Fine-grained Representation Learning and Multi-view Collaborative Augmentation for Recommendation

Huiting Li[1], Wenjun Ma[1,2,3], Weishan Cai[4], and Yuncheng Jiang[1,2] ✉

[1] School of Artificial Intelligence, South China Normal University, Foshan, 528225, China
[2] School of Computer Science, South China Normal University, Guangzhou, 510631, China
[3] Aberdeen Institute of Data Science and Artificial Intelligence, South China Normal University, Foshan, 528225, China
[4] School of Computer Science, Guangdong University of Education, Guangzhou, 510631, China
{huitingli, caiws, jiangyuncheng}@m.scnu.edu.cn, mawenjun@scnu.edu.cn

Abstract. Graph neural networks (GNNs) have recently advanced in processing graph-structured data and are increasingly used in recommendation systems. Recently, many studies have incorporated side information as auxiliary views, such as the user's social connections and the item's knowledge-aware dependencies, to enhance the user-item interaction view. However, current works overlook the differences in learning behavior between auxiliary views and interaction view, and transfer side information from different views separately, which can lead to a semantic gap and fail to explore the collaborative effect of auxiliary views. To address this challenge, we propose FiCoRec, a novel fine-grained augmentation method for recommendation, comprising two key enhancement components: Hierarchical Knowledge Transfer (HKT) and Multi-view Semantic Fusion (MSF). Specifically, HKT designs an interaction semantic decouple (ISD) method to decouple the interaction view embeddings into homogeneous features (hoFs) and heterogeneous features (heFs). Then a hierarchical contrastive learning framework is used to fully capture the local and global semantics from the intermediate-layer to enhance hoFs. MSF explores a collaborative augmentation mechanism by utilizing meta-learning to enhance the interaction view. Extensive experiments conducted on five datasets against seven baseline methods demonstrate that our FiCoRec outperforms the state-of-the-art methods with a margin of 0.33%-2.76%.

Keywords: Heterogeneous graph · Feature decouple · Contrastive learning · Recommendation system.

1 Introduction

Recommender systems have become an essential tool in various domains, providing personalized recommendations to users by predicting their preferences based on past behavior and interactions. In recent years, with the rapid development of Graph Neural Networks (GNNs), many GNN-based methods have emerged and gained significant traction in capturing complex user-item interactions. Unlike traditional collaborative filtering or matrix factorization methods, GNNs leverage iterative message-passing mechanisms to effectively encode the structural relationships between users and items, modeling higher-order connectivity and non-linear relationships [7,26]. Despite the great process, many existing GNN-based recommendation frameworks primarily focus on homogeneous graphs without considering the diverse relational patterns in real-world scenarios which may limit their ability in complex recommendation environments.

To mitigate this limitation, researchers endeavor to exploit heterogeneous side information as auxiliary views, such as social connections and item-wise relations, to enrich the semantics of latent representations and augment the user-item interaction view. GraphRec [4] is the first to utilize GNNs to incorporate social influence to enhance recommendation performance. Subsequently, many studies introduce various techniques to model users' social relations. For instance, DiffNet++ [25] proposes a diffusion neural network to model higher-order social structures, while MHCN [28] leverages hypergraphs to fuse social relations and mine complex connections between users. Furthermore, HGCL [1] extends the heterogeneous relations and considers item dependencies cooperating with users' social influence to fully exploit side information from limited data. Additionally, Recdiff [14] and DSL [22] focus on alleviating noise in social relations by leveraging self-supervised learning (SSL).

Previous approaches primarily focus on extracting informative side information from auxiliary views to enrich the interaction view. However, the current transformation of side information faces two limitations: (1) it provides a coarse-grained transformation, overlooking the differences in learning behavior between auxiliary views and the interaction view, and (2) it transfers side information from different auxiliary views separately, leaving the collaborative effect of multiple auxiliary views insufficiently explored. The first issue arises from the distinct content of different views. In auxiliary views, GNNs operate on homogeneous graphs, capturing intra-group relationships or similarities (e.g., user social connections or item-wise relations). In contrast, the interaction view involves heterogeneous graphs, where various node types introduce greater complexity and diverse information. Consequently, the learning behaviors of GNNs are significantly different between auxiliary views and the interaction view, leading to a semantic gap in the latent representation space. Thus, blindly utilizing the embeddings of auxiliary views to augment the interaction view can introduce unnecessary noise, hindering the utilization of heterogeneous side information. The second issue lies in the isolated transfer of side information from different auxiliary views which ignores their potential collaborative effects, and the ability to jointly provide a more comprehensive understanding of user preferences and item characteristics. These limitations constrain the recommender system's capabilities, ultimately leading to suboptimal performance.

To tackle the aforementioned challenge, we propose FiCoRec, a novel fine-grained augmentation method for recommendation. FiCoRec comprises two key components: hierarchical knowledge transfer (HKT) and multi-view semantic fusion (MSF). On one hand, HKT bridges the semantic gap between auxiliary and interaction views by enabling fine-grained side information transfer. Specifically, as the embeddings of the interaction view aggregate the information about users' and items' characteristics, we initially design interaction semantic decouple (ISD) to decouple them into user-related part and item-related part. For user, the former aims to preserve the user's own intrinsic attribution, e.g., personal information and social relationships. The latter contains information related to items associated with a user, i.e., the user's personalized preferences. For items, the meanings of these parts are reversed. Thus, we denote the two parts as homogeneous features (hoFs) and heterogeneous features (heFs). Building upon this foundation, HKT further designs a hierarchical contrastive learning framework to make full use of the local and global semantics of heterogeneous relations in auxiliary views to enhance hoFs. On the other hand, MSF utilizes meta-learning to consider the collaborative effect of two auxiliary views. It treats heFs as a bridge to integrate the knowledge of two auxiliary views and generate an adaptive mapping by meta-network. Subsequently, the mapping combined with the enhanced hoFs from HKT, jointly enriches the embeddings of the interaction view. Finally, the enhanced embeddings are utilized for downstream recommendation tasks.

The key contributions of our work can be summarized as follows:

1. We propose a novel fine-grained heterogeneous relation augmentation framework (FiCoRec) to boost the recommendation performance by exploiting side information from auxiliary views.
2. We present two customized modules: HKT and MSF. HKT transfers side information in a fine-grained way by utilizing features decoupling, and enables the model to learn local and global semantics of heterogeneous relations in auxiliary views through a hierarchical contrastive learning framework. MSF simultaneously leverages multiple auxiliary views to provide collaborative augmentation.
3. FiCoRec is evaluated on several real-world datasets. Experimental results demonstrate that the FiCoRec framework greatly improves recommendation performance and achieves state-of-the-art results.

2 Related Work

2.1 Heterogeneous Graph Representation Learning

Heterogeneous networks have been widely used since real-world objects and their interactions are often multi-typed. In prior research, shallow network embedding methods leverage single-layer decomposition of certain affinity matrices, e.g., metapath2vec [3], Hin2vec [5]. With the development of GNNs, deep network embedding methods aggregate the information from neighboring nodes to learn structural representation, e.g., HAN [24], MAGNN [6]. In recent years, relevant studies have predominantly explored SSL techniques, incorporating strategies such as contrastive learning, generative pretraining, and mask reconstruction. For instance, MuHca [15] separately employs node embeddings from two views into a contrastive generative adversarial network to implement data augmentation. HGMAE [21] designs two masking strategies and metapath-based edge reconstruction, target attribute restoration, and positional feature prediction to capture graph information.

2.2 GNN-based Recommendation Systems

Graph Neural Networks (GNNs) have been widely adopted in recommendation systems due to their effectiveness in capturing high-order connectivity and user-item interactions. Early works such as NGCF [23] and LightGCN [11] learn user and item embeddings by linearly propagating them on the user-item interaction graph. Subsequent advancements like PinSage [27] introduce an industrial solution that combines random walks and graph convolutions. Later, many GNN-based recommendations are develop to model user-user and user-item graphs via message passing, such as DH-HGCN [10] and GraphRec [4]. Recent efforts focus on utilizing heterogeneous relational data to enhance GNNs, such as social connections between users and the knowledge dependencies of items [12,20]. For example, HGCL [1] transfers heterogeneous relations to the user-item interaction graph with contrastive learning across different auxiliary views. Besides, RecDiff [14] and DSL [22] focus on alleviating the noisy effect in the social graph by leveraging SSL. However, previous works neglect the differences of learning behavior between different graphs, which leads to a semantic gap in the latent representation space and introduces unnecessary noise. Moreover, existing studies transfer side information across different views separately, ignoring the potential collaborative effects. To address the challenges, we design a fine-grained augmentation method for recommendation.

3 Preliminaries

For the interaction view, we denote the user-item interaction graph as $\mathcal{G}_{ui} = \{\mathcal{U}, \mathcal{V}, \mathcal{E}_{ui}\}$, where $\mathcal{U}$ and $\mathcal{V}$ represent the sets of users and items, respectively, $\mathcal{E}_{ui} \subseteq |\mathcal{U}| \times |\mathcal{V}|$ represents all the edges in the user-item interaction graph $\mathcal{G}_{ui}$. And we define the embeddings of the interaction view as $\mathbf{E}^{inter}$. If user i interacts with an item j, that is, an edge (i, j) exists in $\mathcal{E}_{ui}$. Considering the negative sampling process, we denote $(i, j^+, j^-) \in \mathcal{O}$ as a sample for a user with existing and non-existing interactions with items, where $\mathcal{O}$ represents the set of all samples. Let $\mathbf{E}^u$ and $\mathbf{E}^i$ represent the embedding of users and items, respectively. Each user $i \in \mathcal{U}$ or item $j \in \mathcal{V}$ corresponding to a feature vector $\mathbf{e}_i^u$ or $\mathbf{e}_j^i$.

For the auxiliary views, we denote the user-user social connection graph as $\mathcal{G}_{uu} = \{\mathcal{U}, \mathcal{E}_{uu}\}$, and item-item knowledge dependency graph is denoted as

$\mathcal{G}_{ii} = \{\mathcal{V}, \mathcal{E}_{ii}\}$. $\mathcal{G}_{uu}$ represents the social connection between users, and $\mathcal{G}_{ii}$ represents the item-wise relation. Similarly, the embeddings of the two auxiliary views are denoted as $\mathbf{E}^{aux}$, including $\mathbf{E}^{uu}$ and $\mathbf{E}^{ii}$. Each user $i \in \mathcal{U}$ or item $j \in \mathcal{V}$ corresponding to a vector $\mathbf{e}_i^{uu}$ or $\mathbf{e}_j^{ii}$. The main notations and explanations are shown in Table 1.

Table 1. Notations and explanations.

Notations	Explanations
$\mathbf{E}^{inter}$	embedding of interaction view, including $\mathbf{E}^{u}$ and $\mathbf{E}^{i}$
$\mathbf{E}^{aux}$	embedding of auxiliary view, including $\mathbf{E}^{uu}$ and $\mathbf{E}^{ii}$
heF	heterogeneous feature decoupled from $\mathbf{E}^{inter}$, including UheF and IheF
UheF/IheF	heF for user/item
hoF	homogeneous feature decoupled from $\mathbf{E}^{inter}$, including UhoF and IhoF
UhoF/IhoF	hoF for user/item

4 Methodology

We develop FiCoRec which utilizes fine-grained representation learning and a multi-view collaborative augmentation strategy. As shown in Fig. 1, we first learn embeddings from $\mathcal{G}_{ui}$ (interaction view), as well as $\mathcal{G}_{uu}$ and $\mathcal{G}_{ii}$ (auxiliary views), using three-layer GNN encoders, where the encoders for $\mathcal{G}_{uu}$ and $\mathcal{G}_{ii}$ are incorporated in HKT. Then HKT provides a fine-grained representation learning, and MSF employs a multi-view collaborative augmentation strategy to enhance model performance. For HKT in Fig. 1(a), we first decouple $\mathbf{E}^{inter}$ through ISD, which is the foundation of fine-grained representation learning. Furthermore, HKT employs a hierarchical contrastive learning to transfer side information, which is able to capture local and global semantics from $\mathbf{E}^{aux}$. For MSF in Fig. 1(b), it contains two branches for user embedding and item embedding, respectively. Each branch utilizes an anchor-neighbor concatenation (ANC) and a meta-learning framework to aggregate the knowledge from two auxiliary views.

4.1 Interaction Semantic Decouple

The $\mathbf{E}^{inter}$ combines the information about user and item characteristics. Intuitively, by distinguishing these two parts and enhancing them separately with side information, the effects of the semantic gap and the discrepancies in GNN learning behaviors can be alleviated. With this motivation, we design ISD to decouple the $\mathbf{E}^{inter}$ into a user-related part and an item-related part. Then we exploit corresponding side information to enhance both parts. We define two categories of features to describe the aforementioned two parts: homogeneous features (hoFs) and heterogeneous features (heFs), whose meanings are elaborated as follows.

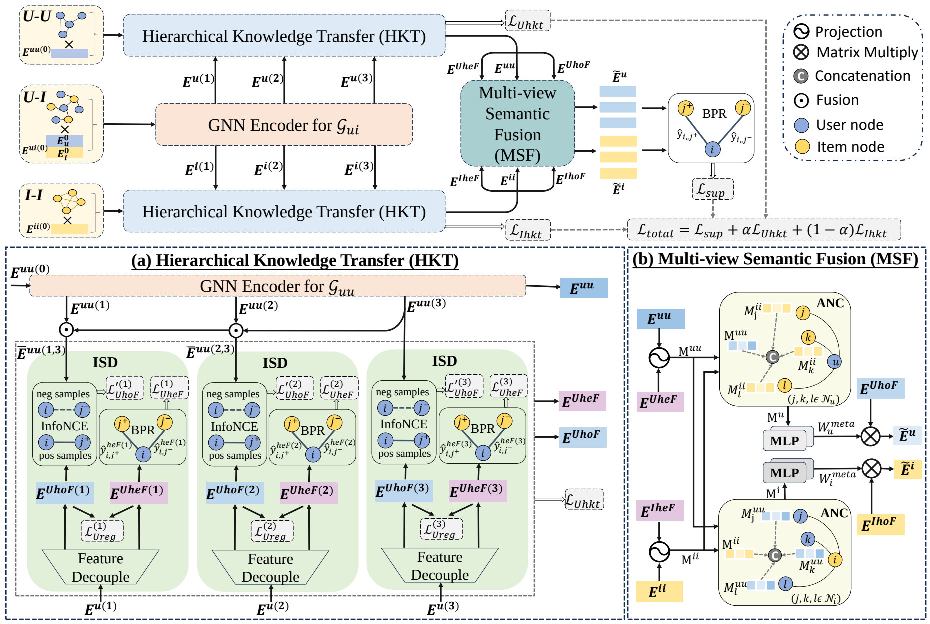


Fig. 1. The upper of the figure is the overview of our proposed FiCoRec framework, while the lower part illustrates the HKT and MSF, respectively. ISD is represented by the green bar in HKT. For ISD, we adopt the same operation for user and item. And for HKT, we design symmetric structures for the two auxiliary views. For simplicity, we illustrate them from the perspective of user.

On one hand, for hoF, we anticipate that it can preserve the intrinsic attribution relevant to either user or item, e.g., social relationships for users and knowledge dependencies for items. Therefore, we define the user-related part of user embeddings as user homogeneous features (UhoF), and for item, the item-related part of item embeddings is defined as item homogeneous features (IhoF). To ensure UhoF and IhoF capture the intrinsic attribution of user and item, we attempt to utilize the $\mathbf{E}^{aux}$ to employ supervision and augmentation. However, given that the learning of graph structures heavily relies on a comprehensive grasp of the local neighborhood nodes, directly minimizing the embeddings of two individual embeddings proves to be insufficiently effective.

Inspired by previous works [1,22], we introduce contrastive self-supervised learning to provide more effective supervision. Intuitively, we hope to maximize the similarity between node pairs corresponding to the same node in $\mathbf{E}^{inter}$ and $\mathbf{E}^{aux}$, i.e., $\mathbf{e}_i^{UhoF}$ and $\mathbf{e}_i^{uu}$ (positive pairs), while for the different nodes, i.e., $\mathbf{e}_i^{UhoF}$ and $\mathbf{e}_j^{uu}$ (negative pairs) are opposite. Here, we adapt InfoNCE loss as the optimization objective and pose the cross-view contrastive learning as the task of classifying positive pairs among hoF and $\mathbf{E}^{aux}$. Then, we denote the contrastive objective of users utilizing layer-k $\mathbf{E}^{aux}$ and $\mathbf{E}^{inter}$ as

$$\mathcal{L}_{UhoF}^{(k)} = \sum_{i\in\mathcal{U}} -\log \frac{\exp(sim(\mathbf{e}_i^{UhoF(k)}, \mathbf{e}_i^{uu(k)})/\tau)}{\sum_{j\in\mathcal{U}} \exp(sim(\mathbf{e}_i^{UhoF(k)}, \mathbf{e}_j^{uu(k)})/\tau)}, \tag{1}$$

where $\mathbf{e}^{UhoF(k)}$ represents the UhoF in layer-k and τ represents the scalar temperature parameter, $sim(\cdot)$ denotes the similarity function. Analogously, the contrastive loss of items $\mathcal{L}_{IhoF}^{(k)}$ can be formulated in a similar way.

On the other hand, for heF, we anticipate that it represents the user's personality or the item's characteristic. For example, the item-related part for user embedding represents the knowledge of item associated with the user, which contains the user's preferences. Based on this characteristic, we define the item-related part for user embedding as user heterogeneous features (UheF). Similarly, the user-related part of the item embedding represents the target audience of an item, which is defined as item heterogeneous features (IheF). Here, we introduce the downstream task as supervision. In particular, we follow previous works [1] and adopt the Bayesian Personalized Ranking (BPR) [18] pair-wise loss function to constrain heF. Similarly, we consider the case of layer-k, that is,

$$\mathcal{L}_{heF}^{(k)} = \sum_{(i,j^+,j^-)\in\mathcal{O}} -\ln(\text{sigmoid}(\hat{y}_{i,j^+}^{heF(k)} - \hat{y}_{i,j^-}^{heF(k)})) \tag{2}$$

where $\hat{y}_{i,j}^{heF(k)} = (\mathbf{e}_i^{UheF(k)})^\top \mathbf{e}_j^{IheF(k)}$ represents the likelihood of user i interacting with item j. Here, $\mathbf{e}^{UheF(k)}$ and $\mathbf{e}^{IheF(k)}$ denote the heF for user and item decoupled from layer-k $\mathbf{E}^{inter}$, respectively.

Meanwhile, hoF and heF represent different parts of features for $\mathbf{E}^{inter}$, respectively. It is vital to minimize the correlation between them to ensure their independence, mitigating the unnecessary impact introduced by the mutual interaction of hoF and heF. Thus, we employ additional constraints on them utilizing orthogonal regularization, that is,

$$\mathcal{L}_{Ureg}^{(k)} = \|(\mathbf{E}^{UhoF(k)})^\top \mathbf{E}^{UheF(k)} - I\|_F^2, \tag{3}$$

where $\mathbf{E}^{UhoF(k)}$ and $\mathbf{E}^{UheF(k)}$ represent the embedding matrix of UhoF and UheF in layer-k, respectively, I denotes the identity matrix, $\|\cdot\|_F^2$ represents the Frobenius norm. Analogously, the regularization loss of items $\mathcal{L}_{Ireg}^{(k)}$ can be formulated in a similar way.

4.2 Hierarchical Knowledge Transfer

From the perspective of message-passing mechanism, deeper features generally contain global semantics, while the low-level features are able to capture local structural information, both of which are crucial for graph machine learning tasks. Thus, HKT is designed to transfer side information from the $\mathbf{E}^{aux}$ of multiple layers. Specifically, we model the HKT component by extending the ISD from single-layer to multiple-layer.

For hoF, considering the connection paths between the auxiliary views and the interaction view, we design a progressive approach to integrate embeddings from different layers for side information transfer. As shown in Fig. 1(a), in addition to aligning the same layer's $\mathbf{E}^{aux}$ and $\mathbf{E}^{inter}$, the deeper $\mathbf{E}^{aux}$ is also employed to augment the low-level $\mathbf{E}^{inter}$ by fusing with low-level $\mathbf{E}^{aux}$. Here, we utilize the attention-based fusion method (ABF) introduced in [2] as the fusion operation, which can adaptively calculate the weight of the fused portions. For layer-k, the equation 1 can reformulate as

$$\mathcal{L}_{UhoF}^{'(k)} = \sum_{i \in \mathcal{U}} -\log \frac{\exp(sim(\mathbf{e}_i^{UhoF(k)}, \bar{\mathbf{e}}_i^{uu(k,L)})/\tau)}{\sum_{j \in \mathcal{U}} \exp(sim(\mathbf{e}_j^{UhoF(k)}, \bar{\mathbf{e}}_j^{uu(k,L)})/\tau)}, \tag{4}$$

where L represents the total number of layers, and $\bar{\mathbf{e}}^{uu(k,L)} = \mathcal{F}(\mathbf{e}_i^{uu(k)}, \mathbf{e}_i^{uu(k,L)})$ denotes the combination of the layer-k embedding, while $\mathbf{e}^{uu(k)}$ and a residual embedding $\mathbf{e}^{uu(k,L)}$ containing the information from the layer-k to the last layer. $\mathcal{F}(\cdot)$ represents the fusion operation. Notably, for the layer-L, $\mathbf{e}^{uu(L,L)}$ is equal to $\mathbf{e}^{uu(L)}$. Similarly, the objective of IhoF, $\mathcal{L}_{IhoF}^{'(k)}$ is defined in a similar way.

From the perspective of the interaction view, the layer-3 $\mathbf{E}^{aux}$ is aggregated with the layer-2 $\mathbf{E}^{aux}$ to enhance the layer-2 $\mathbf{E}^{inter}$. Therefore, the latter can learn side information from multiple layers of the auxiliary view, enhancing the comprehension of local and global semantics. On the other hand, from the perspective of the auxiliary views, the layer-3 embedding learns the residual of the layer-2 embedding, mitigating the over-smoothing problem for GNNs. Thus, HKT enables the two categories of views to augment themselves in a mutual manner.

Considering the connection paths solely affect the calculation of hoF loss, while heF loss and the orthogonal regularization retain their original formulation. Consequently, the optimization objective of HKT module is defined as:

$$\mathcal{L}_{Uhkt} = \frac{1}{L} \sum_{k=1}^{L} (\mathcal{L}_{UhoF}^{'(k)} + \lambda \mathcal{L}_{Ureg}^{(k)} + \mathcal{L}_{heF}^{(k)}). \tag{5}$$

where λ denotes a hyperparameter to determine the weight of the regularization. $\mathcal{L}_{Ihkt}$ and $\mathcal{L}_{Uhkt}$ share the same computation way.

4.3 Multi-view Semantic Fusion

MSF attempts to explore the collaborative augmentation mechanism that integrates multiple auxiliary views. It bridges two auxiliary views through heF and employs a meta-learning framework to aggregate knowledge of two views, generating adaptive mapping from the aggregated knowledge.

For the user side, we involve an aggregate operation to integrate the user-user embeddings and UheF, that is,

$$M^{uu} = \sigma(\Phi(\mathbf{e}^{uu}, \mathbf{e}^{UheF})), \tag{6}$$

where $\sigma(\cdot)$ represents the activation function. Φ actually acts as a projection that maps the features of the user-user view into a shared semantic space. A similar projection is utilized for the item side M^{ii}. In particular, to smoothly merge the two sides' knowledge M^{uu} and M^{ii}, we design an anchor-neighbor concatenation (ANC). Specifically, we treat one side as the anchor (e.g., a user u) and incorporate the neighborhood of the anchor in the interaction view that belongs to the other side (e.g., the item $i \in \mathcal{N}_u$). The aggregated knowledge is calculated as

$$M^u = M^{uu} \| \sum_{j \in \mathcal{N}_u} M_j^{ii}, M^i = M^{ii} \| \sum_{j \in \mathcal{N}_i} M_j^{uu}, \tag{7}$$

where $\mathcal{N}_u$ and $\mathcal{N}_i$ represent the neighborhood of the user u and the item i, respectively, and $\|$ denotes the concatenation operation.

Subsequently, the aggregated knowledge is utilized to generate a parameterized mapping,

$$W_u^{meta} = \mathrm{MLP}_{\theta_1}(M^u)\mathrm{MLP}_{\theta_2}(M^u), \tag{8}$$

where θ_1 and θ_2 represent the parameters of MLPs. W_i^{meta} is calculated in a similar way. The parameterized mapping contains the knowledge from both two auxiliary views, which serve as a knowledge repository. Given a certain user or item, we can extract the knowledge most relevant to them by applying non-linear transformation to W^{meta} and their intrinsic attribution. Therefore, the enhanced embedding including aggregated knowledge from both two auxiliary views is formulated as

$$\widetilde{\mathbf{E}}^u = \sigma(W_u^{meta}\mathbf{E}^{UhoF}), \widetilde{\mathbf{E}}^i = \sigma(W_i^{meta}\mathbf{E}^{IhoF}). \tag{9}$$

The final output $\widetilde{\mathbf{E}}^u$ and $\widetilde{\mathbf{E}}^i$ can be seen as the integration of the enhanced hoF and heF, which is utilized to predict the interaction of user and item. Here, we also employ BPR loss function to measure the performance of the whole model, that is,

$$\mathcal{L}_{sup} = \sum_{(i,j^+,j^-) \in \mathcal{O}} -\ln(\mathrm{sigmoid}(\hat{y}_{i,j^+}, \hat{y}_{i,j^-})), \tag{10}$$

where $\hat{y}_{i,j} = \widetilde{\mathbf{e}}_i^\top \widetilde{\mathbf{e}}_j$ denotes the likelihood of user i interacting with item j calculated by the enhanced embedding $\widetilde{\mathbf{e}}^i$ and $\widetilde{\mathbf{e}}^u$. The final optimization objective of our model is

$$\mathcal{L}_{total} = \mathcal{L}_{sup} + \alpha\mathcal{L}_{Uhkt} + (1-\alpha)\mathcal{L}_{Ihkt}. \tag{11}$$

where α is a hyperparameter that governs the weight of $\mathcal{L}_{Uhkt}$ and $\mathcal{L}_{Ihkt}$

5 Experiments

In this section, we evaluate the effectiveness of our FiCoRec by exploring the following questions:

- **RQ1:** How does the performance of FiCoRec compare to existing approaches?

– **RQ2:** How do different components in our model contribute to recommendation performance?
– **RQ3:** How do key hyperparameters impact model performance?
– **RQ4:** How does ISD benefit side information transformation?

Table 2. Statistics of experimental datasets

Dataset	#User	#Item	#Interaction	Sparsity
Ciao	6776	101415	265308	99.96%
Epinions	15210	233929	630391	99.98%
FilmTrust	441	1853	12190	98.86%
KuaiRec	472	9030	711680	83.70%
LastFM	1427	6113	18092	99.72%

5.1 Experimental Setup

Experimental Datasets. We conduct experiments on five publicly available datasets: Ciao [17], Epinions [19], FilmTrust [9], KuaiRec [8], LastFM[1]. The details of each dataset are presented in Table 2.

Compared Baselines. To verify the effectiveness of our FiCoRec, we compare it with various baselines. The details of baselines are described as below.

– **GraphRec** [4] firstly adapts GNNs to model the user-user social network and the user-item interaction network to capture heterogeneous relations.
– **SMIN** [16] injects social and knowledge-aware relational structures into the user-item interaction modeling by self-supervised learning.
– **MHCN** [28] designs a multi-channel hypergraph convolutional network to enhance social recommendation by deploying high-order user relations.
– **KCGN** [13] designs a multi-task learning framework combining item interdependent knowledge and social influence to enhance social recommendation.
– **DSL** [22] aims to denoise personalized social information, retaining important social relationships for modeling user preferences.
– **HGCL** [1] transfers heterogeneous relational semantics to user-item interaction modeling with contrastive learning across different views.
– **RecDiff** [14] designs a hidden-space diffusion paradigm to alleviate the noisy effect in compressed and dense representation space.

[1] https://grouplens.org/datasets/hetrec-2011/

Implement Details. Following previous works [1], we utilize Hit Ratio ($H@K$) and Normalize Discounted Cumulative Gain ($N@K$) to measure the recommendation accuracy of various methods. Both metrics are employed with $K = 10$. In the evaluation setup, one positive (interacted) item and 99 negative (non-interacted) items are sampled for each user to assess performance. Besides, we conduct experiments in the PyTorch framework. For all the baselines, we re-implement them in the aforementioned five datasets, and a single RTX 3090 GPU is used for training and testing. The Adam optimizer is used for model optimization, with the learning rate searched from 0.001 to 0.1 and a fixed weight decay rate of 0.05. As hyperparameters, we use a base configuration of hidden state dimension $d = 32$, GNN propagation layers $L = 3$, batch size = 8192, and total loss balanced weight $\alpha = 0.8$ across all datasets. Other hyperparameters are tuned individually for each dataset based on its scale and structure.

Table 3. Performance comparison of all methods on different datasets in terms of NDCG and HR. Best and second performances are marked with **bold** and <u>underline</u>. Δavg denotes relative improvements over all baselines on average, and Δsota indicates the improvements of FiCoRec compared to state-of-the-art methods.

Datasets	CiaoDVD		Epinions		FilmTrust		KuaiRec		LastFM	
Metric	H@10	N@10	H@10	N@10	H@10	N@10	H@10	N@10	H@10	N@10
GraphRec	0.6832	0.4679	0.7610	0.5529	0.8125	0.6319	0.6168	0.3524	0.7262	0.5181
SMIN	0.6895	0.4818	0.8079	0.5997	0.8433	0.6797	0.6695	0.3974	0.7387	0.5145
MHCN	0.7034	0.4876	0.8164	0.6056	0.8429	0.6637	0.6428	0.3956	0.7631	0.5529
KCGN	0.6878	0.4792	0.8187	0.6061	0.8705	0.7276	0.6631	0.4096	0.7555	0.5440
DSL	0.6980	0.4898	0.8138	0.5919	<u>0.8795</u>	**0.7349**	0.6716	0.4103	0.7362	0.5268
HGCL	<u>0.7325</u>	<u>0.5184</u>	<u>0.8226</u>	0.6165	0.8750	0.7095	<u>0.6737</u>	<u>0.4151</u>	<u>0.7875</u>	<u>0.5810</u>
RecDiff	0.7016	0.4934	0.8213	<u>0.6178</u>	0.8763	0.7114	0.6725	0.4029	0.7421	0.5393
FiCoRec	**0.7349**	**0.5319**	**0.8260**	**0.6342**	**0.8977**	<u>0.7310</u>	**0.6907**	**0.4175**	**0.8092**	**0.6726**
Δavg(%)	5.12	9.02	2.47	6.68	4.81	5.58	4.98	5.28	7.98	10.04
Δsota(%)	0.33	2.60	0.41	2.65	2.07	-0.53	2.52	0.58	2.76	2.03

5.2 Overall Performance Comparison(RQ1)

In this section, we conduct experiments over five benchmark datasets and compare our FiCoRec to the state-of-the-art baseline methods. The experimental results of all baseline methods and our model are reported in Table 3. Based on the evaluation results, we identify the following key observations:

(1) Our FiCoRec consistently outperforms baseline methods across five datasets. Notably, FiCoRec achieves the highest average improvement of 7.98% and 10.04% over seven baselines for the H@10 and N@10 metrics on the LastFM dataset. We attribute these improvements to the design of interaction semantic decouple, which provides a fine-grained transformation for side information.

(2) From the table, we also observe that FiCoRec outperforms the state-of-the-art model with a margin of 0.33%-2.76% on the H@10. Notably, the improvement of N@10 is more than H@10 on large-scale datasets where the user-item interaction is sparser. This suggests that FiCoRec effectively enhances the ranking ability and robustness of recommender system, enabling more personalized and interest-oriented recommendations for users. Conversely, the improvement on small-scale datasets illustrates an unstable trend. We hypothesize that it is because user interests are relatively simple in small-scale datasets, making it easier for models to capture user preferences, thus limiting the performance gains achieved by FiCoRec.

(3) As can be seen, HGCL leverages side information from both user's social connections and item's knowledge-aware dependencies, achieving remarkable performance among five datasets. Our FiCoRec not only provides a fine-grained transformation of side information but also explores the joint influence of both user and item. The superior performance further demonstrates the effectiveness of the proposed HKT and MSF module.

5.3 Ablation Study(RQ2)

We conduct ablation studies to assess the significance of customized components of our methods. Specifically, we disable the two key components and evaluate the discrepancy in recommendation performance. The experimental results are shown in Table 4. As can be seen, removing any of the two key components leads to performance degradation across the evaluated dataset. Furthermore, we do not include the user-user social connection view or the item-item knowledge-dependency view to consider their influence. For the influence of components, removing HKT causes a significant drop in performance. This is because it leverages side information and different-level semantics to enhance the interaction view. Without HKT, the decoupled hoF and heF fail to fully capture side information from the auxiliary views. In addition, for the influence of the auxiliary view, we can observe that the performance of FiCoRec outperforms the case where one of the auxiliary views is disabled. This emphasizes the importance of side information transformation, which is similar to the analysis in HGCL[1]. In particular, the user-user view plays a more essential role in our model, indicating the social connection between users more likely to directly influence the preference of the user.

5.4 Impact of Hyperparameters(RQ3)

We perform hyperparameter analysis to investigate the impact of hidden state dimension d, total loss balanced weight α and the total number of propagation layers L on the CiaoDVD, FilmTrust, and LastFM datasets. The results are shown in Fig. 2.

For the parameter d, as the value increases from 8 to 32, we can observe a remarkable improvement on both metrics. This is because the model's capacity

Table 4. Ablation studies on key components of FiCoRec

Datasets	CiaoDVD		FilmTrust		LastFM	
Metric	H@10	N@10	H@10	N@10	H@10	N@10
w/o-hkt	0.7137	0.5133	0.8786	0.7163	0.7820	0.5758
w/o-msf	0.7256	0.5209	0.8803	0.7205	0.7836	0.5827
w/o-uu	0.7068	0.5110	0.8709	0.7042	0.7618	0.5640
w/o-ii	0.7178	0.5202	0.8791	0.7192	0.7789	0.5753
FiCoRec	**0.7349**	**0.5319**	**0.8977**	**0.7310**	**0.8092**	**0.5928**

increases as the hidden dimension enlarges. However, excessive enlargement can degrade performance due to overfitting.

For the parameter α applied to $\mathcal{L}_{total}$, the performance fluctuates as α varies, suggesting that varying proportions of information transfer between user-user view and item-item view influence the recommendation performance. When the proportion of users exceeds 0.8, the model's performance declines, indicating that an optimal balance between users and items is crucial for achieving the best performance.

For the parameter L, as the total number of propagation layers increases, the performance gradually improves, suggesting that hierarchical knowledge transfer facilitates the transformation of side information. We can observe that performance degrades when the value of L reaches 4, which may be attributed to the oversmoothing problem for GNNs.

In addition, the model exhibits different sensitivity to hyperparameters, with smooth performance varying for α but more pronounced variations for L. Thus, optimizing each hyperparameter is crucial for achieving the best performance in specific applications.

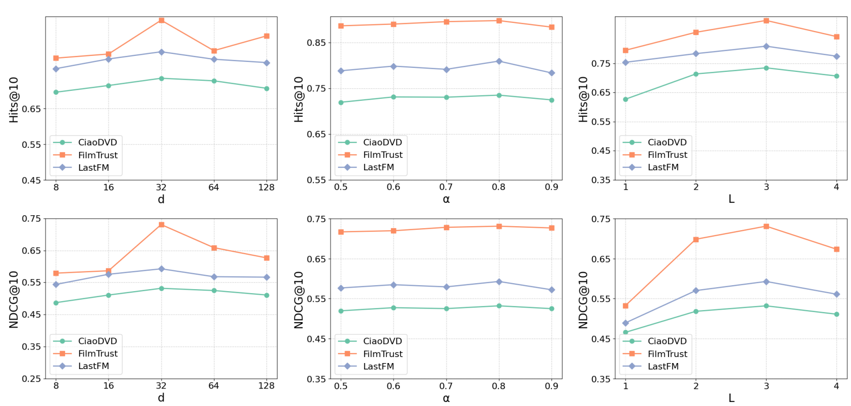
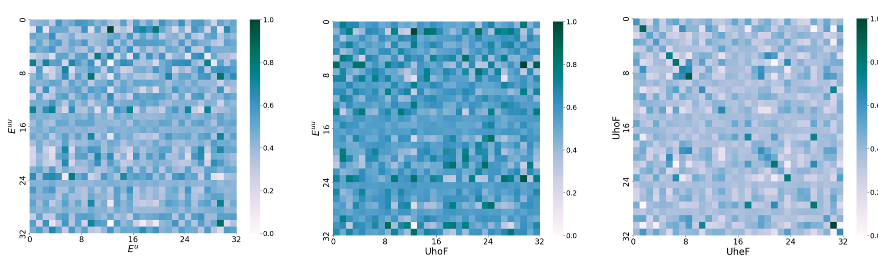


Fig. 2. Hyperparameter studies of the FiCoRec

5.5 Deep Analysis of ISD(RQ4)

To further analyze the effect of the ISD, we visualize the heatmaps of the correlations between the original $\mathbf{E}^{inter}$, the decoupled hoF and heF, and the $\mathbf{E}^{aux}$, which is shown in Fig. 3. Here, we utilize the user embedding, $\mathbf{E}^{u}$ and $\mathbf{E}^{uu}$ on the LastFM dataset as an example. From Fig. 3(a) and (b), we can observe that $\mathbf{E}^{uu}$ shows moderate correlation with $\mathbf{E}^{u}$, but $\mathbf{E}^{uu}$ correlates more with UhoF than $\mathbf{E}^{u}$. This phenomenon indicates two facts: (1) semantic gap exists between $\mathbf{E}^{aux}$ and $\mathbf{E}^{inter}$, hindering the transformation of side information. (2) UhoF obtained through $\mathbf{E}^{uu}$ can effectively capture the user side information. Meanwhile, Fig. 3(c) shows a slight correlation between UheF and UhoF. This suggests that UheF and UhoF learn different feature information, which demonstrates the effectiveness of ISD. Furthermore, we utilize $\mathbf{E}^{u}$ and UheF to predict the user-item interaction without any augmentation and visualize the evaluation results in Fig. 4. As can be seen, the tendency of curves in two figures is comparable, which indicates that the UheF can achieve a performance comparable to the original $\mathbf{E}^{u}$. Notably, the performance of UheF initially exhibits a significant gap compared to $\mathbf{E}^{u}$, which we hypothesize is attributed to the inability of UheF to accurately capture user preference information during the early stages of training. As training progresses, its performance gradually converges to that of $\mathbf{E}^{u}$, eventually stabilizing with a slight margin, indicating the ability of UheF to model user preference.



(a) (b) (c)

Fig. 3. Heatmap of the correlation between $\mathbf{E}^{u}$, $\mathbf{E}^{uu}$, hoF and heF. The axes denote embedding dimensions, and the color intensity reflects similarity magnitude, with darker shades indicating higher similarity. (a) $\mathbf{E}^{uu}$ versus $\mathbf{E}^{u}$, (b) $\mathbf{E}^{uu}$ versus UhoF, (c) UheF versus UhoF.

6 Conclusion

In this paper, we focus on the semantic gap between different views due to the differences in learning behavior and explore the collaborative effect of multiple

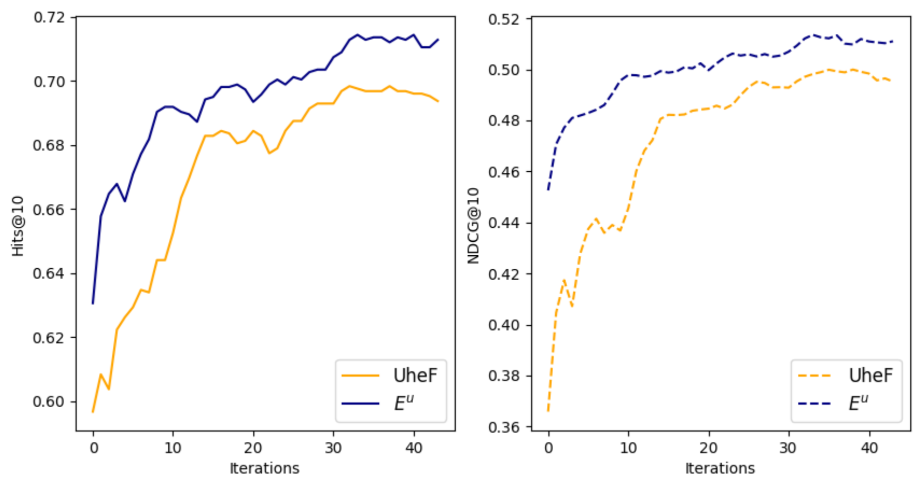


Fig. 4. Evaluation utilizing $\mathbf{E}^u$ and UheF without any augmentation.

views. Thus, we propose a fine-grained augmentation method for recommendation, termed as FiCoRec. It comprises two key components: Hierarchical Knowledge Transfer (HKT) cooperating with interaction feature decoupling process (ISD) and Multi-view Semantic Fusion (MSF). ISD first decouples the interaction view embeddings, and HKT transfers side information knowledge through a hierarchical contrastive learning framework. Then MSF leverages meta-learning to consider the collaborative effect of two auxiliary views. Extensive experimental results confirm that our FiCoRec outperforms the state-of-the-arts. For future work, we aim to explore the enhancement of heterogeneous relationships in multi-behavior scenarios.

Acknowledgments. The works in this paper are supported by the National Natural Science Foundation of China under Grant Nos. 62477015 and 62306079; Key Research and Development Program of Guangdong of China under Grant No. 2023B0303010004; The Innovation Team Project for Universities in Guangdong Province under Grant No. 2023KCXTD011; Guangdong Provincial Natural Science Foundation General Project under Grant No. 2025A1515011637.

References

1. Chen, M., Huang, C., Xia, L., Wei, W., Xu, Y., Luo, R.: Heterogeneous graph contrastive learning for recommendation. In: Proceedings of the sixteenth ACM International Conference on Web Search and Data Mining. pp. 544–552 (2023)
2. Chen, P., Liu, S., Zhao, H., Jia, J.: Distilling knowledge via knowledge review. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5008–5017 (2021)

3. Dong, Y., Chawla, N.V., Swami, A.: metapath2vec: Scalable representation learning for heterogeneous networks. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 135–144 (2017)
4. Fan, W., Ma, Y., Li, Q., He, Y., Zhao, E., Tang, J., Yin, D.: Graph neural networks for social recommendation. In: The World Wide Web conference. pp. 417–426 (2019)
5. Fu, T.y., Lee, W.C., Lei, Z.: Hin2vec: Explore meta-paths in heterogeneous information networks for representation learning. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. pp. 1797–1806 (2017)
6. Fu, X., Zhang, J., Meng, Z., King, I.: MAGNN: Metapath aggregated graph neural network for heterogeneous graph embedding. In: Proceedings of the Web conference 2020. pp. 2331–2341 (2020)
7. Gao, C., Zheng, Y., Li, N., Li, Y., Qin, Y., Piao, J., Quan, Y., Chang, J., Jin, D., He, X., et al.: A survey of graph neural networks for recommender systems: Challenges, methods, and directions. ACM Transactions on Recommender Systems pp. 1–51 (2023)
8. Gao, C., Li, S., Lei, W., Chen, J., Li, B., Jiang, P., He, X., Mao, J., Chua, T.S.: KuaiRec: A fully-observed dataset and insights for evaluating recommender systems. In: Proceedings of the 31st ACM International Conference on Information & Knowledge Management. pp. 540–550 (2022)
9. Guo, G., Zhang, J., Yorke-Smith, N.: A novel bayesian similarity measure for recommender systems. In: Proceedings of the 23rd International Joint Conference on Artificial Intelligence. pp. 2619–2625 (2013)
10. Han, J., Tao, Q., Tang, Y., Xia, Y.: DH-HGCN: Dual homogeneity hypergraph convolutional network for multiple social recommendations. In: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 2190–2194 (2022)
11. He, X., Deng, K., Wang, X., Li, Y., Zhang, Y., Wang, M.: LightGCN: Simplifying and powering gGraph convolution network for recommendation. In: Proceedings of the 43rd International ACM SIGIR conference on Research and Development in Information Retrieval. pp. 639–648 (2020)
12. Huang, C.: Recent advances in heterogeneous relation learning for recommendation. In: International Joint Conference on Artificial Intelligence. pp. 4442–4449 (2021)
13. Huang, C., Xu, H., Xu, Y., Dai, P., Xia, L., Lu, M., Bo, L., Xing, H., Lai, X., Ye, Y.: Knowledge-aware coupled graph neural network for social recommendation. In: Proceedings of the AAAI Conference on Artificial Intelligence. pp. 4115–4122 (2021)
14. Li, Z., Xia, L., Huang, C.: RecDiff: Diffusion model for social recommendation. In: Proceedings of the 33rd ACM International Conference on Information and Knowledge Management. pp. 1346–1355 (2024)
15. Liang, D., Li, B., Li, H., Jiang, Y.: MuHca: Mixup heterogeneous graphs for contrastive learning with data augmentation. In: Pacific Rim International Conference on Artificial Intelligence. pp. 377–388. Springer (2023)
16. Long, X., Huang, C., Xu, Y., Xu, H., Dai, P., Xia, L., Bo, L.: Social recommendation with self-supervised metagraph informax network. In: Proceedings of the 30th ACM International Conference on Information & Knowledge Management. pp. 1160–1169 (2021)
17. Park, C., Kim, D., Xie, X., Yu, H.: Collaborative translational metric learning. In: 2018 IEEE International Conference on Data Mining. pp. 367–376 (2018)

18. Rendle, S., Freudenthaler, C., Gantner, Z., Schmidt-Thieme, L.: BPR: Bayesian personalized ranking from implicit feedback. In: Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence. pp. 452–461 (2009)
19. Richardson, M., Agrawal, R., Domingos, P.: Trust management for the semantic web. In: International Semantic Web conference. pp. 351–368. Springer (2003)
20. Sharma, K., Lee, Y.C., Nambi, S., Salian, A., Shah, S., Kim, S.W., Kumar, S.: A survey of graph neural networks for social recommender systems. ACM Computing Surveys pp. 1–34 (2024)
21. Tian, Y., Dong, K., Zhang, C., Zhang, C., Chawla, N.V.: Heterogeneous graph masked autoencoders. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 9997–10005 (2023)
22. Wang, T., Xia, L., Huang, C.: Denoised self-augmented learning for social recommendation. In: Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence. pp. 2324–2331 (2023)
23. Wang, X., He, X., Wang, M., Feng, F., Chua, T.S.: Neural graph collaborative filtering. In: Proceedings of the 42nd international ACM SIGIR conference on Research and development in Information Retrieval. pp. 165–174 (2019)
24. Wang, X., Ji, H., Shi, C., Wang, B., Ye, Y., Cui, P., Yu, P.S.: Heterogeneous graph attention network. In: The World Wide Web conference. pp. 2022–2032 (2019)
25. Wu, L., Li, J., Sun, P., Hong, R., Ge, Y., Wang, M.: DiffNet++: A neural influence and interest diffusion network for social recommendation. IEEE Transactions on Knowledge and Data Engineering pp. 4753–4766 (2020)
26. Wu, S., Sun, F., Zhang, W., Xie, X., Cui, B.: Graph neural networks in recommender systems: a survey. ACM Computing Surveys pp. 1–37 (2022)
27. Ying, R., He, R., Chen, K., Eksombatchai, P., Hamilton, W.L., Leskovec, J.: Graph convolutional neural networks for web-scale recommender systems. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 974–983 (2018)
28. Yu, J., Yin, H., Li, J., Wang, Q., Hung, N.Q.V., Zhang, X.: Self-supervised multi-channel hypergraph convolutional network for social recommendation. In: Proceedings of the Web conference 2021. pp. 413–424 (2021)