.8324	0.5006	0.4983	0.0719	0.1158
	NCF	0.0982	0.1419	0.6231	0.7321	0.2785	0.3073	0.0511	0.0810
	IDS4NR(+NCF)	0.0906	0.1373	**0.9030**	**0.9747**	0.5349	0.5243	**0.0859**	**0.1287**
	CML	0.0856	0.1257	0.5923	0.7433	0.1916	0.2093	0.0335	0.0544
	IDS4NR(+CML)	0.1123	0.1643	0.7363	0.8156	0.4331	0.4330	0.0838	0.1232
Beauty	GANC	0.0041	0.0081	0.4432	0.4862	**0.8850**	**0.8942**	0.0036	0.0076
	PPNW	0.0117	0.0189	0.4374	0.5415	0.5952	0.5920	0.0088	0.0158
	TailNet	0.0060	0.0089	0.5336	0.6358	0.7290	0.6713	0.0055	0.0086
	MIRec	0.0220	0.0370	0.4847	0.5954	0.5729	0.5533	0.0170	0.0309
	DropoutNet	0.0338	0.0510	0.1930	0.2754	0.0691	0.1214	0.0045	0.0114
	HERS	**0.0432**	**0.0650**	0.4680	0.6232	0.2375	0.2690	0.0192	0.0341
	MetaEmb	0.0272	0.0450	0.6929	0.7676	0.3885	0.3967	0.0198	0.0340
	AGNN	0.0263	0.0403	0.5377	0.6518	0.3124	0.3388	0.0151	0.0259
	DICE	0.0116	0.0199	0.2476	0.3486	0.08716	0.1143	0.0021	0.0049
	LFM	0.0242	0.0401	0.5558	0.6906	0.1978	0.2278	0.0102	0.0197
	IDS4NR(+LFM)	0.0270	0.0426	0.6029	0.7115	0.4006	0.4128	0.0190	0.0321
	NCF	0.0201	0.0317	0.6286	0.7264	0.3037	0.3319	0.0121	0.0210
	IDS4NR(+NCF)	0.0254	0.0413	**0.9026**	**0.9737**	0.5387	0.5363	0.0253	0.0417
	CML	0.0211	0.0314	0.6510	0.7963	0.2204	0.2396	0.0102	0.0168
	IDS4NR(+CML)	0.0293	0.0477	0.7919	0.8608	0.4958	0.4905	**0.0262**	**0.0432**

The best performance among all is in bold while the second best one is marked with an underline.

to investigate the effects of intent disentanglement and feature self-supervision.

- $\text{IDS4NR}_{\text{w/o-SS}}$: This variant removes the feature self-supervision module from the IDS4NR framework.
- $\text{IDS4NR}_{\text{w/o-SS, Exp}}$: Recall that our IDS4NR uses an item novelty score α_v to control the impacts of disentangled representation of popularity and preference intents. This variant is based on $\text{IDS4NR}_{\text{w/o-SS}}$ and further replaces α_v in Eq. (13) with 0.5 for treating all samples equally.
- $\text{IDS4NR}_{\text{w/o-SS, ID}}$: This variant removes the intent disentanglement module completely. To ensure utilizing the content information as the standard IDS4NR, we take the average of collaborative feature and all kinds of attributes of users/items as the input of backbone layer on the basis of $\text{IDS4NR}_{\text{w/o-SS}}$.
- $\text{IDS4NR}_{\text{w/o-SS, Attr}}$: This variant further removes the attribute inputs from $\text{IDS4NR}_{\text{w/o-SS}}$ to examine the impacts of attributes.

For clarity, we choose the CML backbone as the representative to perform the ablation and other parameter studies. We present the ablation results on the three datasets in Table 3.

First, we investigate the impacts of feature self-supervision module. We find that the overall trade-off performance

TABLE 3
Results for Ablation Study (Backbone - *CML*)

Dataset	Model	Accuracy		Coverage		Novelty		Trade-off	
		Rec@5	Rec@10	Cov@5	Cov@10	Nov@5	Nov@10	F1@5	F1@10
MovieLens	IDS4NR	0.0752	0.1302	**0.5519**	0.6613	0.4296	0.4253	**0.0506**	**0.0903**
	$\text{IDS4NR}_{\text{w/o-SS}}$	0.0793	0.1298	0.4295	0.5436	0.4025	0.4237	0.0451	0.0817
	$\text{IDS4NR}_{\text{w/o-SS, Exp}}$	0.0767	0.1303	0.4238	0.5303	0.3687	0.3825	0.0413	0.0760
	$\text{IDS4NR}_{\text{w/o-SS, ID}}$	**0.0966**	**0.1576**	0.3196	0.4260	0.2803	0.3101	0.0373	0.0699
	$\text{IDS4NR}_{\text{w/o-SS, Attr}}$	0.0534	0.0689	0.5222	0.5751	**0.6347**	**0.7436**	0.0439	0.0637
Music	IDS4NR	0.1123	0.1643	0.7363	0.8156	0.4331	0.4330	**0.0838**	**0.1232**
	$\text{IDS4NR}_{\text{w/o-SS}}$	0.1099	0.1591	0.7615	0.8814	0.4151	0.4350	0.0810	0.1230
	$\text{IDS4NR}_{\text{w/o-SS, Exp}}$	0.1087	0.1582	0.7329	0.8548	0.4018	0.4212	0.0772	0.1192
	$\text{IDS4NR}_{\text{w/o-SS, ID}}$	**0.1521**	**0.2198**	0.5362	0.6791	0.2406	0.2768	0.0633	0.1054
	$\text{IDS4NR}_{\text{w/o-SS, Attr}}$	0.0561	0.0771	**0.8380**	**0.9123**	**0.5105**	**0.5729**	0.0512	0.0774
Beauty	IDS4NR	0.0293	0.0477	0.7919	0.8608	0.4958	0.4905	**0.0262**	**0.0432**
	$\text{IDS4NR}_{\text{w/o-SS}}$	0.0260	0.0432	0.8040	0.9082	0.4703	0.4867	0.0227	0.0398
	$\text{IDS4NR}_{\text{w/o-SS, Exp}}$	0.0254	0.0410	0.7879	0.8995	0.4641	0.4804	0.0218	0.0374
	$\text{IDS4NR}_{\text{w/o-SS, ID}}$	**0.0478**	**0.0703**	0.4600	0.6336	0.2238	0.2781	0.0201	0.0378
	$\text{IDS4NR}_{\text{w/o-SS, Attr}}$	0.0137	0.0219	**0.8647**	**0.9345**	**0.5383**	**0.5975**	0.0135	0.0236

for $\text{IDS4NR}_{\text{w/o-SS}}$ declines on all three datasets compared with the standard IDS4NR. This indicates that the feature self-supervision component captures the relationship between the collaborative feature and content feature and models the user preference well for cold-start items, and thus it is an essential part of IDS4NR.

Next, we examine the impacts of intent disentanglement module. It is clear that both the coverage and novelty of $\text{IDS4NR}_{\text{w/o-SS, Exp}}$ decrease a lot if comparing it with $\text{IDS4NR}_{\text{w/o-SS}}$. The reason might be that the intent disentanglement without any constraint cannot learn the specific factors explicitly and will limit the ability of learning representations for users' intents on popular/niche items.

Following, the overall performance of the variant $\text{IDS4NR}_{\text{w/o-SS, ID}}$ decreases dramatically along with a few increased accuracy. After removing the intent disentanglement module, the model degrades into a naive content-based method, and focuses on the items having similar contents with users' historical interactions. With the limited recommendation scope, the accuracy of the model may increase a bit but the coverage and novelty performance become extremely poor.

Finally, the accuracy performance of the variant $\text{IDS4NR}_{\text{w/o-SS, Attr}}$ decreases dramatically though its coverage and novelty increases. These results prove the importance of an item's attributes on its accuracy. For example, a movie might be popular due to a certain director and/or an actor.

4.3.2 Parameter Study

We investigate the impacts of hyper-parameters in IDS4NR, including the dimension of the latent vector D and the weighting factor γ for feature self-supervision loss. We first fix γ to 0.01 and vary D, and then fix D to 50 and vary γ. The results of IDS4NR with CML as backbone layer for the varying parameter D and γ are shown in Figs. 3 and 4, respectively.

We can see from Fig. 3 that accuracy and novelty show opposite trends on all datasets. This is consistent with the problem setting and the results in previous studies. Movielens is a small dataset with relatively dense user-item interactions. The model can already learn a good feature representation even with a small dimensionality (D=10). In contrast, Music and Beauty are even sparser. The larger dimensionality improves the representation ability of features with more latent factors of users and items, and thus the model reaches a better accuracy, but its novelty declines at the same time. Hence we set D as 50 for a better balance in most cases.

Our IDS4NR takes feature self-supervision learning as an auxiliary task to enhance the collaborative features of cold-start items, and the weighting factor γ controls the proportion of self-supervision task. As we can see in Fig. 4, with the

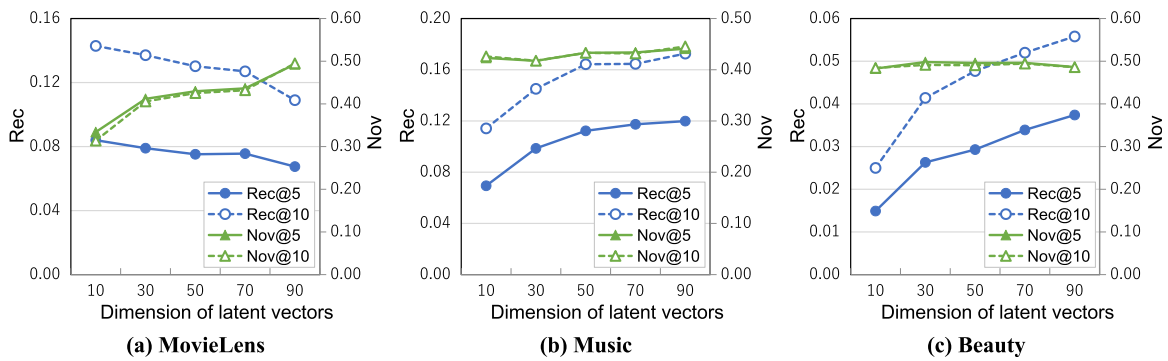


Fig. 3. Impacts of latent vector dimension D (backbone - *CML*).

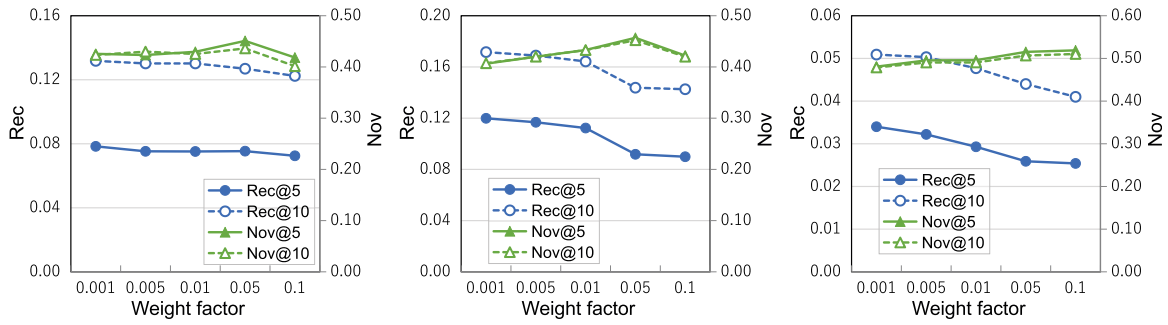


Fig. 4. Impacts of weighting factor γ (backbone - *CML*).

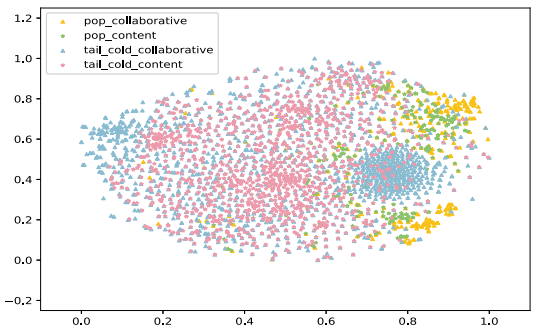
increase of γ, the accuracy on different datasets gradually decreases and the novelty increases, and this trend is more obvious when γ exceeds 0.01. When γ is small, it can improve the performance for cold-start items as an auxiliary task and thus enhances the recommendation novelty. However, when γ is too large, the auxiliary self-supervision task dominates the loss and interferes the learning of the main task, leading to a sharp decline of recommendation accuracy.

4.4 Visualization Analysis

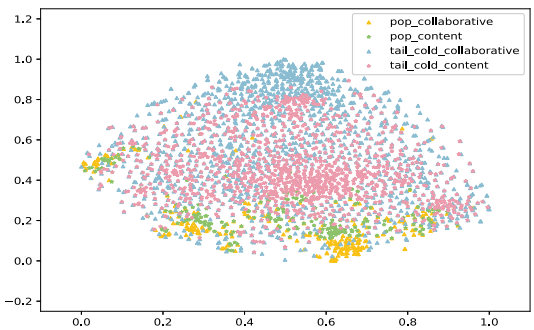
We conduct two types of visualization analysis using t-SNE [58]. One is to observe the representations of collaborative features and those of content features. The other is to observe the disentangled user and item representations.

We first visualize the embeddings of collaborative and content features. We take MovieLens dataset as an example and choose CML as the backbone. The results are shown in Fig. 5 where different features are shown in different colors.

We have two findings for Fig. 5. First, from the results in Fig. 5a before self-supervision, it is clear that two types of embeddings for head items (the green and yellow dots) match well in the figure. Meanwhile, those for tail/cold-start items (the blue and pink dots) are well separated because they don't have or have few collaborative features; both of these demonstrate the correlation between content features and collaborative features. Second, from the results in Fig. 5b after self-supervision, it can be seen that two types of embeddings for both head and tail/cold-start items become closer, showing the effectiveness of self-supervision module on strengthening such correlation.



(a) Before self-supervision



(b) After self-supervision

Fig. 5. Visualization of embeddings for collaborative and content features by IDS4NR.

We then visualize the disentangled embeddings of users' and items' popularity and preference intents. We take Music dataset as an example and choose CML as the backbone. The results are shown in Fig. 6 where different intents are shown in different colors.

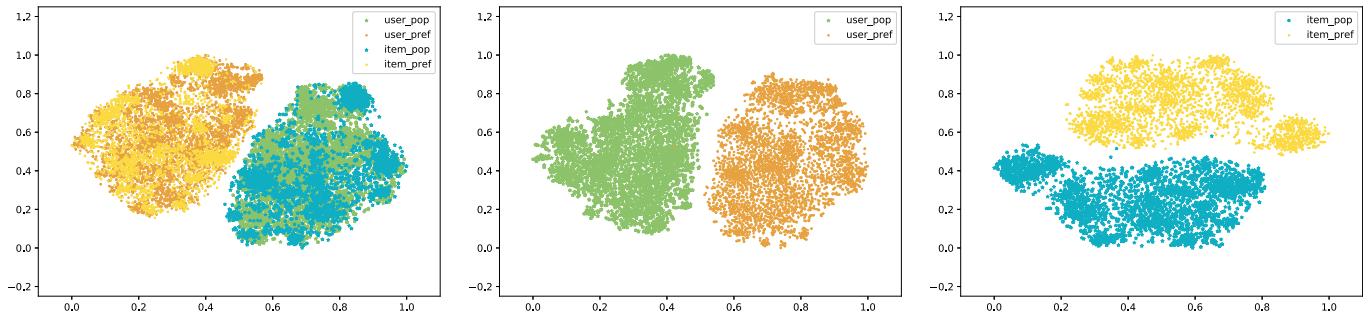
We also have two findings for Fig. 6. First, from the disentangled user and item representations in Fig. 6a, it is clear that the user's popularity is close to the item's popularity, and the same goes for the preference. That is to say, a user's interaction with the item is indeed based on the disentangled two types of intents. Second, from Figs. 6b and 6c, we can see that the disentangled representations of popularity and preference of a user or an item are well separated in both user and item embeddings. This indicates that our intent prototype strategy is effective in disentangling the intents into two factors (popularity and preference).

4.5 Comprehending Novelty in Recommendation (RQ3)

To have a close look at the difference between our proposed method and the existing studies, we select CML and MIRec as the representatives of the base model and the novel recommendation model for a thorough comparison.

From a macroscopic perspective (all users), we study the overall novelty distribution of the recommendation results of these methods. Specifically, following the novelty score defined in Equation (12), we plot the novelty distribution of the top-10 recommendation lists for all users of each method on Music dataset.

As shown in Fig. 7, among the three methods, most of the recommended items of CML fall in the range of low novelty. This is reasonable since CML is a traditional collaborative filtering method and it favors the popular items. MIRec is a novelty-promoting recommendation method and the vast majority of recommended items have medium to high novelty. However, it barely covers the popular items, which leads to the high novelty but low accuracy and coverage. In



(a) Disentangled embedding (b) Disentangled user embedding (c) Disentangled item embedding

Fig. 6. Visualization of disentangled embeddings by IDS4NR.

contrast, the recommended items of our IDS4NR are relatively evenly distributed in different novelty ranges and most of the items are in the middle level. Therefore, from the overall point of view, IDS4NR can achieve the best trade-off to recommend the popular, long-tail, and cold-start items.

From a micro perspective (individual user), we select two users and visualize their recommendation lists to analyze the difference among three methods. Fig. 9 shows the top-10 recommendation list of user-1771 and that of user-87 in the Music dataset.

For user-87, many recommended items by three methods have high novelty. There are four tail items (2436, 2166, 453, and 1465) in the recommendation list of CML, and all the recommended items of MIRec belong to long-tail. The coverage of IDS4NR is similar to that of CML, and is higher than that of MIRec. In addition, item-877 and item-2991 are recommended and hit by IDS4NR, where item-877 is a head item, and item-2991 is a cold-start one, indicating that our IDS4NR model has the ability to find out different types of relevant items, including the cold-start ones.

For user-1771, the recommended items of CML tend to have low novelty, while those of MIRec are mostly in the middle to high degree. Our IDS4NR can cover a great range of items with low or high novelty. It is noted that item-717 and item-759 are recommended and hit by IDS4NR, where item-717 and item-759 belongs to the head and the tail, respectively. In contrast, both of two hits in MIRec are tail items. This reveals that our model does not over-emphasize the novelty to conceal the popular items.

4.6 Analysis on Computational Cost (RQ4)

To investigate the training process of our model, we plot the model's training curves in Fig. 8. As can be seen, the loss declines quickly first and gradually slows down. It finally reaches the convergence when the training epoch increases, which proves that our model is stable and easy to train.

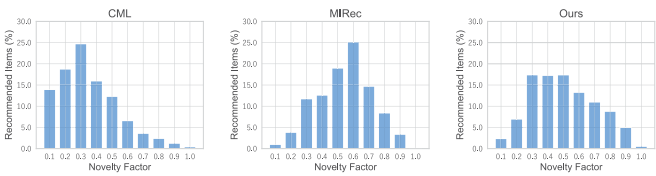


Fig. 7. The overall novelty distribution of recommended items w.r.t. different methods on Music.

We further compare the model's space complexity by presenting the trainable parameter number of our model and that of base models in Table 4.

The parameter number of our IDS4NR and its different base models mainly depends on the user/item embedding matrix. Compared with the base models, our framework introduces additional user/item attributes, so the increased cost lies primarily in the attribute embedding. For the intent disentanglement and self-supervision module in our framework, they need a few weight matrix at a lower cost. Note that NCF has the largest parameter scale on Music and Beauty. This is because it uses the independent embedding layer in its GMF and MLP module, which is equivalent to expanding the dimension of embedding vectors. When we take NCF as the backbone layer, these two modules are shared by the disentangled user/item representation and thus the parameter number decreases.

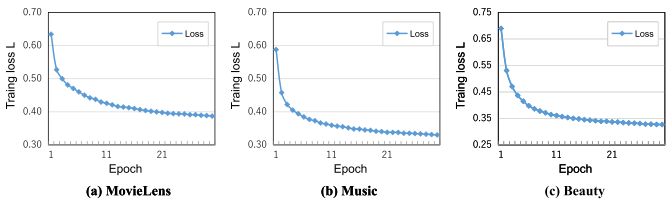


Fig. 8. Training curves of the overall loss. (backbone - *CML*).

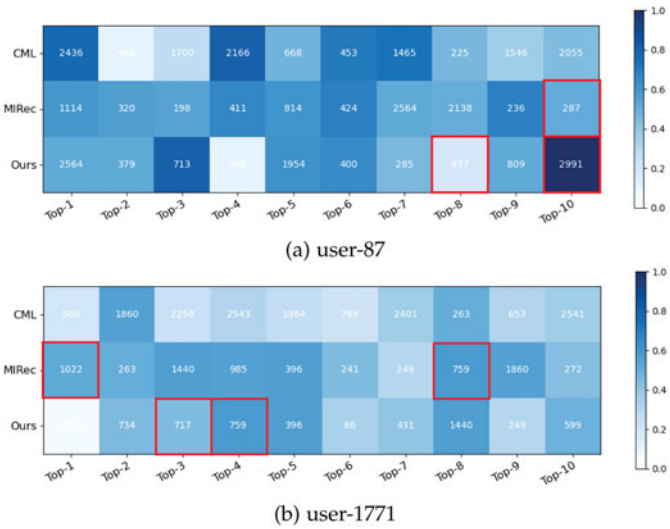


Fig. 9. Visualization of top-10 recommended items and their corresponding novelty score. The hit items are marked with red boxes.

TABLE 4
Parameter Number of Models on Three Datasets

Model	Model size ($\times 10^6$)		
	MovieLens	Music	Beauty
LFM	0.13	0.46	0.72
IDS4NR(+LFM)	0.57	0.78	1.19
NCF	0.28	0.93	1.42
IDS4NR(+NCF)	0.60	0.84	1.24
CML	0.13	0.46	0.70
IDS4NR(+CML)	0.52	0.76	1.16

5 Conclusion

In this paper, we propose a new model IDS4NR for novel recommendation. Our model is distinguished from existing novelty-oriented methods in two key issues. First, we realize that the varying degrees of users' intent to popular and niche items are the inherent factors in determining the trade-off between accuracy and novelty in novel recommendation. We then implement this idea with the disentangled intent integrated end-to-end framework. Second, we extend the definition of novel recommendation to cover the cold-start items. We also propose a self-supervision strategy to solve the missing collaborative feature problem which is a big obstacle when recommending cold-start items. To validate our approach, we conduct extensive comparison experiments and deep analysis studies on three real-world datasets. The results prove that our proposed IDS4NR model achieves a new state-of-the-art trade-off performance on the novel recommendation task for the long-tail and cold-start items.

References

[1] G. Adomavicius and Y. Kwon, "Improving aggregate recommendation diversity using ranking-based techniques," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 5, pp. 896–911, Mar. 2012.

[2] H. Yin, B. Cui, J. Li, J. Yao, and C. Chen, "Challenging the long tail recommendation," *Proc. VLDB Endowment*, vol. 5, no. 9, pp. 896–907, 2012.

[3] Y. Park, "The adaptive clustering method for the long tail problem of recommender systems," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 8, pp. 1904–1915, Aug. 2013.

[4] G. Oestreicher-Singer and A. Sundararajan, "Recommendation networks and the long tail of electronic commerce," *MIS Quart.*, vol. 36, no. 1, pp. 65–83, 2012.

[5] Z. Zolaktaf, R. Babanezhad, and R. Pottinger, "A generic top-N recommendation framework for trading-off accuracy, novelty, and coverage," in *Proc. IEEE Int. Conf. Data Eng.*, 2018, pp. 149–160.

[6] K. Lo and T. Ishigaki, "Matching novelty while training: Novel recommendation based on personalized pairwise loss weighting," in *Proc. IEEE Int. Conf. Data Mining*, 2019, pp. 468–477.

[7] S. Liu and Y. Zheng, "Long-tail session-based recommendation," in *Proc. 14th ACM Conf. Recommender Syst.*, 2020, pp. 509–514.

[8] Y. Zhang, D. Z. Cheng, T. Yao, X. Yi, L. Hong, and E. H. Chi, "A model of two tales: Dual transfer learning framework for improved long-tail item recommendation," CoRR, 2020.

[9] D.-N. Lascu and G. Zinkhan, "Consumer conformity: Review and applications for marketing theory and practice," *J. Marketing Theory Pract.*, vol. 7, no. 3, pp. 1–12, 1999.

[10] J. Park and R. Feinberg, "Eformity: Consumer conformity behaviour in virtual communities," *J. Res. Interactive Marketing*, vol. 4, no. 3, pp. 197–213, 2010.

[11] K. Kapoor, V. Kumar, L. G. Terveen, J. A. Konstan, and P. R. Schrater, ""i like to explore sometimes": Adapting to dynamic user novelty preferences," in *Proc. ACM Conf. Recommender Syst.*, 2015, pp. 19–26.

[12] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 1024–1034.

[13] M. Zhang and Y. Chen, "Inductive matrix completion based on graph neural networks," in *Proc. Int. Conf. Learn. Representations*, 2020.

[14] Y. Zhu *et al.*, "Addressing the item cold-start problem by attribute-driven active learning," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 4, pp. 631–644, Apr. 2020.

[15] M. Aharon, O. Anava, N. Avigdor-Elgrabli, D. Drachsler-Cohen, S. Golan, and O. Somekh, "Excuseme: Asking users to help in item cold-start recommendations," in *Proc. ACM Conf. Recommender Syst.*, 2015, pp. 83–90.

[16] J. Zhang, X. Shi, S. Zhao, and I. King, "STAR-GCN: Stacked and reconstructed graph convolutional networks for recommender systems," in *Proc. Int. Joint Conf. Artif. Intell.*, 2019, pp. 4264–4270.

[17] L. Hu, S. Jian, L. Cao, Z. Gu, Q. Chen, and A. Amirbekyan, "HERS: modeling influential contexts with heterogeneous relations for sparse and cold-start recommendation," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 3830–3837.

[18] T. Qian, Y. Liang, Q. Li, and H. Xiong, "Attribute graph neural networks for strict cold start recommendation," *IEEE Trans. Knowl. Data Eng.*, early access, Nov. 16, 2020, doi: 10.1109/TKDE.2020.3038234.

[19] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30–37, 2009.

[20] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T. Chua, "Neural collaborative filtering," in *Proc. 26th Int. Conf. World Wide Web*, 2017, pp. 173–182.

[21] C. Hsieh, L. Yang, Y. Cui, T. Lin, S. J. Belongie, and D. Estrin, "Collaborative metric learning," in *Proc. Int. Conf. World Wide Web*, 2017, pp. 193–201.

[22] G. Guo, J. Zhang, and N. Yorke-Smith, "Trustsvd: Collaborative filtering with both the explicit and implicit influence of user trust and of item ratings," in *Proc. AAAI Conf. Artif. Intell.*, 2015, pp. 123–125.

[23] X. He and T.-S. Chua, "Neural factorization machines for sparse predictive analytics," in *Proc. 40th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2017, pp. 355–364.

[24] X. Xin, X. He, Y. Zhang, Y. Zhang, and J. Jose, "Relational collaborative filtering: Modeling multiple item relations for recommendation," in *Proc. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2019, pp. 125–134.

[25] M. Vartak, A. Thiagarajan, C. Miranda, J. Bratman, and H. Larochelle, "A meta-learning perspective on cold-start recommendations for items," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 6904–6914.

[26] F. Pan, S. Li, X. Ao, P. Tang, and Q. He, "Warm up cold-start advertisements: Improving CTR predictions via learning to learn ID embeddings," in *Proc. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2019, pp. 695–704.

[27] M. Volkovs, G. Yu, and T. Poutanen, "Dropoutnet: Addressing cold start in recommender systems," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 4957–4966.

[28] I. Higgins *et al.*, "beta-VAE: Learning basic visual concepts with a constrained variational framework," in *Proc. Int. Conf. Learn. Representations*, 2017.

[29] R. Hamaguchi, K. Sakurada, and R. Nakamura, "Rare event detection using disentangled representation learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 9319–9327.

[30] J. Ma, C. Zhou, P. Cui, H. Yang, and W. Zhu, "Learning disentangled representations for recommendation," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2019, pp. 5712–5723.

[31] Y. Zhang, Z. Zhu, Y. He, and J. Caverlee, "Content-collaborative disentanglement representation learning for enhanced recommendation" in *Proc. ACM Conf. Recommender Syst.*, 2020, pp. 43–52.

[32] L. Hu *et al.*, "Graph neural news recommendation with unsupervised preference disentanglement," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 4255–4264.

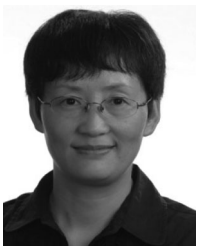
[33] X. Wang, H. Jin, A. Zhang, X. He, T. Xu, and T. Chua, "Disentangled graph collaborative filtering," in *Proc. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2020, pp. 1001–1010.

[34] J. Ma, C. Zhou, H. Yang, P. Cui, X. Wang, and W. Zhu, "Disentangled self-supervision in sequential recommenders," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2020, pp. 483–491.

[35] Y. Zheng, C. Gao, X. Li, X. He, Y. Li, and D. Jin, "Disentangling user interest and conformity for recommendation with causal embedding," in *Proc. Int. Conf. World Wide Web*, 2021, pp. 2980–2991.

[36] M. Venkatesan, "Experimental study of consumer behavior conformity and independence," *J. Marketing Res.*, vol. 3, pp. 384–387, 1966.

[37] X. Liu *et al.*, "Self-Supervised Learning: Generative or Contrastive," CoRR, 2020.
[38] P. Vincent, H. Larochelle, Y. Bengio, and P. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proc. Int. Conf. Mach. Learn.*, 2008, pp. 1096–1103.
[39] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *Proc. Int. Conf. Learn. Representations*, 2014.
[40] T. N. Kipf and M. Welling, "Variational graph auto-encoders," in *Proc. NeurIPS Workshop Bayesian Deep Learn.*, 2016.
[41] A. Razavi, A. van den Oord, and O. Vinyals, "Generating diverse high-fidelity images with VQ-VAE-2," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2019, pp. 14 837–14 847.
[42] I. Misra and L. van der Maaten, "Self-supervised learning of pretext-invariant representations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 6706–6716.
[43] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *Proc. Int. Conf. Learn. Representations*, 2016.
[44] J. Wu *et al.*, "Self-Supervised Graph Learning for Recommendation," CoRR, 2020.
[45] J. Yu, H. Yin, J. Li, Q. Wang, N. Q. V. Hung, and X. Zhang, "Self-supervised multi-channel hypergraph convolutional network for social recommendation," in *Proc. Int. Conf. World Wide Web*, 2021, pp. 413–424.
[46] K. Zhou *et al.*, "S3-Rec: Self-supervised learning for sequential recommendation with mutual information maximization," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage.*, 2020, pp. 1893–1902.
[47] J. Cao, X. Lin, S. Guo, L. Liu, T. Liu, and B. Wang, "Bipartite graph embedding via mutual information maximization," in *Proc. 14th ACM Int. Conf. Web Search Data Mining*, 2021, pp. 635–643.
[48] X. Xia, H. Yin, J. Yu, Q. Wang, L. Cui, and X. Zhang, "Self-supervised hypergraph convolutional networks for session-based recommendation," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 4503–4511.
[49] Y. Hu, Y. Koren, and C. Volinsky, "Collaborative filtering for implicit feedback datasets," in *Proc. IEEE Int. Conf. Data Mining*, 2008, pp. 263–272.
[50] Y. Wu, C. DuBois, A. X. Zheng, and M. Ester, "Collaborative denoising auto-encoders for top-n recommender systems," in *Proc. ACM Int. Conf. Web Search Data Mining*, 2016, pp. 153–162.
[51] Z. Cheng, Y. Ding, X. He, L. Zhu, X. Song, and M. S. Kankanhalli, "A3NCF: An adaptive aspect attention model for rating prediction," in *Proc. Int. Joint Conf. Artif. Intell.*, 2018, pp. 3748–3754.
[52] C. Chen, M. Zhang, Y. Liu, and S. Ma, "Neural attentional rating regression with review-level explanations," in *Proc. Int. Conf. World Wide Web*, 2018, pp. 1583–1592.
[53] C. Chen, M. Zhang, Y. Zhang, Y. Liu, and S. Ma, "Efficient neural matrix factorization without sampling for recommendation," *ACM Trans. Inf. Syst.*, vol. 38, no. 2, pp. 14:1–14:28, 2020.
[54] G. Box and D. Meyer, "An analysis for unreplicated fractional factorials," *Technometrics*, vol. 28, pp. 11–18, Feb. 1986.
[55] Y. Ji, A. Sun, J. Zhang, and C. Li, "A re-visit of the popularity baseline in recommender systems," in *Proc. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2020, pp. 1749–1752.
[56] C. Chen, M. Zhang, Y. Zhang, W. Ma, Y. Liu, and S. Ma, "Efficient heterogeneous collaborative filtering without negative sampling for recommendation," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 19–26.
[57] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, 2015.
[58] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, 2008.



Tieyun Qian (Member, IEEE) received the BS degree in computer science from University of Technology, China, in 1991, and the PhD degree in computer science from the Huazhong University of Science and Technology, China, in 2006. She is currently a professor with the School of Computer Science, Wuhan University, China. Her current research interests include text mining, Web mining, and natural language processing. She has published more than 80 papers in leading conferences and top journals including ACL, AAAI, *Special Interest Group on Information Retrieval*, *ACM Transactions on Information Systems*, and *IEEE Transactions on Knowledge and Data Engineering*. She is a member of ACM, and CCF. She has served as program committee member or area chair of many premium conferences like WWW, AAAI, ACL, IJCAI.

Yile Liang received the BS degree in computer science from Hunan University, China, in 2018, and the master's degree in computer science from Wuhan University, China, in 2021. His current research interests include recommender systems and web mining. He has authored/coauthored several papers in leading conferences and journals such as *IEEE Transactions on Knowledge and Data Engineering*, *Special Interest Group on Information Retrieval*, AAAI, and *Neurocomputing*.

Qing Li (Fellow, IEEE) received the BEng degree from Hunan University (Changsha), and the MSc and PhD degrees from the University of Southern California, Los Angeles, all in computer science. He is currently a chair professor with the Department of Computing, the Hong Kong Polytechnic University. His research interests include multi-modal data management, machine learning, social media, Web services, and e-learning systems. He has authored/co-authored more than 400 publications in these areas. He has actively involved in the research community and served as an associate editor of a number of major technical journals including *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, *ACM Transactions on Internet Technology (TOIT)*, *Data Science and Engineering (DSE)*, *World Wide Web (WWW)*, and *Journal of Web Engineering*, in addition to being a conference and program chair/co-chair of numerous major international conferences. He also sat in the steering committees of DASFAA, ER, ACM RecSys, IEEE U-MEDIA, and ICWL. He is a fellow of IEE/IET, and a distinguished member of CCF (China).

Xuan Ma received the BS degree in computer science from Wuhan University, China, in 2021. She is currently working toward the master's degree with the School of Computer Science, Wuhan University, China. Her current research interests are in the recommender systems and web mining.

Ke Sun received the BS degree in computer science from Wuhan University, China, in 2017. He is currently working toward the PhD degree with the School of Computer Science, Wuhan University, China. His current research interests are in the recommender systems and web mining. He has authored/coauthored several papers in leading conferences and journals such as AAAI, CIKM, and *Information Sciences*.

Zhiyong Peng (Member, IEEE) received the BSc degree from Wuhan University, the MEng degree from the Changsha Institute of Technology of China, in 1985 and 1988, respectively, and the PhD degree from the Kyoto University of Japan, in 1995. He is currently a professor with the school of computer science and vice director with Big Data institute, Wuhan University, member of the seventh discipline assessment groups of the academic degree commission of the state council of China. He worked as a researcher with Advanced Software Technology and Mechatronics Research Institute of Kyoto from 1995 to 1997 and a member of technical staff in Hewlett-Packard Laboratories Japan from 1997 to 2000. His research interests include object deputy databases, Big Data management and analysis. He is a member of IEEE Computer Society, ACM SIGMOD and vice director of Database Society of Chinese Computer Federation. He was general co-chair of WAIM2011, DASFAA2013 and PC co-chair of DASFAA2012, WISE2006, and CIT2004.