Self-supervised Contrastive Enhancement with Symmetric Few-shot Learning Towers for Cold-start News Recommendation

Hao Jiang
Communication University of China, Beijing, China
jianghaocuc@foxmail.com

Chuanzhen Li
Communication University of China, Beijing, China
lichuanzhen@cuc.edu.cn

Juanjuan Cai
Communication University of China, Beijing, China
caijuanjuan@cuc.edu.cn

Runyu Tian
Communication University of China, Beijing, China
tianrunyu@cuc.edu.cn

Jingling Wang
Communication University of China, Beijing, China
wjl@cuc.edu.cn

ABSTRACT

Nowadays, news spreads faster than it is consumed. This, alongside the rapid news cycle and delayed updates, has led to a challenging *news cold-start* issue. Likewise, the *user cold-start* problem, due to limited user engagement, has long hindered recommendations. To tackle both of them, we introduce the **S**ymmetric **F**ew-shot Learning framework for **C**old-start **N**ews **R**ecommendation (SFCNR), built upon self-supervised contrastive enhancement. Our approach employs symmetric few-shot learning towers (SFTs) to transform warm user/news attributes into their behavior/content features during training. We design two innovative feature alignment strategies to enhance towers training. Subsequently, this tower generates virtual features for cold users/news during inference, leveraging tower-stored prior knowledge through a personalized gating network. We assess the SFCNR on four quality news recommendation models, conducting comprehensive experiments on two kinds of News dataset. Results showcase significant performance boosts for both warm and cold-start scenarios compared to baseline models.

CCS CONCEPTS

• **Information systems → Recommender systems**.

KEYWORDS

News Recommendation; Cold-start Recommendation; Few-shot Learning; Contrastive Learning

ACM Reference Format:

Hao Jiang, Chuanzhen Li, Juanjuan Cai, Runyu Tian, and Jingling Wang. 2023. Self-supervised Contrastive Enhancement with Symmetric Few-shot Learning Towers for Cold-start News Recommendation. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM '23), October 21–25, 2023, Birmingham, United Kingdom.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3583780.3615053

CIKM '23, October 21–25, 2023, Birmingham, United Kingdom.

ACM ISBN 979-8-4007-0124-5/23/10…$15.00
https://doi.org/10.1145/3583780.3615053

1 INTRODUCTION

Online news platforms now dominate as users' go-to source for daily news. Given the vast daily news volume, an efficient recommendation system catering to users' preferences is imperative.

Existing news recommendation models typically consist of two core modules: **User Interest Modeling** [2, 10, 15, 16] and **News Understanding Modeling** [38, 39, 41], which emphasize modeling accuracy. Nonetheless, in contrast to other scenarios, news recommendation confronts distinctive challenges such as short update intervals and limited timeliness [7, 26]. Consequently, a notable fraction of candidate news belong to the realm of *cold* news. Moreover, the constant influx of *cold* users to online news platforms results in sparse behavioral data [43]. The absence of historical interactions for *cold* users/news presents hurdles in optimizing their representations during model training, leading to less-than-optimal performance. This underscores that news recommendation grapples with **BOTH** user and news cold-start challenges.

Efforts to tackle cold-start issues in recommendation systems fall into three categories. The first group involves robustly inferring warm embeds for cold user/items through dropout or masking training [34]. The second group aims to enhance learning efficiency with limited interaction data by using meta-learning for cold items [18]. The third group focuses on learning the transformation between side information and behavioral features to aid item embedding initialization [25]. However, these methods are constrained by reliance on model structure and the need for additional information, rendering them inadequate for dynamic news recommendation.

Fortunately, **Zero-Shot Learning** (ZSL), well-established in computer vision [6, 33], emerges as a promising solution to the cold-start challenge. The core of ZSL is to infer knowledge from *Known* classes to *Unknown* ones by leveraging auxiliary information (e.g., attributes or contexts), thereby generating samples for the latter. Similarly, in cold-start scenarios, limited historical interaction data impedes model training for new users and items. Hence, it makes sense to extend the idea of ZSL to the cold-start recommendation, as shown at the top of Figure 1.

Several studies have tackled the cold-start problem using ZSL. LLAE [20] employed linear regression models to capture the transformation from user attributes to behaviors. MAIL [9] introduced user attribute and behavior encoders trained through cross reconstruction. GAZRec [1] utilized the generative adversarial networks to implement the zero-shot learning towers.

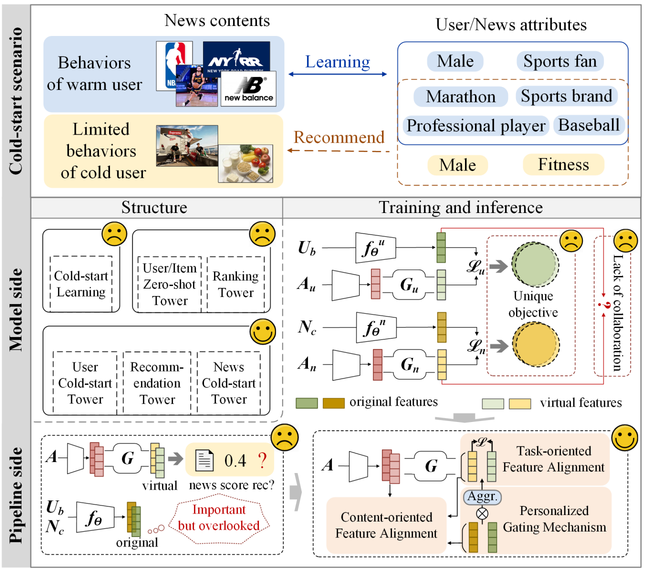


Figure 1: The motivation for the paper.

- **Shortcomings of the traditional ZSL model.** Linear models such as LLAE may struggle to handle complex mappings effectively, and replacing ranking models with zero-shot cold-start modules often focuses on addressing cold-start problems while sacrificing the performance of warm users and items. Moreover, approaches like MAIL and GAZRec have neglected direct collaboration between users and items, and the separation of zero-shot towers can lead to disparities in information distribution. Furthermore, simply aligning virtual and original features can result in overfitting problems. These shortcomings are illustrated in the middle of Figure 1.
- **Shortcomings of the traditional ZSL pipeline.** Previous zero-shot methods often differentiate warm and cold users/items only during inference. Cold users/items' original features are replaced with generated ones, while warm users/items retain their original features. However, these studies mainly focus on scenarios with *pure* cold users and items, rarely found in real-world situations. We propose that despite their scarcity, cold users' raw behaviors are crucial for characterizing interests. Similarly, for cold news, while contextual details might be tough to capture, discrete content features like news titles and entities remain vital descriptors. These constraints are depicted in the lower part of Figure 1.

Therefore, a unified framework that addresses both user and news cold-start scenarios is essential for effective news recommendation. To overcome the limitations of ZSL in tackling the cold-start problem, we propose a **S**ymmetric **F**ew-shot Learning Towers for **C**old-start **N**ews **R**ecommendation (SFCNR) in this paper. Our model comprises three towers: the *base recommendation tower*, the *user cold-start tower*, and the *news cold-start tower*. The base recommendation tower can be any high-performing news recommendation model. For the symmetric few-shot learning towers (SFTs), we divide their role into two phases: training (prior knowledge storage) and inference (prior knowledge transfer). During training, we convert user and news attributes into virtual representations through an attribute encoder, a virtual feature encoder, and a feature alignment network enhanced with contrastive learning. Specifically, to enhance cold-start learning tower training, we align virtual representations with originals internally and externally. We introduce two novel alignment methods: *content-oriented alignment* (internal) and *task-oriented alignment* (external). For inference, cold-start towers generate virtual representations by stored knowledge, followed by a personalized gating mechanism that balances virtual and original representations. Importantly, our framework is model-agnostic, requiring no changes to the original model structure.

To summarize, our major contributions are as follows:

- We propose a model-agnostic framework called SFCNR, which effectively enhances recommendation performance in both user and news cold-start scenarios without modifying the structure of the original recommendation model.
- **On the model side**, we leverage the concepts of both self-supervised and supervised learning to propose two innovative alignments, ensuring that the cold-start towers effectively learn precise and tailored virtual feature representations for the recommendation.
- **On the pipeline side**, we highlight the importance of raw features of cold users and cold news and propose a personalized gating mechanism to perform weighted fusion based on the quality of raw representation learning.
- We conduct experiments on two real-world news recommendation datasets and observe that our model successfully enhances recommendation performance in cold-start scenarios while maintaining performance in warm-start scenarios.

2 RELATED WORK

This section provides a review of existing models related to news recommendation, with a focus on cold-start recommendation and zero-shot recommendation.

2.1 News Recommendation

Existing news recommendation methods can be categorized into three main groups. Feature-based methods [3, 5, 11, 12, 30] rely on interaction information and hand-crafted news features for recommendations. Deep learning-based methods [2, 4, 21, 22, 28, 35, 38–41, 44] utilize complex neural network encoders to represent news and users, enabling more sophisticated recommendation scenarios. Graph-based methods [10, 15, 16, 29, 31] consider higher-order relationships between users and news to alleviate the sparsity of interaction data. While they focus on accurately modeling users and understanding news, they often overlook the cold-start problem, which arises when there are few training samples available.

2.2 Cold-start Recommendation

The cold-start issue persists in recommendation systems due to inadequate samples, resulting in inaccurate user and item representations and subpar suggestions. Current solutions for tackling cold-start challenges fall into four categories: 1) Strengthening CF-based models via methods like data augmentation, DropoutNet [34], MTPR [8], and CC-CC [32], which randomly omit partial embeddings. 2) Enhancing learning efficiency and adaptability in cold-start scenarios using effective interaction data and common prior knowledge, seen in MeLU [18], MWUF [24], and MetaHIN [45]. 3) Expanding user and item representations with auxiliary data [25, 27] and employing graph neural network (GNN) techniques [14, 23] to capture higher-order connectivity information, addressing sparse data concerns. 4) Advancing comprehension of cold users via cross-domain learning [19, 36], transferring knowledge from different domains. However, these methods often focus on either user or item aspects and necessitate model structure or parameter adjustments to tackle the cold-start problem.

2.3 Zero-shot Learning Recommendation

Zero-Shot Learning (ZSL) techniques, commonly used in computer vision, tackle the classification problem for samples of unknown classes, which shares similarities with the cold-start problem in recommendation systems. In the cold-start scenario, both cold users and cold news lack sufficient interaction data as training samples, impeding the capture of user interests and news understanding. Thus, extending the concept of ZSL to cold-start recommendation is a logical step. Several attempts have been made to apply ZSL to address the cold-start problem, including LLAE [20], MAIL [9], and GAZRec [1]. LLAE employs a low-rank linear encoder to map user attribute features to user behavior features. MAIL utilizes two auto-encoders to align different user feature spaces through cross-reconstruction. GAZRec considers cold-start scenarios for both users and news, using generative adversarial networks to train zero-shot learning towers, which are then applied to cold users and news. These methods aim to learn mapping relationships from prior attribute features of warm users/items and apply them to new users/items. However, some of these approaches overly prioritize cold-start scenarios, potentially impacting the recommendation performance for warm users/items, which is unreasonable. Moreover, disregarding the original features of the cold users/items leads to inadequate understanding of the users/items.

3 METHODS

In this section, we first introduce the basic recommendation tower. Then we explore the cold-start learning towers (SFTs). Finally we describe how to train SFTs using the few-shot learning.

3.1 Preliminaries

3.1.1 Definition of Warm/Cold Users and News. For any user, we denote it as $u = \{A_u, B, flag_u\}$ containing both attribute features (A_u) and behavior features (B), and similarly, we define news containing attribute features (A_n) and content features (C) as $n = \{A_n, C, U_n, flag_n\}$, where U_n means the set of users who have clicked on the news, $flag_u$ and $flag_n$ are used in the gating mechanism to indicate the actual type of user and news. The set of users and news is divided into three different types: warm (U_w/N_w), normal (U/N) and cold (U_c/N_c), as given in Equation 1.

$$u \in \begin{cases} U_w & |\mathrm{B}| \geq 10, \\ U & 3 < |\mathrm{B}| < 10, \\ U_c & |\mathrm{B}| \leq 3 \end{cases} \qquad n \in \begin{cases} N_w & |\mathrm{U_n}| \geq 10, \\ N & 0 < |\mathrm{U_n}| < 10, \\ N_c & |\mathrm{U_n}| = 0 \end{cases} \tag{1}$$

3.1.2 Problem Definition. In our framework, the recommendation tower is model-agnostic, aiming to calculate the score of how much a user prefers a candidate news when given the representation of any type of user ($\mathbf{u}$) and news ($\mathbf{n}$). Specially, for the cold user $u \in U_c$ and news $n \in N_c$, we add cold-start learning towers to acquire virtual behavior representation ($\hat{\mathbf{u}_b}$) and virtual content representation ($\hat{\mathbf{n}_c}$) using attribute features representation ($\mathbf{u}_t$) and ($\mathbf{n}_t$) during the training phase, and in the inference phase, a personalized fusion mechanism is used to augment the original behavior representation ($\mathbf{u}_b$) and content representation ($\mathbf{n}_c$) for final recommendation.

3.2 Overall Framework

The structure of our model is depicted in Figure 2, consisting of three main modules: the *Base recommendation tower*, the *User cold-start learning tower*, and the *News cold-start learning tower*, referred to as symmetric few-shot learning towers (SFTs).

The base recommendation tower, which can be any outstanding news recommendation model, generates final recommendations by considering user behaviors and news contents. To address the cold-start problems, we introduce the additional user cold-start learning tower and the news cold-start learning tower, which are symmetrically designed. SFTs learn virtual representations by leveraging attribute features from warm users and news. The learned virtual representations are then utilized to enhance recommendations for cold users and news. This approach effectively tackles the cold-start challenges in news recommendation.

In terms of the pipeline, our model incorporates a personalized gating mechanism to fuse the original features and virtual features. This ensures the preservation of original information while enhancing the virtual features, effectively mitigating the cold-start problems. On the model side, we employ contrastive learning to improve the learning of virtual features. By sampling negative samples, we capture important information related to behavior/content features and attribute features, enabling efficient training.

3.3 Basic Recommendation Tower

3.3.1 Basic Recommendation. The news and user encoders of the top-performing baselines are applied to generate representations of news contents and user behaviors. For different types of user and news, we differentiate to generate the user representation ($\mathbf{u}$) and news representation ($\mathbf{n}$), as shown in Equation 2.

$$\mathbf{u} = \begin{cases} \mathbf{u}_b & \mathrm{flag_u} = 1, \\ Gate(\mathbf{u}_b, \hat{\mathbf{u}_b}) & \mathrm{flag_u} = 0 \end{cases} \qquad \mathbf{n} = \begin{cases} \mathbf{n}_c & \mathrm{flag_n} = 1, \\ Gate(\mathbf{n}_c, \hat{\mathbf{n}_c}) & \mathrm{flag_n} = 0 \end{cases} \tag{2}$$

where $Gate(\cdot)$ indicates the personalized fusion unit to optionally enhance the original information with virtual information. It would be activated automatically when $flag_u = 0$ or $flag_n = 0$, i.e. when users or news are cold. $\mathbf{u}_b$ and $\mathbf{n}_c$ could be the behavior representation and content representation of any type of users and news, while $\hat{\mathbf{u}_b}$ and $\hat{\mathbf{n}_c}$ are virtual representations of cold users and news.

3.3.2 Personalized Gating Mechanism. We believe that the original behaviors of cold users matter a great deal, as incipient behaviors are more indicative of interest without noise. Similarly, for cold news, despite the lack of contextual information, some discrete news content features, such as titles and entities, remain essential to describing the news and should not be ignored. In short, we argue that original user representations reflect real user behaviors, while virtual user representations suggest potential user behaviors, and the same goes for news. Traditional methods directly replace

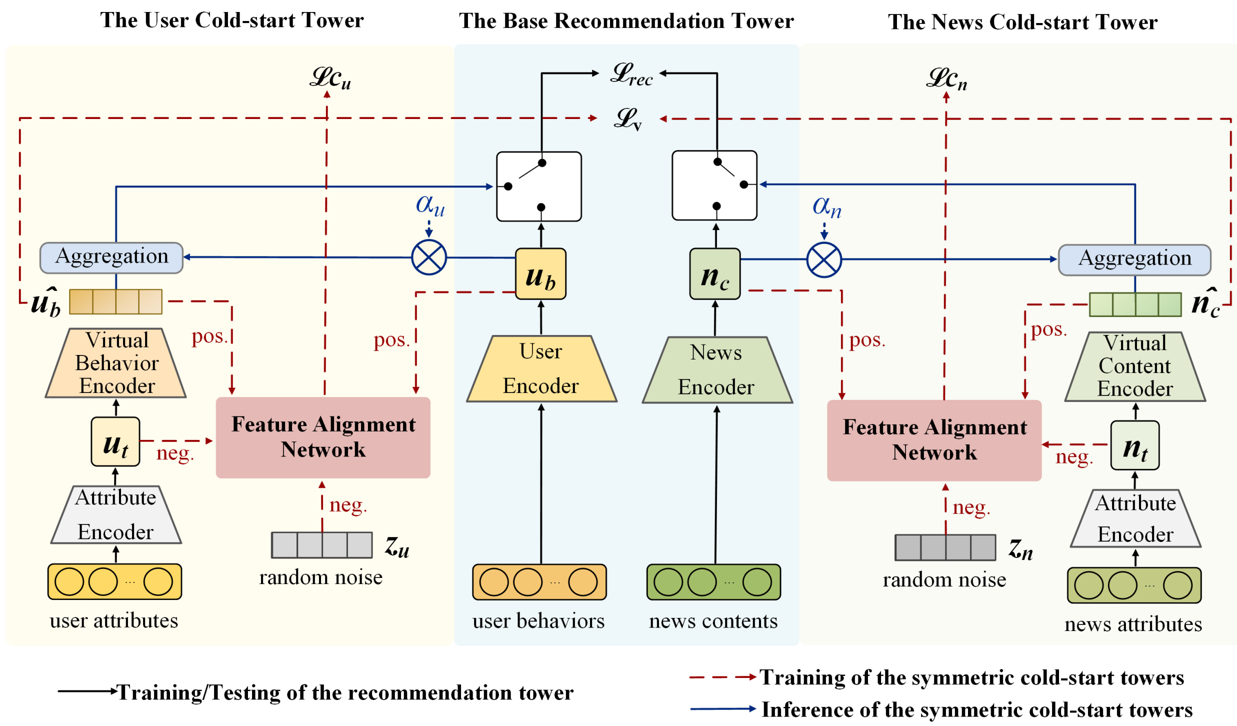


Figure 2: The overall structure of SFCNR.

the representation of cold users or items with generated representations, thus ignoring and wasting real information. We proposes a personalized fusion mechanism ($Gate(\cdot)$) to reserve the original information and enhance it with latent information. In the case of the user we first assess the quality of original representation learning with Equation 3.

$$\alpha_u = \textbf{sigmoid}(\mathbf{q}^{\mathrm{T}}(\mathbf{W}_u \times \mathbf{u}_b + \mathbf{b}_u)), \tag{3}$$

where $\mathbf{u}_b$ is the original behavior representation of the user. $\mathbf{q}^{\mathrm{T}}$, $\mathbf{W}_u$, and $\mathbf{b}_u$ are trainable parameters. Ultimately, the final representation is obtained by a weighted fusion of the original feature representation ($\mathbf{u}_b$) and the virtual feature representation ($\hat{\mathbf{u}_b}$), as shown in Equation 4. The news will be fused in the same way.

$$\mathbf{u} = \alpha_u * \mathbf{u}_b + (1 - \alpha_u) * \hat{\mathbf{u}_b}. \tag{4}$$

3.3.3 Basic Training. During the training of the recommendation model, a negative sampling technique is implemented to balance positive and negative items. Each clicked news is thought to be a positive sample, and **K** news items that were presented in the same session but not clicked by the corresponding user are randomly selected as negative samples. Besides, the order of news sequences is disrupted to avoid possible positional bias. The probability scores of positive and negative news being clicked on are expressed as $\hat{y^+}$ and $[\hat{y_1^-}, \hat{y_2^-}, ..., \hat{y_K^-}]$. By normalising these scores using the *Softmax* function, the problem of predicting the probability of a click is converted into a Perudo **K+1** classification task. The loss function is shown as Equation 5, where S is the set of training samples.

$$\mathcal{L}_{rec} = -\sum_{i \in \mathcal{S}} \log\left(\frac{\exp\left(\hat{y}_i^+\right)}{\exp\left(\hat{y}_i^+\right) + \sum_{j=1}^{K} \exp\left(\hat{y}_{i,j}^-\right)}\right). \tag{5}$$

3.4 Symmetric Few-shot Learning Towers

The symmetric few-shot learning towers consist of two main components: the *user cold-start learning tower*, and the *news cold-start learning tower*. Each of them is identical in structure, which contains an attribute encoder and a virtual feature generator, creating virtual feature representations for cold users or cold news. Furthermore, although the two towers are constructed independently, there is a potential connection between them. To generate high-quality virtual features, we design two types of feature alignment, content-oriented feature alignment and task-oriented feature alignment.

3.4.1 Construction of Attribute Features. In regard to the user (u), we take user ID (ID_u) and categories of his/her preferred news as user attributes (A_u), and in particular, as there are no generic attributes in the dataset, we choose coarse-grained attributes such as topics (T) and subtopics (S) of news to describe the categories, i.e. $A_u = \left[T_1, S_1 \ldots, T_{|B|}, S_{|B|}\right]$, where $|B|$ is the number of user behaviors. As for the news (n), due to the lack of contextual information, we regard news ID (ID_n) and discrete content features like titles (W), entities (E), topics (T) and subtopics (S) as news attributes, i.e. $A_n = [W_1, \ldots, W_M, E_1 \ldots, E_L, T, S]$, where M and L are the number of title words and news entities.

3.4.2 Virtual Feature Generation. In the user cold-start tower, we generate the representation of attribute feature ($\mathbf{u}_t$) through an attribute encoder (G_u). To be specific, the attribute encoder first embeds categories and user ID, followed by a non-linear transformation through a fully connected layer, and then performs an attention mechanism for weighted fusion, as given by Equation 6.

$$\mathbf{u}_t = \textbf{ATT}(\textbf{MLP}_a(\textbf{Emb}_u(\mathbf{A_u})) \,||\, \textbf{MLP}_{id}(\textbf{Emb}_{id}(\mathbf{ID_u}))), \tag{6}$$

where $\mathbf{Emb}_u(\cdot)$ and $\mathbf{Emb}_{id}(\cdot)$ are category embedding and user ID embedding layers, respectively. $\mathbf{ATT}(\cdot)$ denotes the attention fusion layer, and $\mathbf{MLP}(\cdot)$ represents a multi-layer perceptron. $||$ signifies the concatenation. User virtual representation ($\hat{\mathbf{u}_b}$) is then generated through a virtual behavior encoder (V_u), as Equation 7.

$$\hat{\mathbf{u}_b} = \mathrm{Tanh}(\mathbf{MLP}_{u_1}(\mathrm{LeakyReLU}(\mathbf{MLP}_{u_2}(\mathbf{u}_t)))). \tag{7}$$

The encoder applies transformations to the user behavior feature space, generating the virtual representation using two fully connected layers with combined *LeakyReLU* and *Tanh* functions.

In the news cold-start tower, we apply a similar method of attribute feature generation and virtual feature encoding to derive the news virtual feature representation ($\hat{\mathbf{n}_c}$). To differentiate between news attribute and content features, we diverge from the usual multi-head attention mechanism in language models. Instead, we treat title and entity words as news labels, employing an embedding layer and a multi-layer perceptron for processing. This is accomplished using Equation 8.

$$\mathbf{n}_t = \mathbf{ATT}(\mathbf{MLP}_a(\mathbf{Emb}_n(\mathbf{A_n}))\ ||\ \mathbf{MLP}_{id}(\mathbf{Emb}_{id}(\mathbf{ID_n}))), \tag{8}$$

where $\mathbf{Emb}_n(\cdot)$ and $\mathbf{Emb}_{id}(\cdot)$ denotes the attribute embedding layer and the news ID embedding layer. We then design a virtual content feature encoder (V_n) to generate the news virtual feature representation ($\hat{\mathbf{n}_c}$), similar to the generation of user virtual representations. This is expressed by Equation 9:

$$\hat{\mathbf{n}_c} = \mathrm{Tanh}(\mathbf{MLP}_{n_1}(\mathrm{LeakyRelu}(\mathbf{MLP}_{n_2}(\mathbf{n}_t)))). \tag{9}$$

3.4.3 Content-oriented Feature Alignment. Merely aligning virtual and original features can lead to overfitting and inadequate representation in conventional approaches. To address this, we introduce a content-oriented feature alignment method utilizing contrastive learning for the internal alignment of virtual and original features. Specifically, for generated virtual features ($\hat{\mathbf{u}_b} \in \mathbb{R}^d$), we incorporate randomly generated noise vectors ($\mathbf{z}_u \in \mathbb{R}^d$) sampled from a Gaussian distribution ($\mathcal{N}(0,1)$) as negative samples, alongside attribute feature representations ($\mathbf{u}_t \in \mathbb{R}^d$) also as negative samples. Original features ($\mathbf{u}_b \in \mathbb{R}^d$) serve as positive samples. The choice of using attribute features as negative samples aims to differentiate them, ensuring the viability of the transformation. The content-oriented alignment loss function is defined by Equation 10.

$$\mathcal{L}_{\mathrm{c}_u} = -\mathbb{E}_{u \in \mathrm{U}_w}\left[\frac{\exp(\mathrm{sim}(\hat{\mathbf{u}_b}, \mathbf{u}_b)/\tau)}{\exp(\mathrm{sim}(\hat{\mathbf{u}_b}, \mathbf{z}_u)/\tau) + \exp(\mathrm{sim}(\hat{\mathbf{u}_b}, \mathbf{u}_t)/\tau)}\right]. \tag{10}$$

Here, $\mathbf{U}_w$ signifies the set of warm users, sim denotes cosine similarity, and τ stands for the temperature coefficient (set as $\tau = 0.9$). In the training phase, we exclusively utilize warm user samples to train the user cold-start learning tower due to their substantial number of click behaviors. Similarly, we employ the content-oriented alignment loss function, as defined in Equation 11, to approximately train the news cold-start learning tower.

$$\mathcal{L}_{\mathrm{c}_n} = -\mathbb{E}_{n \in \mathrm{N}_w}\left[\frac{\exp(\mathrm{sim}(\hat{\mathbf{n}_c}, \mathbf{n}_c)/\tau)}{\exp(\mathrm{sim}(\hat{\mathbf{n}_c}, \mathbf{z}_n)/\tau) + \exp(\mathrm{sim}(\hat{\mathbf{n}_c}, \mathbf{n}_t)/\tau)}\right], \tag{11}$$

where $\mathbf{N}_w$ denotes the set of warm news, $\hat{\mathbf{n}_c}$ and $\mathbf{n}_c$ represents virtual features and original features of news, respectively, and $\mathbf{n}_t$ refers to news attribute features.

3.4.4 Task-oriented Feature Alignment. In traditional methods, the division of zero-shot towers can cause discrepancies in the distribution of information between virtual features and original features, leading to sub-optimal recommendation performance. To address this, we propose a supervised task for generating virtual features that better align with behavior and content distributions while performing well in recommendation tasks.

Specifically, our task-oriented alignment method applies the same negative sampling strategy as for the base recommendation to guarantee the efficiency of model training, and a cross-entropy loss function to achieve the task-oriented feature alignment, as shown in Equation 12.

$$\mathcal{L}_v = \sum_{u \in \mathrm{U_w}\ and\ n \in \mathrm{N_w}} -\log\left(\frac{\exp(\mathrm{y}(\hat{\mathbf{u}_b}, \hat{\mathbf{n}_c^i}))}{\exp(\mathrm{y}(\hat{\mathbf{u}_b}, \hat{\mathbf{n}_c^i})) + \sum_{j=1}^{K}\exp(\mathrm{y}(\hat{\mathbf{u}_b}, \hat{\mathbf{n}_c^j}))}\right), \tag{12}$$

where $\mathbf{U}_w$ and $\mathbf{N}_w$ denotes the set of warm users and warm news, $\hat{\mathbf{u}_b}$ and $\hat{\mathbf{n}_c}$ indicates user and news virtual representation, $\mathrm{y}(\cdot)$ represents the virtual feature predictor implemented by dot product, and $\mathbf{K}$ refers to the number of negative samples.

3.5 Training Tricks

Our model adopts an end-to-end training method to ensure the collaborative effect among the modules. According to Equation 13, the overall loss function is the sum of each loss function.

$$\mathcal{L} = \mathcal{L}_v + \mathcal{L}_{\mathrm{c}_n} + \mathcal{L}_{\mathrm{c}_u} + \mathcal{L}_{rec}. \tag{13}$$

Specifically, the cold-start learning tower is divided into two phases: training and inference. In the training phase, only warm users and news are used to train the towers. In the inference phase, a personalized gating mechanism is employed to combine original features with virtual features. These fused representations are then applied to cold users and news to improve recommendation performance in cold-start scenarios. Additionally, the parameters of the recommendation tower are fine-tuned by training on the entire dataset to enhance the model's adaptability to diverse user and news profiles.

4 EXPERIMENT

The experiments are designed to answer the following questions:

- **RQ 1**: Can our model effectively improve the performance of existing models in user/news cold-start scenarios?
- **RQ 2**: Does our method have advantages over other zero-shot models?
- **RQ 3**: Can the features learned by the cold-start towers enhance news recommendation performance?
- **RQ 4**: In the real-world dataset, what are the differences between recommendations made for cold users/news in the original model and in the model with the cold-start learning towers added?

4.1 Experimental Settings

4.1.1 Datasets. The dataset used in the experiments is Microsoft's MIND news dataset[1] [42]. It is an open-source dataset that includes logs of users from the MSN News website[2] for a period of one

[1] https://msnews.github.io/index.html
[2] https://www.msn.com/en-us/news

month, consisting of one million records containing news content and user click behavior information. We preprocess a version of the dataset to construct **MIND-small** for model training and testing. Additionally, we supplement it with an auxiliary news dataset, **Adressa**[3] [13], which is a Norwegian language dataset containing records of user interactions over a one-week period. The detailed statistics of these two pre-processed are described in Table 1.

Table 1: Detailed statistics of datasets.

Dataset	#**Users**	#**News**	#**Impressions**	#**Clicks**
MIND-small	47,391	54,997	50,000	2,369,550
Adressa	24,818	564	18,639	111,840

4.1.2 Parameter Settings. During model training, the **Adam** [17] optimizer is used, with a learning rate of 1×10^{-4}, 5×10^{-5}, and 2×10^{-5} for different experiments, and it is found that 1×10^{-4} is the best. The batch size is set to 50, and the dropout rate is set to 0.2. The negative sampling ratio K is set to 4. We train the models on 70% of the data and reserve the remaining 30% for testing. The selected evaluation metrics are AUC, MRR, NDCG@5, and NDCG@10.

4.2 Overall Performance (RQ1)

In this section, we compare four excellent existing baseline models with models that incorporate the user cold-start learning tower and the news cold-start learning tower under cold-start recommendation scenarios. The comparison methods include:

- **NPA**[39] (denoted as **Original-NPA**), a news recommendation method with personalized attention mechanism to select important words and news articles according to user preferences. Specifically, user ID is used as the query vector for the word-level and news-level attention networks.
- **NRMS**[41] (denoted as **Original-NRMS**), a news recommendation method which adopts multi-head self-attention mechanism in user and news encoders to capture interactions between words in news titles, and between news browsed by users.
- **NAML**[38] (denoted as **Original-NAML**), a news recommendation method with attentive multi-view learning to incorporate different kinds of news information such as titles, bodies and topic categories into the representations of news articles.
- **MNN4Rec** (denoted as **Original-MNN4Rec**), a news recommendation model which utilizes both news external knowledge and news internal features for news representation and enhances user interest using multi-level user behavior modeling.
- **SFTs-NPA**, **SFTs-NRMS**, **SFTs-NAML**, **SFTs-MNN4Rec**, model that adopts NPA, NRMS, NAML and MNN4Rec as the basic recommendation model respectively, and adds the SFTs[4].

4.2.1 Recommendation Scenarios. To verify the effectiveness of this framework in different scenarios, we create four recommendation scenarios: warm user recommendation scenario **(WU)**, cold user recommendation scenario **(CU)**, warm news recommendation scenario **(WN)**, and cold news recommendation scenario **(CN)**. The size of the samples is shown in Table 2.

[3]https://reclab.idi.ntnu.no/dataset/
[4]https://github.com/JiangHaoPG11/SFCNR_code

Table 2: Number of samples in four different scenarios.

Sample size	**WU**	**CU**	**WN**	**CN**
MIND-small	89,074	861,291	1,443,271	308,765
Adressa	49,422	62,376	42,367	69,431

4.2.2 Overall Performance. In various recommendation scenarios, our framework leads to significant improvements in recommendation performance. Specifically, in the **MIND-Small** dataset, the models demonstrate an improvement ranging from **1%** to **3%**, while in the **Adressa** dataset, the improvement ranges from **4%** to **14%**. These results are presented in Tables 3, 4, 5, and 6, with the growth rate denoted as **IMP**. Notably, the most substantial effect is observed in the **CU** and **CN**, where the four models achieve an average improvement (**Avg. IMP**) of **3.347%** and **3.258%** in terms of AUC.

Discussions. We have three discussions as followed:

- We observe that the baseline model NRMS demonstrates the most significant improvement when incorporating the cold-start learning towers in the warm user recommendation scenario **(WU)**. This can be attributed to the fact that NRMS relies solely on headline information from clicked news and utilizes a multi-head self-attention mechanism to process user behavior representation. As a result, it becomes highly sensitive to cold-start scenarios when click behavior is insufficient, and no auxiliary information is available to enhance the representation of user behaviors **(CU)**. In contrast, our model is capable of learning virtual user behavior features with similar interests from representations of user attributes, thereby enhancing the recommendation performance.
- We have observed that the model does not exhibit significant improvement in the cold news recommendation scenario (**CN**). This can be attributed to two main reasons: firstly, the diversity and complexity of news content representations pose challenges in learning suitable virtual content representations from discrete news attributes. Secondly, the limited number of cold news samples from the same user also impacts the model's performance. Furthermore, the model tends to prioritize warm news, where the contextual information of the text is more likely to be captured. However, our approach focuses on generating virtual representations from generic attribute features, which can greatly enhance news understanding and improve recommendation performance, even in warm news recommendation scenarios (**WN**).
- This improvement can be attributed to the training of the cold-start learning tower, which specifically emphasizes the behavior representation of warm users and the content representation of warm news through the content-oriented feature alignment module. As a result, the model becomes more adept at adapting to and aligning with the virtual representation of new users and news.

Conclusions. Compared to traditional news recommendation models such as NRMS, NAML, NPA, and MNN4Rec, our framework offers effective enhancements in recommendations for both user and news cold-start scenarios while preserving the original model structure, without compromising the performance of the recommendation tower. Moreover, unlike other cold-start methods, we also emphasize recommendations for warm users and warm news through a task-oriented feature alignment approach. This approach ensures that cold and warm representations have similar distributions, leading to a better fit for the recommendation task.

Table 3: The performance of different methods in WU scenario (recommend to warm user).

Type	Model	MIND-small				Adressa			
		AUC	MRR	NDCG@5	NDCG@10	AUC	MRR	NDCG@5	NDCG@10
NPA [39]	Original-NPA	0.6138	0.3094	0.3305	0.3857	0.8946	0.8235	0.8800	0.8961
	SFTs-NPA	0.6244	0.3150	0.3366	0.3933	0.9541	0.9123	0.9335	0.9347
NRMS [41]	Original-NRMS	0.6325	0.3222	0.3462	0.3984	0.8988	0.8881	0.8946	0.9148
	SFTs-NRMS	0.6578	0.3376	0.3620	0.4177	0.9708	0.9482	0.9611	0.9613
NAML [38]	Original-NAML	0.6635	0.3456	0.3741	0.4264	0.9145	0.8368	0.8966	0.9087
	SFTs-NAML	0.6625	0.3402	0.3675	0.4217	0.9439	0.9087	0.9308	0.9315
MNN4Rec	Original-MNN4Rec	0.6698	0.3421	0.3738	0.4255	0.9712	0.9431	0.9516	0.9576
	SFTs-MNN4Rec	0.6726	0.3436	0.3698	0.4284	0.9788	0.9583	0.9689	0.9690
Growth rate	**Avg. IMP**	**1.988%**	**2.595%**	**2.245%**	**2.578%**	**4.665%**	**6.939%**	**4.627%**	**3.273%**

Table 4: The performance of different methods in CU scenario (recommend to cold user).

Type	Model	MIND-small				Adressa			
		AUC	MRR	NDCG@5	NDCG@10	AUC	MRR	NDCG@5	NDCG@10
NPA [39]	Original-NPA	0.6069	0.2779	0.2966	0.3474	0.8713	0.5431	0.8151	0.8544
	SFTs-NPA	0.6124	0.2821	0.3050	0.3580	0.9495	0.6449	0.8919	0.9109
NRMS [41]	Original-NRMS	0.6107	0.2636	0.2926	0.3432	0.8658	0.6373	0.8528	0.8769
	SFTs-NRMS	0.6453	0.2862	0.3111	0.3667	0.9340	0.6634	0.8985	0.9245
NAML [38]	Original-NAML	0.6478	0.3005	0.3299	0.3853	0.8866	0.5495	0.8298	0.8674
	SFTs-NAML	0.6567	0.3016	0.3310	0.3803	0.9166	0.6437	0.8696	0.9032
MNN4Rec	Original-MNN4Rec	0.6489	0.3001	0.3302	0.3834	0.9329	0.6537	0.8881	0.9136
	SFTs-MNN4Rec	0.6669	0.3031	0.3287	0.3875	0.9443	0.6654	0.9037	0.9279
Growth rate	**Avg. IMP**	**3.347%**	**4.167%**	**2.927%**	**3.360%**	**5.364%**	**10.443%**	**5.333%**	**4.433%**

Table 5: The performance of different methods in WN scenario (recommend warm news).

Type	Model	MIND-small				Adressa			
		AUC	MRR	NDCG@5	NDCG@10	AUC	MRR	NDCG@5	NDCG@10
NPA [39]	Original-NPA	0.5923	0.2954	0.3157	0.3675	0.7884	0.8287	0.9249	0.9314
	SFTs-NPA	0.6017	0.3008	0.3227	0.3749	0.7923	0.8224	0.9205	0.9263
NRMS [41]	Original-NRMS	0.6101	0.3062	0.3309	0.3790	0.8084	0.8428	0.9375	0.9434
	SFTs-NRMS	0.6353	0.3227	0.3485	0.3995	0.8645	0.8703	0.9621	0.9655
NAML [38]	Original-NAML	0.6396	0.3278	0.3565	0.4053	0.7965	0.8344	0.9295	0.9361
	SFTs-NAML	0.6391	0.3247	0.3517	0.4033	0.8370	0.8499	0.9450	0.9491
MNN4Rec	Original-MNN4Rec	0.6470	0.3279	0.3591	0.4081	0.9136	0.8525	0.8620	0.9538
	SFTs-MNN4Rec	0.6474	0.3275	0.3560	0.4079	0.9279	0.8786	0.8782	0.9718
Growth rate	**Avg. IMP**	**1.894%**	**2.087%**	**1.989%**	**2.271%**	**3.895%**	**1.560%**	**1.339%**	**1.151%**

Table 6: The performance of different methods in CN scenario (recommend cold news).

Type	Model	MIND-small				Adressa			
		AUC	MRR	NDCG@5	NDCG@10	AUC	MRR	NDCG@5	NDCG@10
NPA [39]	Original-NPA	0.5219	0.1566	0.1694	0.2114	0.5167	0.0088	0.0103	0.0111
	SFTs-NPA	0.5649	0.1821	0.2079	0.2508	0.5554	0.0093	0.0106	0.0114
NRMS [41]	Original-NRMS	0.5251	0.0947	0.102	0.1105	0.7023	0.0145	0.0155	0.0157
	SFTs-NRMS	0.5534	0.1583	0.1856	0.2249	0.9137	0.0143	0.0152	0.0154
NAML [38]	Original-NAML	0.6249	0.2047	0.2372	0.2927	0.5234	0.0103	0.0119	0.0123
	SFTs-NAML	0.6282	0.2163	0.2492	0.2938	0.5655	0.0118	0.0131	0.0134
MNN4Rec	Original-MNN4Rec	0.6286	0.2147	0.2534	0.2973	0.5767	0.0127	0.0139	0.0142
	SFTs-MNN4Rec	0.6339	0.2205	0.2582	0.3040	0.8791	0.0138	0.0147	0.0150
Growth rate	**Avg. IMP**	**3.258%**	**6.100%**	**8.041%**	**5.485%**	**14.708%**	**2.724%**	**1.405%**	**1.100%**

Table 7: Ablation experiments in user cold-start scenario.

Type	Model	AUC	MRR	NDCG@10	$\mathbf{IMP_{AUC}}$
Base.	NRMS	0.6107	0.2636	0.3432	/
	NRMS-att	0.5454	0.2328	0.2873	/
	NAML	0.6478	0.3005	0.3853	/
	NAML-att	0.5913	0.2593	0.3241	/
Abla.	MAIL-NRMS	0.6111	0.2642	0.3426	0.065%
	MAIL-NAML	0.6508	0.2937	0.3771	0.463%
	GAZRec-NRMS	0.6403	0.2888	0.3707	4.847%
	GAZRec-NAML	0.6515	0.3064	0.3896	0.571%
Ours.	**SFTs-NRMS**	**0.6453**	**0.2862**	**0.3667**	**5.666%**
	SFTs-NAML	**0.6567**	**0.3016**	**0.3803**	**1.374%**

Table 8: Ablation experiments in news cold-start scenario.

Type	Model	AUC	MRR	NDCG@10	$\mathbf{IMP_{AUC}}$
Base.	NRMS	0.5351	0.1587	0.2234	/
	NRMS-att	0.5124	0.0835	0.0996	/
	NAML	0.6249	0.2047	0.3262	/
	NAML-att	0.5775	0.2545	0.2927	/
Abla.	MAIL-NRMS	0.5450	0.1690	0.2305	1.850 %
	MAIL-NAML	0.6230	0.2155	0.2890	-0.304%
	GAZRec-NRMS	0.5470	0.1705	0.2315	2.224%
	GAZRec-NAML	0.6264	0.2155	0.2934	0.240%
Ours.	**SFTs-NRMS**	**0.5534**	**0.1583**	**0.2249**	**3.420%**
	SFTs-NAML	**0.6282**	**0.2163**	**0.2938**	**0.528%**

4.3 Ablation Experiments (RQ2)

In this section, we present extensive experiments to showcase the superiority of our model compared to other zero-shot cold-start models. We design comprehensive ablation experiments where various zero-shot learning frameworks are incorporated into the baseline models. These augmented models are then compared against the original models as well as models that solely rely on user and news attribute features for recommendations. Below, we introduce the four comparison models used in the experiments.

- **Baselines**: We consider two high-performing models, **NRMS** [41], which employs multi-head self-attention to capture user behavior interactions, and **NAML** [38], which fuses user behavior using attention. These models provide recommendations by combining user and news representations from the encoders.
- **Baseline Variations**: To further explore the role of cold-start learning towers, we analyse the impact of user behavior representations and news content representations on the model, as reflected in experiments using only the generated representations of users and news for recommendations.
- **MAIL** [9]: A zero-shot framework for solving user cold-start problem, with the user attribute encoder and the user behavior encoder trained by cross-reconstruction.
- **GAZRec** [1]: A framework that combines the ideas of adversarial learning to warm-start zero-shot learning towers with old users and news.

We conduct experiments and evaluations on the **MIND-small** dataset, considering both user cold-start and news cold-start scenarios. Specifically, Table 7 shows the experimental results on the cold user dataset, and Table 8 shows the experimental results on the cold news dataset.

4.3.1 User cold-start scenario. In the user cold-start scenario, we observe limitations in the user encoding model based on the multi-head self-attention mechanism, especially when there is a scarcity of user clicks and only the headline information of news is available. In this context, NRMS proves to be sensitive and ineffective compared to the attention-weighted fusion method, i.e. NAML. However, with the integration of the cold-start learning towers, our model effectively learns potential user behavior from the representation of user interest, resulting in a significant **5.666%** improvement in the AUC metric. Even for NAML, which incorporates auxiliary information about clicked news to augment user interest during user sequence modeling, the addition of cold-start learning towers still yields a notable **1.374%** improvement in the $\mathbf{IMP_{AUC}}$ metric.

In comparison to other zero-shot learning frameworks, our proposed framework demonstrates outstanding performance in this scenario. NRMS and NAML with the MAIL framework exhibit mere improvements of **0.065%** and **0.463%**, respectively, while applying the GAZRec framework to NRMS and NAML leads to improvements of **4.847%** and **0.571%** in news recommendation metrics.

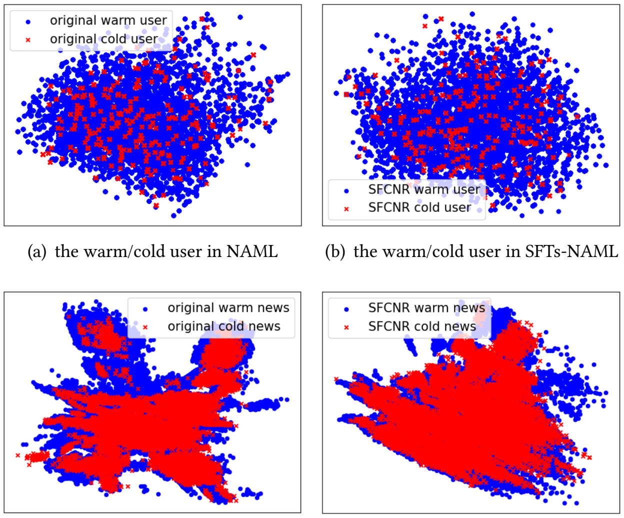
Those phenomenons can be attributed to the following reasons: firstly, our model employs contrastive learning for training the cold-start towers, eliminating the need to treat the distributions of original and virtual representations as labels during training. Consequently, the obtained virtual features better conform to the real distribution. Moreover, to ensure the virtual features are well-suited for the recommendation task, we design a task-oriented feature alignment module. In contrast, the aforementioned models solely focus on the cold-start scenario, disregarding the relationship and collaboration between users and news in different scenarios.

4.3.2 News cold-start scenario. In the cold news recommendation scenario, our approach excels compared to attribute-based methods (NAML) and performs exceptionally well for text-based methods (NRMS). The textual representation of cold news is usually sub-optimal, but the inclusion of virtual features greatly improves recommendation performance.

Despite the challenges, our model surpasses baseline models, achieving impressive improvements of **3.420%** and **0.528%** in recommendation metrics for NRMS and NAML, respectively. These enhancements outperform the improvements achieved by GAZRec and MAIL. By combining the understanding of news content from basic models with virtual representations, our approach effectively enhances the recommendation of new news.

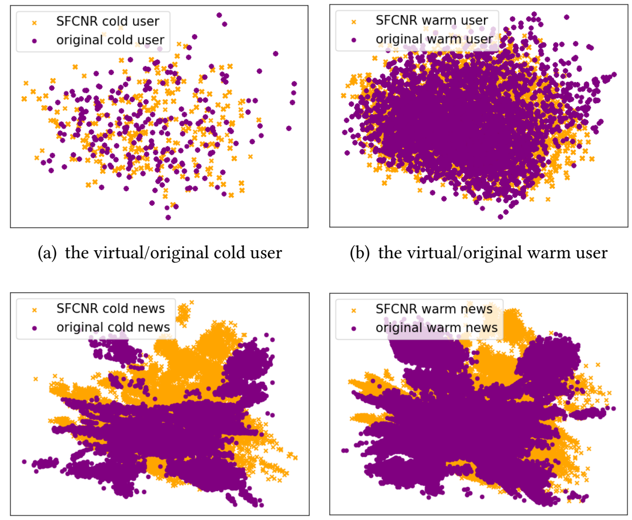
4.4 Case and Visualization Study (RQ3&4)

In this section, we employ the method of PCA [37] to reduce the dimensionality of the original user and news representations, as well as the representations learned from cold-start towers. We visualize these representations on a 2D plane to examine their distribution. Our findings demonstrate that the user and news representations learned from the cold-start towers exhibit a closer fit to the distribution of their original representations.



(a) the warm/cold user in NAML
(b) the warm/cold user in SFTs-NAML
(c) the warm/cold news in NAML
(d) the warm/cold news in SFTs-NAML

Figure 3: The visualization of the representation of warm users/news and cold users/news in origin model and SFCNR.



(a) the virtual/original cold user
(b) the virtual/original warm user
(c) the virtual/original cold news
(d) the virtual/original warm news

Figure 4: The visualization of the original and virtual representations of different types of user and news in the SFCNR.

4.4.1 Cold Representation vs. Warm Representation. To assess the effectiveness of the cold-start towers, we visualize the representations of warm/cold users and warm/cold news in both the original model and the SFCNR framework, as depicted in Figure 3.

User side: We observed a significant enhancement in the alignment between the representation of cold users (red crosses) and warm users (blue dots). The improved alignment indicates that the initial distribution of cold user representations is concentrated, hindering differentiation of their interests and impacting recommendation performance (Figure 3 (a)). However, with the integration of the user cold-start learning tower, the distribution of cold users becomes closely aligned with that of warm users (Figure 3 (b)). Moreover, the personalized gating mechanism in SFTs facilitates a broader distribution space for user representations, enabling a more comprehensive modeling of user interests.

News side: This effect is more prominent in news, as the distribution of cold news (red crosses) in the original model is completely disconnected from warm news (blue dots). The presence of discrete features confines the representation of cold news, hindering comprehensive understanding and efficient distribution (Figure 3 (c)). However, integrating the news cold-start learning tower harmonizes these distributions, improving the understanding and recommendation of cold news (Figure 3 (d)).

4.4.2 Original Representation vs. Virtual Representation. To evaluate the comprehension of user interests and news content by the cold-start towers, we compare the original and virtual representations in the SFCNR framework, as illustrated in Figure 4.

User side: For cold users, the virtual representation (orange dots) exhibit greater dispersion compared to the original representation (purple dots), indicating that the model achieves a deeper understanding of cold users' interests in SFCNR (Figure 4 (a)). Furthermore, the distribution space of virtual representation vectors is also more uniformly spread for warm users (Figure 4 (b)).

News side: Owing to the absence of contextual information for capturing textual features, the original representation vectors (depicted as purple dots) cluster distinctly for various news categories. However, through the integration of the cold-start learning tower, the model adeptly comprehends both warm and cold news, leading to a more balanced distribution of the acquired representations (Figure 4 (c)-(d)).

5 CONCLUSIONS

To address the critical cold-start challenges in news recommendation, we propose the Self-supervised Few-shot Learning framework enhanced by contrastive learning (SFCNR). This framework adeptly handles both user and news cold-start problems. Our approach utilizes Symmetric Few-shot Learning Towers (SFTs) to bridge attribute and behavior (content) features. We learn this link from warm users and news, simulating cold-start situations with limited interactions, and then apply it to cold users and news. For cold-start tower training, we devise content-oriented and task-oriented feature alignments. These methods ensure that virtual features align with the original distribution while adapting to the recommendation task. We validate our framework's efficacy through extensive experiments, including comparisons, ablations, and visualizations. The outcomes showcase improved performance in cold-start scenarios upon integrating our cold-start learning towers with leading news recommendation models. Our framework surpasses existing zero-shot cold-start approaches. Moreover, visualizations affirm successful alignment of virtual features with the original distribution, signifying enhanced representation of cold users and news.

ACKNOWLEDGMENTS

This work is supported by the National Key R&D Program of China under grants number 2020YFF0305300 and SQ2020YFF0426386.

REFERENCES

[1] Alshehri, M. A., and Zhang, X. Generative adversarial zero-shot learning for cold-start news recommendation. In *CIKM* (2022), pp. 26–36.

[2] An, M., Wu, F., Wu, C., Zhang, K., Liu, Z., and Xie, X. Neural news recommendation with long-and short-term user representations. In *ACL* (2019), pp. 336–345.

[3] Capelle, M., Frasincar, F., Moerland, M., and Hogenboom, F. Semantics-based news recommendation. In *WIMS* (2012), pp. 1–9.

[4] Castelvecchi, D. Can we open the black box of ai? *Nature News 538*, 7623 (2016), 20.

[5] Das, A. S., Datar, M., Garg, A., and Rajaram, S. Google news personalization: Scalable online collaborative filtering. In *WWW* (2007), p. 271–280.

[6] Ding, Z., Shao, M., and Fu, Y. Low-rank embedded ensemble semantic dictionary for zero-shot learning. In *CVPR* (2017), pp. 2050–2058.

[7] Donelson, D. C., McInnis, J. M., Mergenthaler, R. D., and Yu, Y. The timeliness of bad earnings news and litigation risk. *ACCOUNT REV 87*, 6 (2012), 1967–1991.

[8] Du, X., Wang, X., He, X., Li, Z., Tang, J., and Chua, T.-S. How to learn item representation for cold-start multimedia recommendation? In *ICMg* (2020), pp. 3469–3477.

[9] Feng, P. J., Pan, P., Zhou, T., Chen, H., and Luo, C. Zero shot on the cold-start problem: Model-agnostic interest learning for recommender systems. In *CIKM* (2021), pp. 474–483.

[10] Ge, S., Wu, C., Wu, F., Qi, T., and Huang, Y. Graph enhanced representation learning for news recommendation. In *WWW* (2020), pp. 2863–2869.

[11] Gershman, A., Wolfe, T., Fink, E., and Carbonell, J. G. News personalization using support vector machines.

[12] Goossen, F., IJntema, W., Frasincar, F., Hogenboom, F., and Kaymak, U. News personalization using the cf-idf semantic recommender. In *WIMS* (2011), pp. 1–12.

[13] Gulla, J. A., Zhang, L., Liu, P., Özgöbek, Ö., and Su, X. The adressa dataset for news recommendation. In *WI* (2017), pp. 1042–1048.

[14] Hao, B., Zhang, J., Yin, H., Li, C., and Chen, H. Pre-training graph neural networks for cold-start users and items representation. In *WSDM* (2021), pp. 265–273.

[15] Hu, L., Li, C., Shi, C., Yang, C., and Shao, C. Graph neural news recommendation with long-term and short-term interest modeling. *IPM 57*, 2 (2020), 102142.

[16] Hu, L., Xu, S., Li, C., Yang, C., Shi, C., Duan, N., Xie, X., and Zhou, M. Graph neural news recommendation with unsupervised preference disentanglement. In *ACL* (2020), pp. 4255–4264.

[17] Kingma, D. P., and Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[18] Lee, H., Im, J., Jang, S., Cho, H., and Chung, S. Melu: Meta-learned user preference estimator for cold-start recommendation. In *KDD* (2019), pp. 1073–1082.

[19] Li, C.-T., Hsu, C.-T., and Shan, M.-K. A cross-domain recommendation mechanism for cold-start users based on partial least squares regression. *TIST 9*, 6 (2018), 1–26.

[20] Li, J., Jing, M., Lu, K., Zhu, L., Yang, Y., and Huang, Z. From zero-shot learning to cold-start recommendation. In *AAAI* (2019), vol. 33, pp. 4189–4196.

[21] Liu, D., Lian, J., Wang, S., Qiao, Y., Chen, J.-H., Sun, G., and Xie, X. Kred: Knowledge-aware document representation for news recommendations. In *RecSys* (2020), pp. 200–209.

[22] Liu, J., Dolan, P., and Pedersen, E. R. Personalized news recommendation based on click behavior. In *IUI* (2010), pp. 31–40.

[23] Liu, S., Ounis, I., Macdonald, C., and Meng, Z. A heterogeneous graph neural model for cold-start recommendation. In *RDIR* (2020), pp. 2029–2032.

[24] Lu, Y., Fang, Y., and Shi, C. Meta-learning on heterogeneous information networks for cold-start recommendation. In *KDD* (2020), pp. 1563–1573.

[25] Mohammadi, S. A., and Andalib, A. Using the opinion leaders in social networks to improve the cold start challenge in recommender systems. In *ICWR* (2017), IEEE, pp. 62–66.

[26] Özgöbek, Ö., Gulla, J. A., and Erdur, R. C. A survey on challenges and methods in news recommendation. In *WEBIST* (2014), pp. 278–285.

[27] Prando, A. V., Contratres, F. G., de Souza, S. N. A., and de Souza, L. S. Content-based recommender system using social networks for cold-start users. In *KDIR* (2017), pp. 181–189.

[28] Qi, T., Wu, F., Wu, C., and Huang, Y. Personalized news recommendation with knowledge-aware interactive matching. In *SIGIR* (2021), pp. 61–70.

[29] Qi, T., Wu, F., Wu, C., Yang, P., Yu, Y., Xie, X., and Huang, Y. Hierec: Hierarchical user interest modeling for personalized news recommendation. In *ACL* (2021), pp. 5446–5456.

[30] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. Grouplens: An open architecture for collaborative filtering of netnews. In *CSCW* (1994), pp. 175–186.

[31] Santosh, T., Saha, A., and Ganguly, N. Mvl: Multi-view learning for news recommendation. In *SIGIR* (2020), pp. 1873–1876.

[32] Shi, S., Zhang, M., Yu, X., Zhang, Y., Hao, B., Liu, Y., and Ma, S. Adaptive feature sampling for recommendation with missing content feature values. In *CIKM* (2019), pp. 1451–1460.

[33] Verma, V. K., Brahma, D., and Rai, P. Meta-learning for generalized zero-shot learning. In *AAAI* (2020), vol. 34, pp. 6062–6069.

[34] Volkovs, M., Yu, G., and Poutanen, T. Dropoutnet: Addressing cold start in recommender systems. *NIPS 30* (2017).

[35] Wang, H., Zhang, F., Xie, X., and Guo, M. Dkn: Deep knowledge-aware network for news recommendation. In *WWW* (2018), pp. 1835–1844.

[36] Wang, X., Peng, Z., Wang, S., Yu, P. S., Fu, W., and Hong, X. Cross-domain recommendation for cold-start users via neighborhood based feature mapping. In *DASFAA* (2018), Springer, pp. 158–165.

[37] Wold, S., Esbensen, K., and Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst. 2*, 1-3 (1987), 37–52.

[38] Wu, C., Wu, F., An, M., Huang, J., Huang, Y., and Xie, X. Neural news recommendation with attentive multi-view learning. In *IJCAI* (2019), pp. 3863–3869.

[39] Wu, C., Wu, F., An, M., Huang, J., Huang, Y., and Xie, X. Npa: Neural news recommendation with personalized attention. In *KDD* (2019), pp. 2576–2584.

[40] Wu, C., Wu, F., An, M., Huang, Y., and Xie, X. Neural news recommendation with topic-aware news representation. In *ACL* (2019), pp. 1154–1159.

[41] Wu, C., Wu, F., Ge, S., Qi, T., Huang, Y., and Xie, X. Neural news recommendation with multi-head self-attention. In *EMNLP* (2019), pp. 6389–6394.

[42] Wu, F., Qiao, Y., Chen, J.-H., Wu, C., Qi, T., Lian, J., Liu, D., Xie, X., Gao, J., Wu, W., et al. Mind: A large-scale dataset for news recommendation. In *ACL* (2020), pp. 3597–3606.

[43] Zheng, G., Zhang, F., Zheng, Z., Xiang, Y., Yuan, N. J., Xie, X., and Li, Z. Drn: A deep reinforcement learning framework for news recommendation. In *WWW* (2018), pp. 167–176.

[44] Zhu, Q., Zhou, X., Song, Z., Tan, J., and Guo, L. Dan: Deep attention neural network for news recommendation. In *AAAI* (2019), vol. 33, pp. 5973–5980.

[45] Zhu, Y., Xie, R., Zhuang, F., Ge, K., Sun, Y., Zhang, X., Lin, L., and Cao, J. Learning to warm up cold item embeddings for cold-start recommendation with meta scaling and shifting networks. In *RDIR* (2021), pp. 1167–1176.