Information Processing and Management 61 (2024) 103701

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/ipm

Cross-view hypergraph contrastive learning for attribute-aware recommendation

Ang Ma [a], Yanhua Yu [a,*], Chuan Shi [a], Zirui Guo [a], Tat-Seng Chua [b]

[a] College of Computer Science, Beijing University of Posts and Telecommunications, Beijing, 100876, China
[b] Sea-NExT Joint Lab, National university of Singapore, Singapore

ARTICLE INFO

Keywords:
Recommender systems
Attribute interactions
Graph neural networks
Hypergraph neural networks
Contrastive learning

ABSTRACT

Recommender systems typically model user–item interaction data to learn user interests and preferences. However, user interactions are often sparse and noisy. Moreover, existing works have constraints in modeling high-order interactions flexibly and sufficiently. We propose a novel Cross-view Hypergraph Contrastive Learning for Attribute-aware Recommendation (CHCLA) to deal with such issues. For considering the high-order interactions, CHCLA models user–item interaction through a graph convolutional network and user/item attribute information through hypergraph convolutional networks, learning user and item representations in two distinct ways. Furthermore, for the problem of sparse and noisy supervised signals, CHCLA uses self-supervised cross-view contrastive learning by generating self-augmented contrastive views between user–item interaction graph and user/item attribute interaction hypergraph to improve the robustness of graph representations learning. CHCLA not only captures user behavior representations but also accommodates inherent user and item attribute preferences. Extensive experiments illustrate that CHCLA performs better than advanced approaches, the NDCG@10 of CHCLA on MovieLens, Book-crossing and Taobao datasets is improved by 1.29%, 2.11% and 8.09% respectively, offering a promising avenue for further research and advancement in recommender systems.

1. Introduction

Personalized recommender systems seek to alleviate information overload on the Internet, and assist users in identifying items of interest. It has been extensively utilized in various fields, for instance E-commerce (e.g., Amazon, Alibaba) (Su, Zhang, Erfani and Gan, 2021), social network (e.g., Facebook, Twitter) (Bu et al., 2010), course recommendation (e.g., MOOC) (Wang et al., 2022) and content-sharing websites (e.g., Google News, Bing News) (Ge, Wu, Wu, Qi, & Huang, 2020; Xia, Yin, Yu, Wang et al., 2021).

In recommendations, collaborative filtering (CF) is a widely used approach. To make the predictions, it was assumed that users with similar behaviors would have similar preferences. Matrix factorization (MF) learns the embeddings of users and items on user–item interactions (e.g., purchase, click, add to cart). However, these approaches consider only low-order interactions between users and items. Powered by graph neural networks (GNNs), researches such as GCMC (Berg, Kipf, & Welling, 2017), NGCF (Wang, He, Wang, Feng, & Chua, 2019), LightGCN (He et al., 2020), etc. focus on leveraging GNNs to model higher-order user–item interactions. The above works mainly considered the structural data of the user–item interactions. However, as shown in Fig. 1(a), attributes (e.g., user gender, age, product type, and brand) are essential side-information (Ni et al., 2023), which can alleviate the cold-start

* Corresponding author.
E-mail addresses: ang@bupt.edu.cn (A. Ma), yuyanhua@bupt.edu.cn (Y. Yu), shichuan@bupt.edu.cn (C. Shi), zrguo@bupt.edu.cn (Z. Guo), dcscts@nus.edu.sg (T.-S. Chua).

https://doi.org/10.1016/j.ipm.2024.103701
Received 26 September 2023; Received in revised form 8 December 2023; Accepted 28 February 2024
Available online 10 March 2024

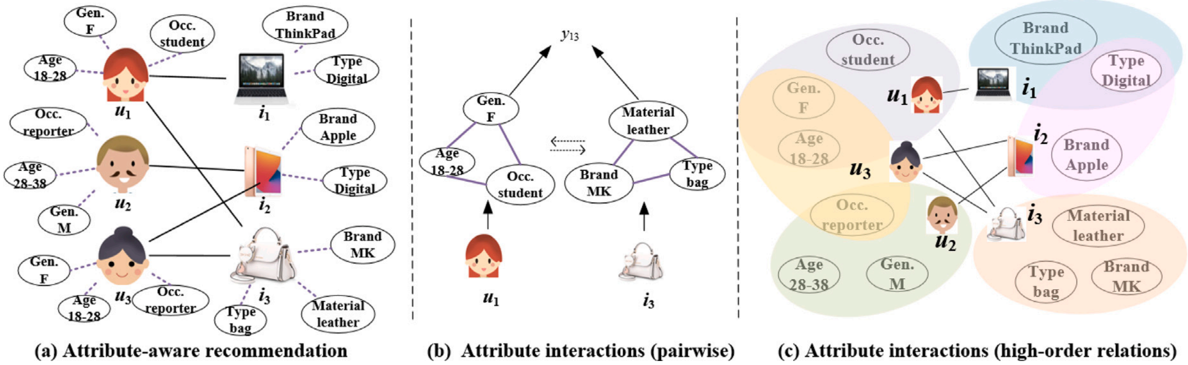


Fig. 1. Illustration of the attribute-aware recommendation scenario and the differences between existing work and CHCLA.

problem. It has already been proven that adding attribute information to recommendations can enhance performance (Li et al., 2021; Luo et al., 2022). One of the key issues is how to use user and item attributes to enhance the recommendation performance. Fi-GNN (Li, Cui, Wu, Zhang, & Wang, 2019) represents the attributes in a graph structure, nodes in a graph correspond to attributes and edges correspond to attribute interactions. For example, in movies recommendations, *Paul* is an *action movie*'s director. (*Paul*, *action movies*) can be regarded as an attribute interaction. Recently, several approaches (Su, Zhang, Erfani and Xu, 2021; Su, Zhang, Erfani, Gan, 2021) model the attribute interactions by graph, as shown in Fig. 1(b), and use graph neural networks to capture such pairwise interactions for recommendation. Nevertheless, the graph structure has limits on high-order relations modeling, as only pairwise connections can be represented in a graph. Hypergraph (Gao et al., 2020) is a generalization of a graph that has significant strengths in modeling the high-order relations that can be more sophisticated than pairwise relations, as shown in Fig. 1(c).

Motivated by these observations, we attempted to design hypergraphs to model user and item attribute interactions for high-order relations. However, this is not an easy task due to the following challenges.

– Sparse supervised signal. In real-world recommender systems, the users–item interactions usually contain their preferences. However, such user–item interactions are frequently sparse, which causes a cold-start problem when the user or item has no or very few interactions.
– Noise interaction data. User–item interactions contain noise. Recommender systems typically use implicit feedback from user behaviors, such as clicks, purchases, etc. as training data. Previous work (Xia, Huang, Xu et al., 2022), however, highlighted certain distinctions between implicit input and genuine user preferences.
– Inflexible and insufficient modeling of high-order relations. High-order user–item interactions are crucial. Existing studies have sought to utilize graphs to model high-order relations; however these methods are limited because graphs can only express pairwise relations.

In recommendation systems, the phenomenon of popularity bias (Yu et al., 2022) has emerged as a critical aspect that significantly influences the effectiveness and fairness of recommendations. Popularity bias refers to the inherent tendency of recommendation algorithms to favor popular items over niche or less well-known ones. Over time, this results in a cumulative advantage for popular items, making them even more popular, while less popular ones continue to decline in popularity. Understanding and addressing popularity bias is of paramount importance as it directly impacts user experience, system diversity, and the potential for serendipitous discovery. Unraveling the intricacies of popularity bias not only contributes to enhancing the accuracy of recommendation algorithms but also holds the key to fostering a more inclusive and diverse recommendation.

To address these challenges, we proposed a novel approach called **C**ross-view **H**ypergraph **C**ontrastive **L**earning for **A**ttribute-aware Recommendation (CHCLA), which learns better high-order representations for users and items. CHCLA employs a cross-view contrastive learning framework with self-augmentation, leverages a graph to model user–item interaction information, and utilizes hypergraphs to model user/item attribute information. Specifically, to obtain local user/item representations, we used a graph convolutional network to model pairwise user–item interactions. Next, for high-order relations, hypergraph structures are employed to model user and item attributes. Thus, the global user/item representations were obtained. The local and global representations were then combined to form the final representations of the user/item for prediction. Ultimately, local and global structures take the place of unified data augmentation, such as corruption or dropping, for contrastive learning. In the experiments, we evaluated CHCLA on three benchmark datasets. The experiment results illustrate that CHCLA achieves the most advanced performance.

In summary, the primary contributions are listed as follows:

– We emphasized the significance of explicitly modeling both user/item attribute information and user–item interactions information in recommendation. We used hypergraphs to model user and item attributes separately, learn high-order interactions information, and use the graph structure to model pairwise user–item interaction information. The CHCLA explores both attribute and interaction preferences to achieve a better balance.

– We introduce cross-view contrastive learning, which enhances the robustness of graph collaborative representations learning by generating self-augmented contrastive views between the user–item interaction graph and the user/item attribute interaction hypergraph.
– We conducted an extensive number of experiments. The results illustrate that CHCLA performs better than advanced baselines in producing accurate and post-hoc interpretability predictions.

2. Related work

In this section, relevant studies in three primary domains are presented, including attribute-aware collaborative filtering, graph-based recommendation, and graph-based contrastive learning.

2.1. Attribute-aware collaborative filtering

One of the most extensively utilized methods in recommendation is collaborative filtering (CF). The key to CF approaches is designing appropriate embedding functions to learn user and item representations. FM (Rendle, 2010) performs attribute interactions as a dot product of embeddings and aggregates linearly. AFM (Xiao et al., 2017) uses an attention mechanism to consider the significance of second-order attribute interactions. The FFM (Juan, Zhuang, Chin, & Lin, 2016) and FNFM (Zhang, Shen, Huang, Li, & Pan, 2019) further take the field information into account and introduce field-aware embeddings. The above studies only modeled second-order interactions and their representational power was limited owing to the linearity modeling. To overcome these limitations, FNN (Zhang, Du, & Wang, 2016) uses deep neural networks (DNNs) to automatically learn non-linear feature interactions. Some studies (Guo, Tang, Ye, Li, & He, 2017; He & Chua, 2017) combine FM and neural networks and use Bi-Interaction operation to learn more useful feature interactions. Nevertheless, a straightforward unstructured feature combination would necessarily restrict the ability to explicitly and flexibly model complex relations between various fields.

2.2. Graph-based recommendation

Recently, researchers (Berg et al., 2017; He et al., 2020; Li et al., 2019; Su, Zhang, Erfani, Xu, 2021; Wang et al., 2019) have investigated graph structures to model user–item interactions. NGCF (Wang et al., 2019) and LigtGCN (He et al., 2020) learn the embeddings of the user and item on user–item interaction with graph neural networks. GMCF (Berg et al., 2017), Fi-GNN (Li et al., 2019), and l_0-SIGN (Su, Zhang, Erfani, Xu, 2021) use GNNs to model attribute interactions for recommendations. To capture each feature that may behave differently while interacting with others, GraphFwFM (Zhai, Yang, & Zhang, 2023) leverages gated graph neural networks to establish feature representations of causal relationships among field features. In session-based recommendation systems, COTREC (Xia, Yin, Yu, Shao and Cui, 2021) is an effective method to solve the problem of data sparsity. COTREC constructs two types of graphs (Item view and Session view) to capture the internal and external correlation between different items from the perspectives of intra-session and inter-session respectively, and establishes two graph encoders for comparative learning, thus generating self-supervised signals. In MKCGN (Chang et al., 2023), a meta-relation-guided graph neural network is employed to capture collaborative signals emanating from the user and item domains. Unfortunately, the lack of modeling of both high-level user–item interaction and user/item attributes interaction has limited performance improvement.

Hypergraph neural networks, which model the complex high-order relations in graphs, have been extensively used. Some recent works capture user–item interaction patterns through constructed hypergraph structures and node-hyperedge dependency matrices. DHCF (Ji et al., 2020) and HCCF (Xia, Huang, Xu et al., 2022) models the hybrid multi-order relations between user and item. In sessions recommendation, DHCN (Xia, Yin, Yu, Wang et al., 2021) adpots hypergraph to represent the high-order relationships among items inside sessions. CoHHN (Zhang et al., 2022) integrated price factors in session recommendation, using hypergraphs to model items in a session and their characteristics (prices and categories). A two-channel aggregation mechanism and attention layer are designed to extract users' price preferences and interest preferences. And CoHHN adpots a co-guided learning to model the complex relationship between price preferences and interest preferences. In social recommendation, MHCN (Yu et al., 2021) is proposed to obtain comprehensive user representations aggregating the embeddings learned by multiple channels, where each channel encodes a hypergraph describing a common high-order user relation. However, there has been limited research on modeling user and item attribute interaction with hypergraphs.

2.3. Graph-based contrastive learning

In recent years, contrastive learning (CL) (You et al., 2020) has been extensively utilized in computer vision, natural language processing, and recommender systems. Because CL is a self-supervised method that does not require data annotation, it is a natural solution for data sparsity in recommendations. A standard method of CL in recommendation is to augment the user–item network with structural perturbations to generate multiple views and then maximize the similarity between multiple views of one node in comparison to those of other nodes. SGL (Wu et al., 2021) takes node and edge dropout as data augmentations and utilizes InfoNCE loss for CL. SimGCL (Yu et al., 2022) is a simple CL method that uses uniform noise in the embedding space for graph augmentation. The SHT (Xia, Huang and Zhang, 2022) improves the robustness of the graph collaborative filtering paradigm by combining hypergraph neural networks with topologically aware Transformers. To improve the semantic representation of nodes for sequential recommendation, MASR (Duan, Zhu, Liang, Zhu, & Liu, 2023) proposed a semantic-enriched contrastive learning method to alleviate noise from the structure. In cross-domain recommendation, MCCLK (Zou et al., 2022) integrates the information of multiple views using CL rather than dynamic linear weighting to improve the expressive ability of networks. Inspired by these works, we utilize contrastive learning to promote user–item interaction learning and attribute interaction learning.

2.4. Difference

Our work exhibits notable distinctions from existing research in multiple aspects: (1) While previous research has utilized graph structures to model attributes feature interactions, many of these works primarily explore pairwise feature interactions (node-node relationships). In contrast, our approach employs hypergraphs to model higher-order interactions involving multiple attribute nodes, which can enrich learned representations with a more comprehensive set of information. (2) Although the use of hypergraphs in recommendation systems has been discussed in previous works, existing approaches mostly focus on user–item interactions. Consequently, their effectiveness in attribute-aware recommendations currently remains unexplored. (3) We use a comprehensive strategy that considers both user–item interactions and interactions between user/item attributes. To enhance the integration of these two aspects, we employ contrastive learning. It is worth noting that our contrastive learning approach deviates from traditional methods that augment graphs with structural perturbations to generate multiple views. Instead, we utilized these user–item interaction graph and user/item attribute interaction graph as self-augmented contrastive views. In comparison to existing techniques, CHCLA not only captures user behavior representations through user–item interactions but also incorporates inherent attribute preference representations of users and items via attribute interactions, which can significantly enhance recommendation performance.

3. Problem formulation

First, we describe two forms of essential structural data, namely the user–item interaction graph and the user/item attribute hypergraph. We then introduce the problem formulation of attribute-aware recommendations.

3.1. Interaction graph and attribute hypergraph

User–Item Interaction Graph. The interactions of user and item can be modeled naturally using a bipartite graph. Users or items are denoted as nodes and the interactions between them are denoted as edges. Previous studies (He et al., 2020; Wang et al., 2019) show that high-order interactions contain rich semantics for recommendations. In Fig. 1(a), the path $u_1 \rightarrow i_3 \rightarrow u_3$ reveals the behavioral similarity between u_1 and u_3, as i_3 is the common item that both two users have interacted; A longer path $u_1 \rightarrow i_3 \rightarrow u_3 \rightarrow i_2$ suggests that u_1 likely choose i_2, because her similar user u_3 has already chosen i_2 before.

User/Item Attribute Hypergraph. A hypergraph is formed using a set of nodes and hyperedges, with each hyperedge connecting any number of nodes. Users or items can be described using multiple attributes and different users or items may have the same attributes. Therefore, each node is defined as an attribute and each hyperedge is a user (or item). In the hypergraph message passing architecture, user/item attribute hypergraph learning can capture complicated high-order relations.

3.2. Attribute-aware recommendation

Let $\mathcal{D} = \left\{\left(u_i, v_j, y_n\right)\right\}_{1\le n\le N}$ be a set of data samples and $\mathcal{U} = \left\{u_1, u_2, \ldots, u_{|\mathcal{U}|}\right\}$, $\mathcal{V} = \left\{v_1, v_2, \ldots, v_{|\mathcal{V}|}\right\}$ represent the user set and the item set, respectively. We denote $\mathcal{X}^U$ as the user attributes set and $\mathcal{X}^V$ as the item attributes set. Each user u_i consists of a set of attributes $\mathcal{X}_i^U \subseteq \mathcal{X}^U$ and each item v_j consists of a set of attributes $\mathcal{X}_j^V \subseteq \mathcal{X}^V$. $y_n \in \{0, 1\}$ indicates the feedback labels. For example, $\mathcal{X}^U = \{Female, Male, 18-28, 29-38\}$, the user attributes set of u_1 is denoted by $\mathcal{X}_1^U = \{Female, 18-28\}$, while the user attributes set of u_2 is denoted by $\mathcal{X}_2^U = \{Male, 29-38\}$.

Given the user set $\mathcal{U}$, the item set $\mathcal{V}$, user attributes set $\mathcal{X}^U$ and item attributes set $\mathcal{X}^V$, attribute-aware recommendation aims to create a prediction model $F(u_i, v_j)$ so that for the input data u_i, v_j, the output of the model is a prediction of the user's true feedback y_n on the item.

4. Methodology

In this section, we begin with an overview of the CHCLA before focusing on the details. Then, we present an in-depth description of each part of the model.

4.1. Overview

CHCLA is primarily composed of four parts: local collaborative interaction learning, global hypergraph interaction learning, aggregation and prediction, and cross-view contrastive learning. Fig. 2 shows an overview of CHCLA. CHCLA takes full advantage of user–item interaction and attribute interactions to improve the recommendation performance. To achieve this, we leverage a graph neural network to obtain information about local pairwise interactions between user and item (Fig. 2(a)). Next, we utilize the hypergraph structure to model higher-order attribute interactions and comprehensively capture global attribute-level collaboration information (Fig. 2(b)). A fusion operation is then utilized on the output of user/item representations to make predictions (Fig. 2(c)). To address the issue of sparse supervised signal, we propose cross-view contrastive learning that generates self-augmented contrastive views from local and global learning (Fig. 2(d)).

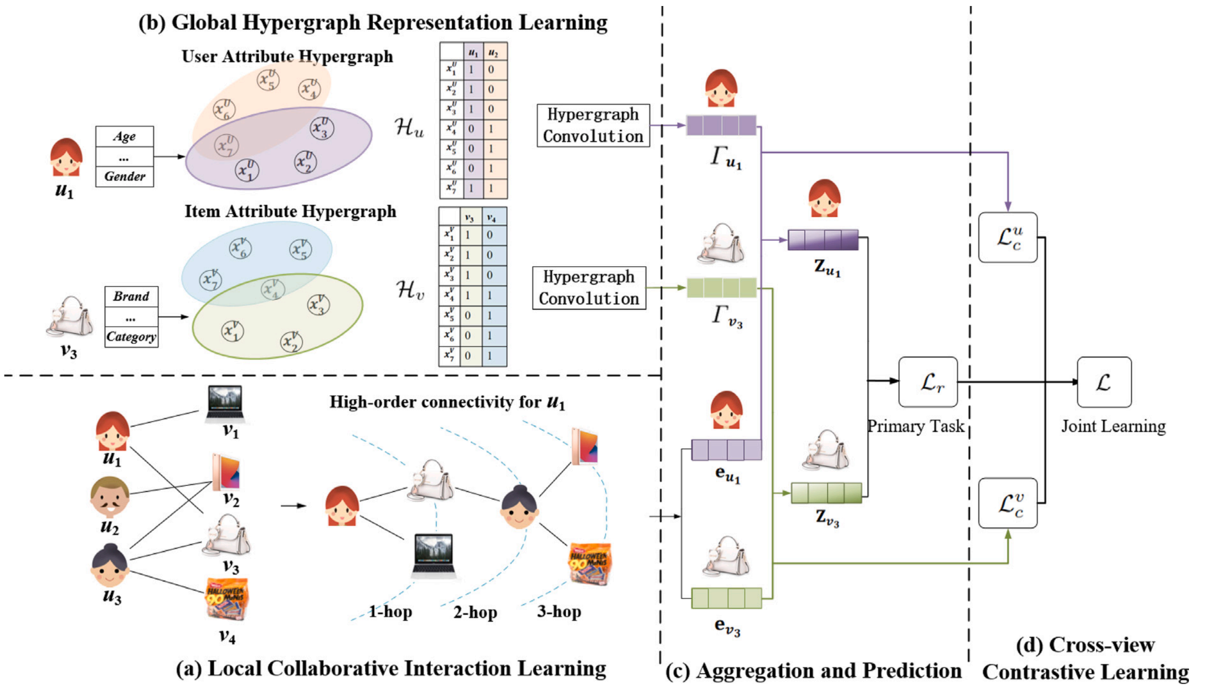


Fig. 2. An Overview of CHCLA.

4.2. Local collaborative interaction learning

Local collaborative interaction learning learns the representations of users and items from the pairwise user–item interactions. We characterize user u_i (or item v_j) with the embedding vector $\mathbf{e}_{u,i} \in \mathbb{R}^d$ $\left(\mathbf{e}_{v,j} \in \mathbb{R}^d\right)$, where d denotes the embedding size. Following the success of the simplified graph convolutional network in LightGCN (He et al., 2020), feature transformation and nonlinear activation were dropped in favor of the straightforward weighted sum aggregator. The graph convolution operation is defined as follows:

$$\begin{aligned}
\mathbf{e}_{u,i}^{(k+1)} &= \sum_{c \in \mathcal{N}_{u,i}} \frac{1}{\sqrt{\left|\mathcal{N}_{u,i}\right|}\sqrt{\left|\mathcal{N}_{u,c}\right|}} \mathbf{e}_{u,c}^{(k)}, \\
\mathbf{e}_{v,j}^{(k+1)} &= \sum_{c \in \mathcal{N}_{v,j}} \frac{1}{\sqrt{\left|\mathcal{N}_{v,j}\right|}\sqrt{\left|\mathcal{N}_{v,c}\right|}} \mathbf{e}_{v,c}^{(k)},
\end{aligned} \tag{1}$$

where $\mathcal{N}_{u,i}$ and $\mathcal{N}_{v,j}$ represent the neighbors of user u_i and item v_j, respectively. $\mathbf{e}_{u,i}^{(k+1)}$ is the representation of the target node and its neighboring nodes in the $(k+1)$th layer.

4.3. Global hypergraph representation learning

Global hypergraph representation learning uses hypergraph structures to model high-order attribute interactions between users and items. Learn representations of users and items under the assumption that users with the same attributes may have comparable interests and preferences. When a new user or item appears, recommendations can be made based on the information provided by the attributes.

A hypergraph comprises nodes and hyperedges, with each hyperedge connecting several nodes. It describes the high-order relationships between attribute interactions. Specifically, the hypergraph incidence matrix for user/item is defined as $\mathcal{H}_u \in \mathbb{R}^{\left|\mathcal{X}^U\right| \times |U|}$ and $\mathcal{H}_v \in \mathbb{R}^{\left|\mathcal{X}^V\right| \times |V|}$, where $\left|\mathcal{X}^U\right|$ and $\left|\mathcal{X}^V\right|$ are the total number of user attribute nodes and item attribute nodes, respectively, and $|U|$ and $|V|$ are the total users and items, respectively. We constructed the hypergraph incidence matrix for users $\mathcal{H}_u$ according to the attribution relations between each user u_i and their attribute set $\mathcal{X}_i^U$. For example, in Fig. 2(b), user 1 has attributes $\mathcal{X}_1^u = \left\{x_1^U, x_2^U, x_3^U, x_7^U\right\}$, then the elements in $\mathcal{H}_u(1,1), \mathcal{H}_u(2,1), \mathcal{H}_u(3,1), \mathcal{H}_u(7,1)$ are recorded as 1.

$$\mathcal{H}_u(l,i) = \begin{cases} 1, if\, x_l^U \in \mathcal{X}_i^U \\ 0, else \end{cases}, \tag{2}$$

where $\mathcal{H}_u(l,i)$ represents the element of the hypergraph incidence matrix in the lth row and ith column, denoted as 1 if user attribute $x_{u,l}$ is included in the attribute set $\mathcal{X}_i^U$ of u_i, and 0 otherwise. The hypergraph incidence matrix for items $\mathcal{H}_v$ can be obtained in a similar manner. The hypergraph message passing can be denoted as,

$$X_u^{(k+1)} = \sigma\left(\mathcal{H}_u \Lambda_u^{(k)}\right) = \sigma\left(\mathcal{H}_u f(\mathcal{H}_u^T X_u^{(k)})\right), \tag{3}$$

where $X_u^{(k)} \in \mathbb{R}^{\left|\mathcal{X}^U\right| \times d}$ denotes the kth layer's representation of user attributes nodes, σ denotes the activation function (e.g. LeakyReLU), $\Lambda_u^{(k)} \in \mathbb{R}^{\left|\mathcal{X}^U\right| \times d}$ denotes the hyperedge embeddings for users. $f(\cdot)$ is a neural method that nonlinearly models the hyperedge information. $X_v^{(k+1)}$ can be calculated similarly.

4.4. Aggregation and prediction

Before diving into the fusion of user–item interaction learning and attribute interaction learning, we first provide the aggregation at each layer.

Local Aggregation. For local collaborative interaction learning, after K layers, we can obtain the user/item embeddings obtained at each layer, denoted by $e_u^{(0)}, e_u^{(1)}, \dots, e_u^{(K)}$ (for item, $e_v^{(0)}, e_v^{(1)}, \dots, e_v^{(K)}$). The local user/item representation at the user–item interaction level is then formed by combining the embeddings generated at each layer,

$$\mathbf{e}_u = \sum_{k=0}^{K} \mathbf{e}_u^{(k)}; \quad \mathbf{e}_v = \sum_{k=0}^{K} \mathbf{e}_v^{(k)}, \tag{4}$$

Global Aggregation. Similarly, after K layers, we can get the user/item attributes embeddings obtained at each layer denoted as, $X_u^{(0)}, X_u^{(1)}, \dots, X_u^{(K)}$ (for item, $X_v^{(0)}, X_v^{(1)}, \dots, X_v^{(K)}$). Embeddings obtained at each level were combined,

$$X_u = \sum_{k=0}^{K} X_u^{(k)}; \quad X_v = \sum_{k=0}^{K} X_v^{(k)}. \tag{5}$$

Then according to the hypergraph incidence matrix, we can obtain the global user and item representation,

$$\Gamma_u = \mathcal{H}_u^\top X_u; \quad \Gamma_v = \mathcal{H}_v^\top X_v. \tag{6}$$

Local and Global Fusion. To combine the local interaction learning with the global attribute interaction learning, we conducted the local and global information fusion. After performing multilayer aggregation, we obtain two-level representations for the user, namely $\mathbf{e}_u$, Γ_u, analogous to item, namely $\mathbf{e}_v$, Γ_v. The final user and item representations are defined as,

$$\mathbf{z}_u = \psi(\mathbf{e}_u, \Gamma_u); \quad \mathbf{z}_v = \psi(\mathbf{e}_v, \Gamma_v), \tag{7}$$

where $\psi(\cdot)$ is a fusion function, such as the sum/mean operation, concatenation operation, attention operation, etc. In CHCLA, we adopted the concatenation as the fusion function owing to its simplicity and effectiveness.

Prediction The model prediction of user u_i and item v_j is obtained by calculating the inner product of the user and item final representations:

$$\hat{y}_{i,j} = \mathbf{z}_{u,i}^T \mathbf{z}_{v,j}. \tag{8}$$

4.5. Cross-view contrastive learning

Here we describe the use of cross-view contrastive learning to augment user and item representations with sparse supervision signals.

Contrastive Learning methods learn how to represent a node by comparing the positive and negative pairs. It focuses on learning the common characteristics between similar instances and distinguishing the differences between non-homogeneous instances. Many studies (Wu et al., 2021; Yu et al., 2022) conduct contrastive learning between original graph and corrupted graphs (e.g. dropping nodes or edges, adding uniform noises).

Different from these works, we consider different views (e.g. local and global) of the same user/item as positive samples and views of different users/items as negative samples. In this way, CHCLA can learn the representations via comparing the generated positive pairs $(\mathbf{e}_{u,i}, \Gamma_{u,i})$ and negative pairs $(\mathbf{e}_{u,i}, \Gamma_{u,i'})$. With the InfoNCE (Yu et al., 2022), the contrastive loss of user and item representations was defined.

$$\mathcal{L}_c^u = \sum_{i=0}^{|\mathcal{U}|} -\log \frac{\exp\left(s\left(\mathbf{e}_{u,i}, \Gamma_{u,i}\right)/\tau\right)}{\sum_{i'=0}^{|\mathcal{U}|} \exp\left(s\left(\mathbf{e}_{u,i}, \Gamma_{u,i'}\right)/\tau\right)}, \tag{9}$$

$$\mathcal{L}_c^v = \sum_{j=0}^{|\mathcal{V}|} -\log \frac{\exp\left(s\left(\mathbf{e}_{v,j}, \Gamma_{v,j}\right)/\tau\right)}{\sum_{j'=0}^{|\mathcal{V}|} \exp\left(s\left(\mathbf{e}_{v,j}, \Gamma_{v,j'}\right)/\tau\right)}, \tag{10}$$

where $s(\cdot)$ denotes the similarity function (e.g. cosine similarity), τ denotes the temperature hyperparameter that adjust the attention to challenging samples. The sample was distinguished from those that were most similar to the sample with decreasing temperature. Cross-view contrastive learning was used to solve the issue of sparse supervised signals by utilizing self-augmented contrastive views generated by local learning and global learning.

Table 1
The statistics of data set.

Data	#Interaction	#User	#Item	#U-Attribute	#I-Attribute
MovieLens1M	1,149,238	5950	3514	30	6944
Book-Crossing	1,050,834	4873	53,168	87	43,157
Taobao	2,599,463	4532	371,760	36	434,254

4.6. Joint learning

To combine the recommendations with self-supervised contrastive learning, a joint learning strategy was used to optimize the entire model.

In the recommendation task, the binary cross-entropy loss (Su, Zhang, Erfani, Gan, 2021) was adopted,

$$\mathcal{L}_r = -\frac{1}{N}\sum_{n=1}^{N}[y_n \cdot \log\left(\hat{y}_n\right) + \left(1 - y_n\right) \cdot \log\left(1 - \hat{y}_n\right)], \tag{11}$$

where N is the total number of the data samples, and y_n is the ground truth label of nth data sample.

By combining the user and item contrastive losses with the binary cross-entropy loss, the model parameters can be learned by minimizing the final loss function.

$$\mathcal{L} = \mathcal{L}_r + \beta(\mathcal{L}_c^u + \mathcal{L}_c^v) + \lambda\|\Theta\|_2^2, \tag{12}$$

where β and λ are two hyper parameters that control the loss of contrastive and L_2 regularization term, respectively. Θ are all model parameters.

5. Experiment

First, the datasets and baselines used in our experiment are introduced. Next, the overall performance is thoroughly illustrated. We then evaluated how the key parts of our model contributed to the performance, and the effects of the model's hyperparameters. Next, we evaluated whether the CHCLA could alleviate popularity bias and data sparsity. We also offered a visualization of the user and item representations in the last section.

5.1. Experimental setups

5.1.1. Datasets

Three benchmark datasets were utilized for the experiments: MovieLens 1M (Harper & Konstan, 2015), Book-crossing (Ziegler, McNee, Konstan, & Lausen, 2005), and Taobao (Zhou et al., 2018). The statistics of the dataset are illustrated in Table 1. MovieLens 1M is a dataset of movie recommendations containing anonymous implicit and explicit movie ratings made by users who signed up for MovieLens in 2000. The dataset included movies and users with their associated attributes. In previous works (Su, Zhang, Erfani, Gan, 2021; Su, Zhao, Erfani, Gan, & Zhang, 2022), the datasets were enriched with additional movie-related attributes from IMDB, such as directors and casts. Following these studies, we utilized enriched data. Book-crossing is a book recommendation dataset that contains users' explicit and implicit ratings. Users and books with the associated attributes are present in instances. Taobao is a product recommendation dataset that records clicks on advertisements presented on website. Each sample included a user (or item) with the associated attributes. More details about the approaches for preprocessing the datasets can be found in previous studies (Su, Zhang, Erfani, Gan, 2021; Su et al., 2022).

5.1.2. Implementations

All experiments in this paper were conducted using Python 3.7, PyTorch 1.8.1 and the CUDA 11.1. Following the settings in Su et al. (2022), the datasets were split into training, validation, and test sets using a 7:1.5:1.5 random split in our implementations. The test data were only utilized to evaluate the models, whereas the validation data were only utilized to determine the appropriate parameter settings. The learnable parameters were initialized and optimized using the Xavier initializer. Adam (Kingma & Ba, 2015) is employed as the optimizer, and the learning rate is set at 1×10^{-3}. The embedding size is fixed at 64, and 2 propagation layers are used for GNNs and HyperGNNs. The regularization weight λ is configured as 1×10^{-5}, the batch size is configured as 1024, and the dropout ratio is chosen among $\{0.2, 0.4, 0.6, 0.8\}$. The temperature parameter is chosen among $\{0.1, 0.5, 1, 5\}$. Followed by Su, Zhang, Erfani, Gan (2021) and Su et al. (2022), we use AUC, Recall@10, Recall@20, NDCG@10, and NDCG@20 as evaluation metrics. The experiments were repeated five times and the values in the tables represent the average results.

Table 2
Overall performance of CHCLA with the baselines.

	MovieLens 1M				Book-crossing				Taobao			
	Recall@10	Recall@20	NDCG@10	NDCG@20	Recall@10	Recall@20	NDCG@10	NDCG@20	Recall@10	Recall@20	NDCG@10	NDCG@20
FM	0.9368	0.9570	0.9241	0.9322	0.8933	0.9394	0.8647	0.8849	0.2284	0.3988	0.1696	0.2366
AFM	0.9522	0.9652	0.9427	0.9469	0.8964	0.9391	0.8679	0.8862	<u>0.2333</u>	0.4061	0.1697	0.2379
NFM	0.9557	0.9665	0.9478	0.9507	0.9049	0.9457	0.8758	0.8942	0.2311	<u>0.4091</u>	<u>0.1711</u>	<u>0.2401</u>
DeepFM	0.9533	0.9659	0.9441	0.9481	0.9011	0.9421	0.8729	0.8908	0.2266	0.4015	0.1660	0.2347
FNFM	0.9524	0.9647	0.9445	0.9481	0.9042	0.9457	0.8777	0.8960	0.2260	0.3854	0.1710	0.2339
Fi-GNN	0.9524	0.9661	0.9416	0.9465	0.9115	0.9490	0.8832	0.9007	0.2306	0.3928	0.1696	0.2333
GMCF	0.9521	0.9655	0.9435	0.9479	<u>0.9318</u>	<u>0.9594</u>	<u>0.9131</u>	<u>0.9248</u>	0.2166	0.3929	0.1582	0.2275
HIRS	<u>0.9538</u>	<u>0.9664</u>	<u>0.9459</u>	<u>0.9500</u>	0.9267	0.9576	0.9066	0.9199	0.2286	0.4076	0.1692	0.2398
CHCLA	**0.9634***	**0.9722***	**0.9581***	**0.9602***	**0.9449***	**0.9663***	**0.9324***	**0.9407***	**0.2444***	**0.4097***	**0.1849***	**0.2503***
Improv.	1.01%	0.59%	1.29%	1.07%	1.41%	0.72%	2.11%	1.72%	4.75%	0.15%	8.09%	3.41%

* CHCLA outperforms the strongest baseline through Wilcoxon signed rank test with the significance of $\alpha = 5\%$.

Table 3
AUC of the comparative methods.

	MovieLens 1M	Book-crossing	Taobao
FM	0.8781	0.7372	0.6178
AFM	0.9001	0.7415	0.6466
NFM	0.9063	0.7745	0.6505
DeepFM	0.9022	0.7569	0.6128
FNFM	0.9005	0.7687	0.6099
Fi-GNN	0.9029	0.7947	0.5815
GMCF	0.9005	<u>0.8316</u>	0.6506
HIRS	<u>0.9040</u>	0.8202	<u>0.6517</u>
CHCLA	**0.9199**	**0.8479**	**0.6649**
Improv.	1.76%	1.97%	1.59%

5.1.3. Baselines

We compare CHCLA with benchmarks that consider both item and user attributes:

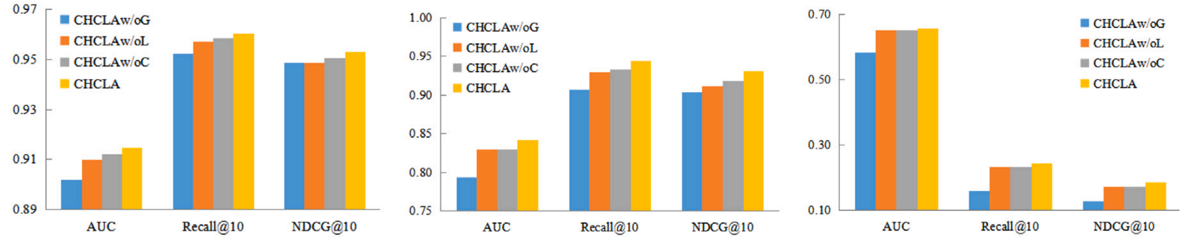
– Factorization Machine (FM) (Rendle, 2010) takes into account second-order attribute interactions by inner product and all the results are added to make the final prediction.
– Attentional Factorization Machine (AFM) (Xiao et al., 2017) is an FM extension that uses an attention mechanism to account the significance of various second-order feature interactions.
– Neural Factorization Machine (NFM) (He & Chua, 2017) uses a Bi-Interaction Pooling layer for modeling second-order interactions, while using concatenated second-order combination features by DNNs to model high-order interactions.
– DeepFM (Guo et al., 2017) uses MLP and FM to combine the interaction results. Particularly, it uses MLP to describe high-order interactions whereas FM is used to model interactions between second-order features.
– Field-aware Neural Factorization Machine (FNFM) (Zhang et al., 2019) uses the Bi-Interaction Concatenation layer for modeling second-order interactions whereas DNNs are used to learn the higher-order feature combination.
– Fi-GNN (Li et al., 2019) models attribute interactions using multi-head self-attention and models data as feature graphs.
– Graph Matching based Collaborative Filtering (GMCF) (Su, Zhang, Erfani, Gan, 2021) models and aggregates two kinds of attribute interactions with a graph matching structure to make recommendations.
– Hypergraph Infomax Recommender System (HIRS) (Su et al., 2022) uses hypergraph to generate advantageous feature interactions of any order for recommendation.

5.2. Overall performance

On three benchmark datasets, we compared CHCLA with the baselines. The results are displayed in Tables 2 and 3. The results of the best baseline are underlined and the best results are in bold. The Improv shows the CHCLA improvement along with the best baseline.

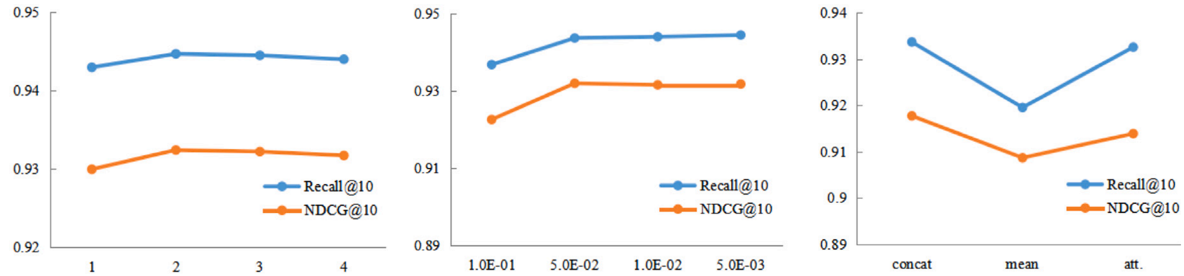
The following insightful observations are taken from Tables 2 and 3:

– CHCLA consistently performs better than all baselines on the three datasets. This is because CHCLA extracts user and item representations at two levels, namely user–item interaction and user/item attributes interaction. These results also demonstrate the ability of CHCLA to effectively use hypergraphs to model high-order attribute information, thus making accurate predictions.



(a) MovieLens 1M (b) Book-crossing (c) Taobao

Fig. 3. Ablation study on key parts of CHCLA.



(a) Number of layers (b) Contrastive loss weight (c) Fusion Method

Fig. 4. Hyperparameter study of CHCLA on Book-crossing.

- HIRS, which also considers hypergraph to model attribute information, outperforms GMCF on two datasets. This demonstrates that hypergraph structures are useful for modeling high-order interactions. The CHCLA distinguishes the attributes of users and items, and uses hypergraph to capture high-order attribute interactions. In addition, cross-view contrastive learning preserves both user–item interaction and attribute interaction from two levels. Hence, CHCLA captures more comprehensive information than the other methods.
- Graph-based models outperform traditional CF models on most datasets, suggesting the effectiveness of modeling pairwise relations using a graph architecture for recommendation. However, hypergraph-based approaches outperform graph-based methods, demonstrating that flexibility and effectiveness of modeling high-order interactions in a hypergraph architecture.

5.3. Ablation studies

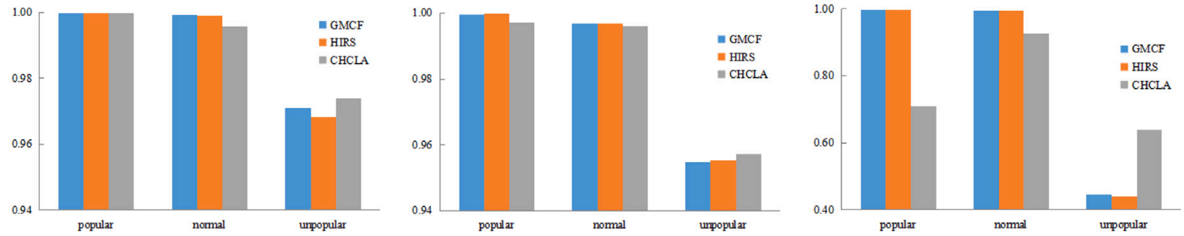
CHCLA comprises three primary parts: local collaborative interaction learning, global hypergraph representation learning, and cross-view contrastive learning. Here, we investigate how the key components of our model contributed to the performance.

- CHCLA$_{w/oG}$: This variation eliminates the global hypergraph representation learning and only use user–item interaction data for predictions.
- CHCLA$_{w/oL}$: This variation eliminates the local collaborative interactions learning and simply use user and item attribute data for predictions.
- CHCLA$_{w/oC}$: This variation eliminates the cross-view contrastive learning. To make predictions, all that is needed to do is to combine the user and item representations learned at the local and global stages, respectively.

Fig. 3 shows the results for the three variations and CHCLA. Compared with the competitive model variants, CHCLA exhibited the best performance. This demonstrates the significance of using hypergraph structures to model higher-order attribute interactions and using self-supervised contrastive learning to facilitate learning at the local and global levels. CHCLA$_{w/oC}$ is worse than CHCLA, indicating that cross-view contrastive learning can strengthen embedding space of user and item through self-discrimination from local and global views. CHCLA$_{w/oG}$ is the least competitive model, which reflects the superiority of learning attribute information. CHCLA$_{w/oL}$ is superior to CHCLA$_{w/oG}$, which displays hypergraph naturally can model higher-order interactions and hypergraph convolutional networks enable the high-order interactions among nodes by leveraging the node-hyperedge-node transform.

5.4. Hyperparameter study

We evaluated the effects of the model's hyperparameters, including layer size K, contrastive loss weight β and fusion method $\psi(\cdot)$. The evaluation results are shown in Fig. 4.



(a) MovieLens 1M (b) Book-crossing (c) Taobao

Fig. 5. Performance of Recall@20 over different item groups.

We vary layer size K range of $\{1,2,3,4\}$, contrastive loss weight β range of $\{0.1,0.01,0.001,0.0001\}$ and display the performance changing curves on book-crossing dataset in Fig. 4(a)(b). $K = 2$ and $\beta = 0.005$ offers the best model performance. The model's performance does not significantly degrade as the layers increases, further indicating that the hypergraph structure is effective in reducing the over-smoothing effect in GNN-based models. Specifically, the representations of users and items were refined using both local and global learning. β controls the contrastive loss weight in joint learning. Reducing the contrastive loss to a level approximately equivalent to the recommendation loss can improve the model's performance (Zou et al., 2022).

The user/item representations from local learning and global learning were combined using fusion methods. Except for concatenation, we further applied the element-wise mean, attention, etc. to fuse representations. Fig. 4(c) shows that concatenation provides the best performance. The local and global stages learn user and item representations from various levels, respectively. The concatenation operation can maintain features at different levels to the greatest extent, whereas overly complex feature fusion methods not only make the model more complex but also worsen its performance.

5.5. Ability to debias

To verify whether the CHCLA can alleviate popularity bias, we divided the item set into three subgroups according to item popularity (Yu et al., 2022). 80% of items with the fewest interactions are labeled 'Unpopular', 5% of the items with the highest interactions are called 'Popular', and the remaining items are called 'Normal'. We then performed experiments to determine the Recall@20 contributed by each group. The results are shown in Fig. 5.

It is clear that all improvements in the CHCLA originate from unpopular items. Its significant advantage on unpopular items (also known as long-tail items) recommendation largely compensates for the disadvantage on popular items. The GMCF and HIRS tend to recommend popular items and obtain the best recall values. It might not be a good idea to recommend popular items because users have presumably already encountered them. In this regard, CHCLA performs noticeably better than GMCF and HIRS, and its outstanding ability to find long-tail products perfectly satisfies user requirements.

5.6. Robustness in alleviating data sparsity

In many real recommendation scenarios, data (e.g., ratings or clicks) sent to recommender systems are sparse. The performance of recommender systems is known to decrease when such sparse data are fed into them (Feng, Liang, Song, & Wang, 2020). We sought to determine whether local and global learning using contrastive learning could yield benefits for highly sparse data. Sparsity level is altered by eliminating ratings at random (Feng et al., 2020; Ovaisi et al., 2022).

We removed 10%, 30%, 50%, 70%, and 90% of rating interactions uniformly and randomly. Original and sparsified training sets were used to train all models, and the same test sets were used for evaluation (Ovaisi et al., 2022). Figs. 6 and 7 show the percentage changes in AUC and NDCG@20 on MovieLens 1M and Book-crossing, respectively. This phenomenon matches our expectation that the AUC and NDCG@20 both decrease as the sparsity increases. After deleting 10%, 30%, 50%, 70%, and 90% of the ratings, the proposed approach significantly outperformed the second-ranked method of NDCG@20 on MovieLens 1M by 0.5%, 0.4%, 0.4%, 0.3% and 1.0%, respectively. The results demonstrated that for each sparsity level, CHCLA outperforms the other approaches in terms of the AUC and NDCG@20 values. By removing 90% of the ratings, the CHCLA clearly outperformed the alternative methods on both datasets, demonstrating that our method can alleviate the sparsity problem.

5.7. Visualization

To qualitatively evaluate the embeddings, we utilized t-SNE (Van der Maaten & Hinton, 2008) to display user and item embeddings. Fig. 8 illustrates the 2D visualization of the embeddings on MovieLens 1M with CHCLA$_{w/oG}$, CHCLA$_{w/oL}$ and CHCLA, where blue nodes indicate users, whereas orange nodes represent items. CHCLA$_{w/oG}$ produces embeddings in the same space for both users and items, making it visually difficult to distinguish between the node embeddings. CHCLA$_{w/oL}$ outperforms CHCLA$_{w/oG}$ in terms of separability because CHCLA$_{w/oL}$ integrates the attributes information of user and item. Visually, CHCLA provides the clearest separation between the two types of nodes.

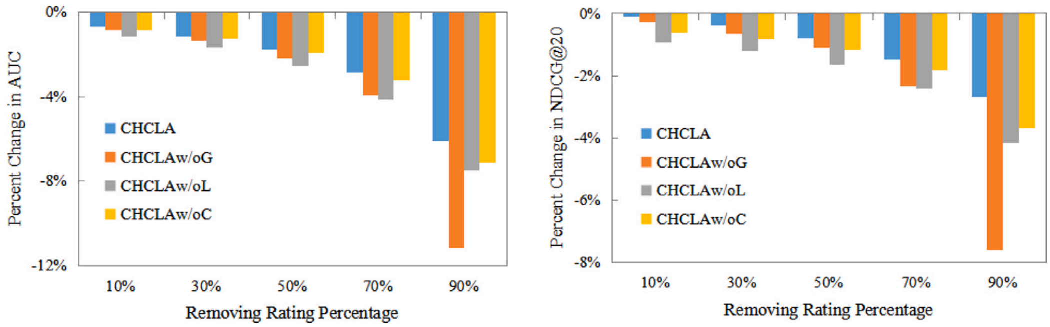


Fig. 6. AUC and NDCG@20 in various levels of sparsity on MovieLens 1M.

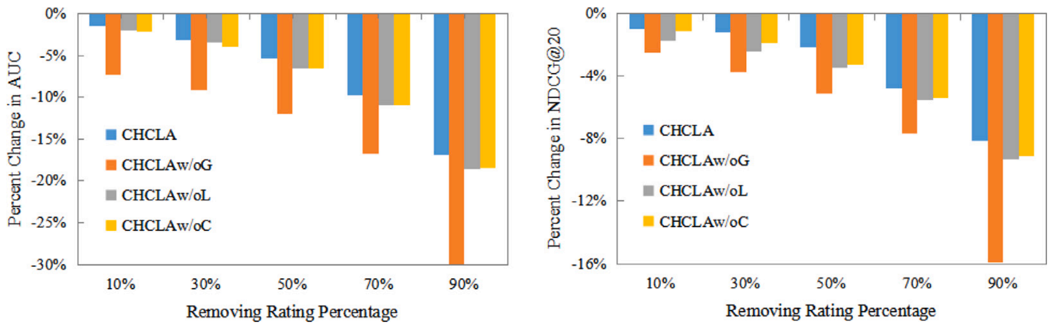
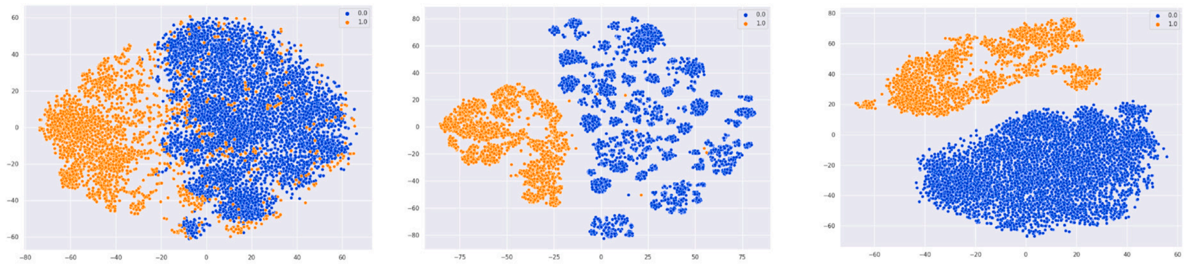


Fig. 7. AUC and NDCG@20 in various levels of sparsity on Book-crossing.



(a) $\text{CHCLA}_{w/oG}$ (b) $\text{CHCLA}_{w/oL}$ (c) CHCLA

Fig. 8. Visualization of node embeddings on MovieLens 1M (blue: users, orange: items). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

6. Conclusion

In this study, we present the CHCLA model, which introduces cross-view contrastive learning and hypergraph structures to handle the challenges of cold start and sparse supervised signals in recommendation and enhance the ability to model flexible and sufficient high-order interactions. Focusing on modeling user–item interaction and attribute interaction information, we designed a graph neural network to capture the user–item pairwise interaction and hypergraph neural networks to learn high-order attribute interaction. Furthermore, we devised cross-view contrastive learning with self-augmentation to enhance the robustness while solving the sparsely supervised signal issue. Experiments on three benchmark datasets illustrated that CHCLA performs better than the alternatives in terms of recommendations. Following the ablation study, we discovered that both the user–item interaction information and attributes interaction information have distinct impacts on recommendation results, thus warranting a meaningful discussion on their influence in user interest modeling. In addition, CHCLA also has some practical significance. For example, it can accurately and quickly analyze the items that users may buy through semantic information such as labels and attributes of items and users, combined with user history, to improve user satisfaction and shopping experience. CHCLA can be effectively applied to address cold start problems encountered in recommendation systems.

CRediT authorship contribution statement

Ang Ma: Investigation, Methodology, Writing – original draft, Writing – review & editing. **Yanhua Yu:** Funding acquisition, Supervision, Writing – review & editing. **Chuan Shi:** Supervision, Writing – review & editing. **Zirui Guo:** Investigation. **Tat-Seng Chua:** Funding acquisition, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This study is supported by the National Natural Science Foundation of China (grant ID U22B2019, grant ID U1936104); National Key Research and Development Program of China (grant ID 2020YFB2104503).

References

Berg, R. v. d., Kipf, T. N., & Welling, M. (2017). Graph convolutional matrix completion. arXiv preprint arXiv:1706.02263, URL: https://arxiv.org/abs/1706.02263.

Bu, J., Tan, S., Chen, C., Wang, C., Wu, H., Zhang, L., et al. (2010). Music recommendation by unified hypergraph: combining social media information and music content. In *Proceedings of the 18th ACM international conference on multimedia* (pp. 391–400). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/1873951.1874005.

Chang, Y., Zhou, W., Cai, H., Fan, W., Hu, L., & Wen, J. (2023). Meta-relation assisted knowledge-aware coupled graph neural network for recommendation. *Information Processing & Management*, *60*(3), Article 103353. http://dx.doi.org/10.1016/j.ipm.2023.103353, URL: https://www.sciencedirect.com/science/article/pii/S0306457323000900.

Duan, H., Zhu, Y., Liang, X., Zhu, Z., & Liu, P. (2023). Multi-feature fused collaborative attention network for sequential recommendation with semantic-enriched contrastive learning. *Information Processing & Management*, *60*(5), Article 103416. http://dx.doi.org/10.1016/j.ipm.2023.103416, URL: https://www.sciencedirect.com/science/article/pii/S030645732300153X.

Feng, C., Liang, J., Song, P., & Wang, Z. (2020). A fusion collaborative filtering method for sparse data in recommender systems. *Information Sciences*, *521*, 365–379. http://dx.doi.org/10.1016/j.ins.2020.02.052, URL: https://www.sciencedirect.com/science/article/pii/S0020025520301353.

Gao, Y., Zhang, Z., Lin, H., Zhao, X., Du, S., & Zou, C. (2020). Hypergraph learning: Methods and practices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(5), 2548–2566. http://dx.doi.org/10.1109/TPAMI.2020.3039374.

Ge, S., Wu, C., Wu, F., Qi, T., & Huang, Y. (2020). Graph enhanced representation learning for news recommendation. In *Proceedings of the web conference 2020* (pp. 2863–2869). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3366423.3380050.

Guo, H., Tang, R., Ye, Y., Li, Z., & He, X. (2017). DeepFM: a factorization-machine based neural network for CTR prediction. In *Proceedings of the 26th international joint conference on artificial intelligence* (pp. 1725–1731). ijcai.org.

Harper, F. M., & Konstan, J. A. (2015). The movielens datasets: History and context. *Acm Transactions on Interactive Intelligent Systems*, *5*(4), 1–19. http://dx.doi.org/10.1145/2827872.

He, X., & Chua, T.-S. (2017). Neural factorization machines for sparse predictive analytics. In *Proceedings of the 40th international ACM SIGIR conference on research and development in information retrieval* (pp. 355–364). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3077136.3080777.

He, X., Deng, K., Wang, X., Li, Y., Zhang, Y., & Wang, M. (2020). Lightgcn: Simplifying and powering graph convolution network for recommendation. In *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval* (pp. 639–648). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3397271.3401063.

Ji, S., Feng, Y., Ji, R., Zhao, X., Tang, W., & Gao, Y. (2020). Dual channel hypergraph collaborative filtering. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 2020–2029). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3394486.3403253.

Juan, Y., Zhuang, Y., Chin, W.-S., & Lin, C.-J. (2016). Field-aware factorization machines for CTR prediction. In *Proceedings of the 10th ACM conference on recommender systems* (pp. 43–50). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/2959100.2959134.

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In *Proceedings of the 3rd international conference on learning representations*. ICLR, URL: https://arxiv.org/pdf/1412.6980.pdf.

Li, Z., Cui, Z., Wu, S., Zhang, X., & Wang, L. (2019). Fi-gnn: Modeling feature interactions via graph neural networks for ctr prediction. In *Proceedings of the 28th ACM international conference on information and knowledge management* (pp. 539–548). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3357384.3357951.

Li, S., Lei, W., Wu, Q., He, X., Jiang, P., & Chua, T.-S. (2021). Seamlessly unifying attributes and items: Conversational recommendation for cold-start users. *ACM Transactions on Information Systems (TOIS)*, *39*(4), 1–29. http://dx.doi.org/10.1145/3446427.

Luo, C., Headden, W., Avudaiappan, N., Jiang, H., Cao, T., Yin, Q., et al. (2022). Query attribute recommendation at Amazon Search. In *Proceedings of the 16th ACM conference on recommender systems* (pp. 506–508). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3523227.3547395.

Ni, X., Nie, J., Zuo, Z., Xie, H., Liang, X., Jiang, M., et al. (2023). DITN: User's indirect side-information involved domain-invariant feature transfer network for cross-domain recommendation. *Information Processing & Management*, *60*(6), Article 103494. http://dx.doi.org/10.1016/j.ipm.2023.103494, URL: https://www.sciencedirect.com/science/article/pii/S0306457323002315.

Ovaisi, Z., Heinecke, S., Li, J., Zhang, Y., Zheleva, E., & Xiong, C. (2022). Rgrecsys: A toolkit for robustness evaluation of recommender systems. In *Proceedings of the 15th ACM international conference on web search and data mining* (pp. 1597–1600). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3488560.3502192.

Rendle, S. (2010). Factorization machines. In *Proceedings of the 10th international conference on data mining* (pp. 995–1000). IEEE, http://dx.doi.org/10.1109/ICDM.2010.127.

Su, Y., Zhang, R., Erfani, S., & Xu, Z. (2021). Detecting beneficial feature interactions for recommender systems. In *Proceedings of the 35th AAAI conference on artificial intelligence* (pp. 4357–4365). AAAI Press.

Su, Y., Zhang, R., M. Erfani, S., & Gan, J. (2021). Neural graph matching based collaborative filtering. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval* (pp. 849–858). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3404835.3462833.

Su, Y., Zhao, Y., Erfani, S., Gan, J., & Zhang, R. (2022). Detecting arbitrary order beneficial feature interactions for recommender systems. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining* (pp. 1676–1686). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3534678.3539238.

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, *9*(86), 2579–2605, URL: https://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf.

Wang, X., He, X., Wang, M., Feng, F., & Chua, T.-S. (2019). Neural graph collaborative filtering. In *Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval* (pp. 165–174). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3331184.3331267.

Wang, X., Ma, W., Guo, L., Jiang, H., Liu, F., & Xu, C. (2022). Hgnn: Hyperedge-based graph neural network for mooc course recommendation. *Information Processing & Management*, *59*(3), Article 102938. http://dx.doi.org/10.1016/j.ipm.2022.102938, URL: https://www.sciencedirect.com/science/article/pii/S0306457322000607.

Wu, J., Wang, X., Feng, F., He, X., Chen, L., Lian, J., et al. (2021). Self-supervised graph learning for recommendation. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval* (pp. 726–735). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3404835.3462862.

Xia, L., Huang, C., Xu, Y., Zhao, J., Yin, D., & Huang, J. (2022). Hypergraph contrastive collaborative filtering. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval* (pp. 70–79). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3477495.3532058.

Xia, L., Huang, C., & Zhang, C. (2022). Self-supervised hypergraph transformer for recommender systems. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining* (pp. 2100–2109). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3534678.3539473.

Xia, X., Yin, H., Yu, J., Shao, Y., & Cui, L. (2021). Self-supervised graph co-training for session-based recommendation. In *Proceedings of the 30th ACM international conference on information & knowledge management* (pp. 2180–2190). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3459637.3482388.

Xia, X., Yin, H., Yu, J., Wang, Q., Cui, L., & Zhang, X. (2021). Self-supervised hypergraph convolutional networks for session-based recommendation. In *Proceedings of the 35th AAAI conference on artificial intelligence* (pp. 4503–4511). AAAI Press.

Xiao, J., Ye, H., He, X., Zhang, H., Wu, F., & Chua, T.-S. (2017). Attentional factorization machines: Learning the weight of feature interactions via attention networks. In *Proceedings of the 26th international joint conference on artificial intelligence* (pp. 3119–3125). ijcai.org.

You, Y., Chen, T., Sui, Y., Chen, T., Wang, Z., & Shen, Y. (2020). Graph contrastive learning with augmentations. In *Proceedings of the 33th international conference on neural information processing systems* (pp. 5812–5823). MIT Press.

Yu, J., Yin, H., Li, J., Wang, Q., Hung, N. Q. V., & Zhang, X. (2021). Self-supervised multi-channel hypergraph convolutional network for social recommendation. In *Proceedings of the web conference 2021* (pp. 413–424). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3442381.3449844.

Yu, J., Yin, H., Xia, X., Chen, T., Cui, L., & Nguyen, Q. V. H. (2022). Are graph augmentations necessary? simple graph contrastive learning for recommendation. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval* (pp. 1294–1303). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3477495.3531937.

Zhai, P., Yang, Y., & Zhang, C. (2023). Causality-based CTR prediction using graph neural networks. *Information Processing & Management*, *60*(1), Article 103137. http://dx.doi.org/10.1016/j.ipm.2022.103137, URL: https://www.sciencedirect.com/science/article/pii/S0306457322002382.

Zhang, W., Du, T., & Wang, J. (2016). Deep learning over multi-field categorical data: –A case study on user response prediction. In *Proceedings of the 38th European conference on IR research* (pp. 45–57). Springer.

Zhang, L., Shen, W., Huang, J., Li, S., & Pan, G. (2019). Field-aware neural factorization machine for click-through rate prediction. *IEEE Access*, *7*, 75032–75040. http://dx.doi.org/10.1109/ACCESS.2019.2921026.

Zhang, X., Xu, B., Yang, L., Li, C., Ma, F., Liu, H., et al. (2022). Price does matter! modeling price and interest preferences in session-based recommendation. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval* (pp. 1684–1693). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3477495.3532043.

Zhou, G., Zhu, X., Song, C., Fan, Y., Zhu, H., Ma, X., et al. (2018). Deep interest network for click-through rate prediction. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 1059–1068). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3219819.3219823.

Ziegler, C.-N., McNee, S. M., Konstan, J. A., & Lausen, G. (2005). Improving recommendation lists through topic diversification. In *Proceedings of the 14th international conference on world wide web* (pp. 22–32). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/1060745.1060754.

Zou, D., Wei, W., Mao, X.-L., Wang, Z., Qiu, M., Zhu, F., et al. (2022). Multi-level cross-view contrastive learning for knowledge-aware recommender system. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval* (pp. 1358–1368). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3477495.3532025.