A Multi-view Graph Contrastive Learning Framework for Cross-Domain Sequential Recommendation

Zitao Xu
College of Computer Science and Software Engineering, Shenzhen University
Shenzhen, China
xuzitao2018@email.szu.edu.cn

Weike Pan*
College of Computer Science and Software Engineering, Shenzhen University
Shenzhen, China
panweike@szu.edu.cn

Zhong Ming*
College of Computer Science and Software Engineering, Shenzhen University
Shenzhen, China
mingz@szu.edu.cn

ABSTRACT

Sequential recommendation methods play an irreplaceable role in recommender systems which can capture the users' dynamic preferences from the behavior sequences. Despite their success, these works usually suffer from the sparsity problem commonly existed in real applications. Cross-domain sequential recommendation aims to alleviate this problem by introducing relatively richer source-domain data. However, most existing methods capture the users' preferences independently of each domain, which may neglect the item transition patterns across sequences from different domains, i.e., a user's interaction in one domain may influence his/her next interaction in other domains. Moreover, the data sparsity problem still exists since some items in the target and source domains are interacted with only a limited number of times. To address these issues, in this paper we propose a generic framework named multi-view graph contrastive learning (MGCL). Specifically, we adopt the contrastive mechanism in an intra-domain item representation view and an inter-domain user preference view. The former is to jointly learn the dynamic sequential information in the user sequence graph and the static collaborative information in the cross-domain global graph, while the latter is to capture the complementary information of the user's preferences from different domains. Extensive empirical studies on three real-world datasets demonstrate that our MGCL significantly outperforms the state-of-the-art methods.

CCS CONCEPTS

• **Information systems** → **Recommender systems**.

KEYWORDS

Cross-domain recommendation, Sequential recommendation, Graph neural networks, Contrastive learning

*co-corresponding authors

RecSys '23, September 18–22, 2023, Singapore, Singapore

ACM ISBN 979-8-4007-0241-9/23/09…$15.00
https://doi.org/10.1145/3604915.3608785

ACM Reference Format:
Zitao Xu, Weike Pan, and Zhong Ming. 2023. A Multi-view Graph Contrastive Learning Framework for Cross-Domain Sequential Recommendation. In *Seventeenth ACM Conference on Recommender Systems (RecSys '23), September 18–22, 2023, Singapore, Singapore.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3604915.3608785

1 INTRODUCTION

Sequential recommendation aims to predict the next item that a user is most likely to interact with based on his/her historical behavior sequences, by capturing the user's dynamic preferences and extracting the item transition patterns across sequences. Existing sequential recommendation methods often employ recurrent neural networks (RNNs) and attention mechanisms to model users' sequential information [10, 11, 13, 28], which have achieved an impressive progress. However, most of the works only focus on users' behavior sequences in a single domain, and suffers from the common cold-start and data sparsity problems in recommender systems.

Cross-domain recommendation has been proposed to alleviate the data sparsity problem in the target domain by introducing some relatively richer source-domain data [40]. Some recent works have been devoted to knowledge transfer between a source domain and a target domain [12, 22, 36]. However, existing methods do not take into account the sequential information of users' behaviors in different domains, and thus may not model the users' dynamic preferences or the sequential dependencies across behavior sequences well.

To address the above issues, in this paper we study a new and emerging problem, i.e., cross-domain sequential recommendation (CDSR). Scenarios of CDSR are commonly existed in real-world applications. Fig. 1 illustrates the interaction sequences of a user in a book domain and a movie domain. From the top half of Fig. 1, we can see that the user has recently read several "Sherlock Holmes" novels, and according to the idea of sequential recommendation, we will recommend the next "Sherlock Holmes" novel for him/her. From the bottom half of Fig. 1, we find that the user has recently tended to watch suspense movies, but there is no strong connection between these movies, so we can only randomly recommend a suspense movie for him/her. However, if we consider the user's interaction sequences in both domains, we will find that the user has recently watched the "Sherlock Holmes" movie in the movie domain and has also read several "Sherlock Holmes" novels in the book domain, therefore it would be a better choice to recommend the next "Sherlock Holmes" movie to him.

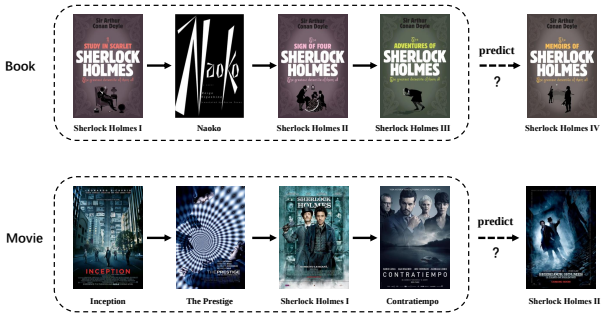


Figure 1: An example of a user's behavior sequences in a book domain (the top half) and a movie domain (the bottom half).

π-Net [21] is one of the earliest works for the CDSR task. It adopts RNNs to capture the sequential information and devises a cross-domain transfer unit to extract and share the user information between two domains at each timestamp. CD-SASRec [1] is an improved version of SASRec [13] which employs attention mechanisms in the CDSR task. DA-GCN [6] uses graph neural networks (GNNs) to model the complicated interaction relationships, which constructs a domain-aware graph to link different domains. RecGURU [15] unifies user embeddings from different domains via an adversarial learning approach and generates a generalized user representation.

Despite the progress of these studies, most existing CDSR methods focus on learning the user preferences of different domains separately and then performing knowledge transfer between these domains, while capturing preferences is usually independent of each domain. This neglects the item transition patterns across sequences from different domains. In real-world applications, the next item that a user may interact with in the target domain is likely to be related to the item he/she recently interacted with in a related source domain, and if we can further explore the joint sequential information across sequences from different domains, the recommendation performance will be improved.

Moreover, the data sparsity problem still exists in the CDSR task since some items in the target and source domains are interacted with only a limited number of times. Self-supervised learning derives the supervised signals from the data itself, which demonstrates its ability to address the data sparsity problem [19]. Contrastive learning is one of the important techniques in self-supervised learning, which can learn discriminative embeddings without explicit extra labels by maximizing the mutual information between positive pairs [32]. Contrastive learning has received increasing attention in sequential recommendation [20, 33, 35, 37]. However, most existing methods only consider a single sequence of a user's own, and cannot adequately extract the self-supervised signals when the interaction data is not sufficient.

In the CDSR task, there are item transition patterns across sequences, and the skewness of the length distribution has to be considered as well. Moreover, there are correlations of a user's preferences in the target and source domains, and the item characteristics of different domains may be not consistent, so how to perform knowledge transfer between domains is another issue to be concerned. It is thus challenging to design a generic contrastive learning framework for CDSR task.

To address the above issues, we treat the studied problem from the perspective of the intra-domain item representation view and the inter-domain user preference view. Specifically, we construct a target-domain sequence graph and a source-domain sequence graph for each user based on his/her historical sequences, with the purpose of capturing the sequential information in each domain. Moreover, we construct a cross-domain global graph by aggregating all users' behaviors in both the target and source domains, which aims to capture the static collaborative information across users' sequences in multiple domains. From the intra-domain item representation view, we apply the contrastive mechanism to the corresponding item embedding of user behavior sequences in the user sequence graph and the cross-domain global graph, so that the graph encoders can learn the complementary information, and extract the self-supervised signals to alleviate the data sparsity problem. Then, we aggregate the item embeddings learned from the two graphs and adopt the sequence encoder to capture the user's sequential preferences in both domains. From the inter-domain user preference view, considering that there are also transition patterns across sequences from different domains, and the user preferences in different domains may be similar in a period of time. Therefore, we apply the contrastive mechanism to the target-domain sequential preferences and the source-domain sequential preferences, so that the knowledge can be transferred across domains, enabling a user's preferences from different domains to complement each other and exact the self-supervised signals.

We summarize our main contributions as follows:

- We propose a generic contrastive learning framework named multi-view graph contrastive learning (MGCL) for cross-domain sequential recommendation, which tackles the problem from the perspective of an intra-domain item representation view and an inter-domain user preference view.
- We construct some user sequence graphs and a cross-domain global graph to learn complicated item representations, and adopt the contrastive mechanism to capture dynamic sequential information, static collaborative information and transition patterns across different domains.
- We conduct extensive empirical studies on three real-world datasets, where the results show that our MGCL significantly outperforms the state-of-the-art baselines. We also conduct ablation studies to demonstrate the effectiveness of each key component.

2 RELATED WORK

In this section, we briefly describe the related works from the following categories: (i) cross-domain general recommendation, (ii) sequential recommendation, and (iii) cross-domain sequential recommendation.

2.1 Cross-Domain General Recommendation

Cross-domain recommendation aims to alleviate the data sparsity issue in a target domain by transferring knowledge from a source domain. The most important concern of cross-domain recommendation is determining what knowledge to transfer between domains and how to transfer the knowledge. EMCDR [22] is a classic mapping-based method which transfers the overlapped users'

preferences between different domains by learning a mapping function. DDTCDR [16] introduces a latent orthogonal mapping to capture user preferences over multiple domains while preserving relations between users across different latent spaces. CMF [27] is another classic approach based on multi-domain collaborative training, which factorizes matrices from multiple domains and share the users' latent factors. CoNet [12] develops a collaborative cross-network to allow dual knowledge transfer across different domains. CCDR [36] performs contrastive learning on user and item attributes between different domains to enable more diverse knowledge transfer. However, these methods are not suitable for sequential recommendation since they all ignore the order in users' behaviors.

2.2 Sequential Recommendation

Sequential recommendation is proposed to capture the sequential patterns among users' historical interactions. RNNs are widely adopted in single-domain sequential recommendation [10, 11] due to their natural instincts to model a sequential data step-by-step. Caser [30] proposes a CNN-based method which adopts both horizontal and vertical convolutional filters to learn the sequential patterns. Another representative technique in sequential recommendation is the attention mechanism [9, 13, 28]. For example, SASRec [13] uses a stacked structure of self-attention blocks to model users' behavior sequences, which can capture the long-range dependencies across sequences. Recently, GNNs have attracted much attention [34, 38] for their ability to capture higher-order relationships among items. SRGNN [34] employs gated GNN in session graphs to capture complex item transitions. There are also some works that adopt contrastive mechanisms to alleviate the data sparsity issue [20, 33, 35, 37]. CL4SRec [37] and CoSeRec [20] propose some data augmentation approaches to construct contrastive tasks. MCLSR [33] employs interest-level and feature-level contrastive mechanisms to learn the co-action information between users and items. Although these studies have made great progress, none of them has considered knowledge transfer under cross-domain situations.

2.3 Cross-Domain Sequential Recommendation

π-Net [21] is one of the earliest works for CDSR in a shared-account scenario. Specifically, it employs GRUs to capture the sequential information in each domain, and transfer knowledge between different domains by a cross-domain transfer unit. PSJNet [29] is an improved method based on π-Net, which proposes a split-and-join framework to learn the cross-domain representations of users. CDSASRec [1] extends SASRec [13] to the cross-domain setting, which fuses the source-domain aggregated vector into the target-domain item embedding. DA-GCN [6] is a GNN-based model in CDSR, which constructs a domain-aware graph to model the multiple associations among items from different domains. There are also some hybrid models that combine different techniques to enhance the capability in capturing the item dependencies among sequences and the complex associations between domains. RecGURU [15] proposes an adversarial learning method to unify user representations from different domains into a generalized user representation. SEMI [14] employs contrastive learning to pre-train encoders for modeling users' behavior sequences from a multimodality view. DDGHM [39] proposes a dual dynamic graphical model with hybrid metric training to exploit the evolving patterns of users' behaviors and enhance representation learning to address the sparsity problem. C^2DSR [3] adopts a graphical and attentional encoder to capture the item relationships, and devises two sequential objectives with a contrastive objective to jointly learn the single-domain and cross-domain user representations.

Although these methods have made great progress, there is still a data sparsity problem because some items in the target and source domains are interacted with only a few times. Besides, most existing methods ignore the correlations between sequences from different domains. To address these problems, we construct a cross-domain global graph based on the interactions of all users in both domains, which aims to capture the static collaborative information among all sequences. Moreover, we apply the contrastive mechanism to the corresponding item embeddings and learn the self-supervised signals to further alleviate the sparsity problem. We also apply the contrastive mechanism to users' preferences from different domains, which extracts the complementary information and enables knowledge transfer across domains.

3 PROPOSED METHOD

In this section, we formally define the CDSR task and introduce the components of our proposed MGCL in detail.

3.1 Problem Definition

For cross-domain sequential recommendation, we have a set of users $\mathcal{U}$, and denote the set of items in the target domain as $\mathcal{I}^X$. Moreover, we have a source domain with same users and a different item set $\mathcal{I}^Y$. We define the target-domain behavior sequence of each user $u \in \mathcal{U}$ as $\mathcal{X} = \{x_1, x_2, \ldots, x_L\}$ (ordered by the interaction time), which consists of L items from $\mathcal{I}^X$. If the sequence length is shorter than L, a padding item will be repeatedly appended at the beginning of the sequence. Moreover, $\mathcal{X}_t = \{x_1, x_2, \ldots, x_t\}$, $1 \leqslant t \leqslant L$ denotes a truncated behavior sequence at time step t with regard to sequence $\mathcal{X}$. Similarly, we denote a truncated item sequence $\mathcal{Y}_{t'} = \{y_1, y_2, \ldots, y_{t'}\}$ for the source domain, where t' is the most recent time step at which the same user interacted with an item in the source domain before the real moment corresponding to the time step t in the target domain. This is to ensure causality of the user behaviors from the source domain to the target domain. The goal of CDSR is to predict the next possible preferred item in the target domain (i.e., x_{t+1}) according to $\mathcal{X}_t$ and $\mathcal{Y}_{t'}$. In the lower left part of Fig. 2 are the hybrid input sequences of the target and source domains ordered by the interaction time.

3.2 An Overview of MGCL

The overall framework of our proposed MGCL is illustrated in Fig. 2. We treat the studied problem from the perspective of an intra-domain item representation view and an inter-domain user preference view.

From the intra-domain item representation view (the top part of Fig. 2), we first construct a target-domain sequence graph and a source-domain sequence graph for each user according to his/her behavior sequences. Moreover, we construct a cross-domain global

graph by aggregating all users' interactions in both domains. Then we adopt the corresponding graph encoder on each graph to achieve message propagation and generate the item embeddings. After that, we apply a nonlinear projection on the output of each graph encoder layer. To make the sequential information and the collaborative information complement each other, we further adopt intra-domain graph contrastive learning on the projected item representations.

From the inter-domain user preference view (the bottom part of Fig. 2), we aggregate the item representations and adopt a sequence encoder to capture the user's sequential preferences in both domains. Then we employ a mapping unit to transfer the source-domain knowledge to the target domain. To learn the complementary information of the user's preferences from different domains, we apply inter-domain preference contrastive learning on the target-domain sequential preference and the mapped source-domain sequential preference. Finally, we concatenate the preferences and predict the next possible preferred item for the user.

3.3 Intra-Domain Item Representation View

Graph neural networks (GNNs) are effective in capturing structured information and higher-order item transitions. However, most existing sequential recommendation methods only focus on a single sequence of each user in a single domain. Such approaches ignore the correlations between sequences from different domains. To tackle this problem, we construct a cross-domain global graph based on the interactions of all users in both domains, which aims to capture the static collaborative information among all sequences. Moreover, we apply some contrastive mechanism to the corresponding item embeddings and learn the self-supervised signals to alleviate the data sparsity problem.

3.3.1 User sequence graph embedding learning. In this part, we separate the target-domain and source-domain sequences into graph-structured data and employ a user sequence graph encoder to generate the item embeddings and capture complex item transitions.

Target-domain sequence graph. Taking the target domain as an example, a user's sequence $\mathcal{X}_t$ in the target domain can be modeled as a directed graph. In the sequence graph, each node represents an item $x_i \in \mathcal{X}_t$ and each edge (x_{i-1}, x_i) denotes that a user clicks x_i after x_{i-1} in the sequence $\mathcal{X}_t$. We denote $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_t\}$ as the embeddings of the target-domain sequence $\mathcal{X}_t$, where $\boldsymbol{x}_i \in \mathbb{R}^d$ is a d-dimensional learnable vector.

User sequence graph encoder layer. Then, we present how to achieve message propagation and update the item embeddings in target-domain sequence graph. For graphs with different structures, we should adopt suitable graph encoders in order to fully extract the connections between the nodes and capture the structural information. Inspired by SRGNN [34], we employ gated GNNs as the user sequence graph encoder, which uses GRUs to model the sequential data step-by-step and control the flow of information between nodes, enabling the model to selectively aggregate and update node embedding. After encoding, the item embeddings in the target-domain sequence graph can be denoted as $\boldsymbol{H}^{S,X} = \{\boldsymbol{h}_1^{S,X}, \boldsymbol{h}_2^{S,X}, \ldots, \boldsymbol{h}_t^{S,X}\}$.

Similarly, we can construct a source-domain sequence graph, and obtain the source-domain item embeddings in the sequence graph $\boldsymbol{H}^{S,Y} = \{\boldsymbol{h}_1^{S,Y}, \boldsymbol{h}_2^{S,Y}, \ldots, \boldsymbol{h}_{t'}^{S,Y}\}$ after a graph encoder layer.

3.3.2 Global graph embedding learning. In this part, we aggregate all user sequences in a cross-domain global graph, and employ a global graph encoder to update the item embeddings and capture the collaborative information.

Cross-domain global graph. The cross-domain global graph is a undirected bipartite graph where a user node represents the corresponding user u_i in the user set $\mathcal{U}$. Based on the users' interaction sequences in the target and source domains, each user node is linked to the item nodes that he/she has interacted with. In such graph construction, different user sequences can be associated through the item nodes, and the item nodes from different domains can also be associated through the user nodes, so that the knowledge can be transferred across different sequences and domains

Global graph encoder layer. To fully extract the collaborative information in the cross-domain global graph, we adopt graph convolutional networks (GCNs) as the global graph encoder. Similar to LightGCN [7], an efficient model in non-sequential recommendation, we aggregates the features of neighboring nodes and obtain new representation for each node. And the propagation mechanism of GCNs enables knowledge transfer between items across different domains, effectively leveraging the rich source-domain information to alleviate the data sparsity problem in the target domain. The message propagation strategy can be formalized as:

$$\boldsymbol{E}^{(k+1)} = (\boldsymbol{D}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{D}^{-\frac{1}{2}})\boldsymbol{E}^{(k)}, \tag{1}$$

where $\boldsymbol{A}$ denotes the adjacency matrix of the cross-domain global graph and $\boldsymbol{D}_{ii} = \sum_{j=0} \boldsymbol{A}_{ij}$ is the corresponding diagonal degree matrix. Notice that k is the depth of the graph encoder layers, and $\boldsymbol{E}^{(0)}$ represents the embedding matrix of all nodes.

Then for a given user u with his/her corresponding behavior sequences in the target domain and source domain (i.e., $\mathcal{X}_t$ and $\mathcal{Y}_{t'}$), we can obtain the corresponding item embeddings $\boldsymbol{H}^{G,X} = \{\boldsymbol{h}_1^{G,X}, \boldsymbol{h}_2^{G,X}, \ldots, \boldsymbol{h}_t^{G,X}\}$ and $\boldsymbol{H}^{G,Y} = \{\boldsymbol{h}_1^{G,Y}, \boldsymbol{h}_2^{G,Y}, \ldots, \boldsymbol{h}_{t'}^{G,Y}\}$ by selecting from the learned embedding matrix $\boldsymbol{E}^{(k)}$ according to the item indices.

3.3.3 Intra-domain graph contrastive learning. The user sequence graph contains the dynamic sequential information of one single user behavior sequence, while the cross-domain global graph contains the static collaborative information across sequences and domains. To jointly learn the complementary information from different graphs, we apply the contrastive mechanism to the corresponding item embeddings of the user behavior sequences in the user sequential graph and the cross-domain global graph.

Projection. We apply a nonlinear projection on the output of each graph encoder layer, which is shown to be effective in SimCLR [4]. The auxiliary projection module maps the item representations to the space where a contrastive loss is applied, which makes contrastive learning more flexible and powerful. Notice that for the later recommendation task we use the item representation before linear projection since the introduction of contrastive learning may lead to information loss.

Taking the target domain as an example, we feed each target-domain item representation $\boldsymbol{h}_i^{S,X} \in \boldsymbol{H}^{S,X}$ and $\boldsymbol{h}_i^{G,X} \in \boldsymbol{H}^{G,X}$ into a

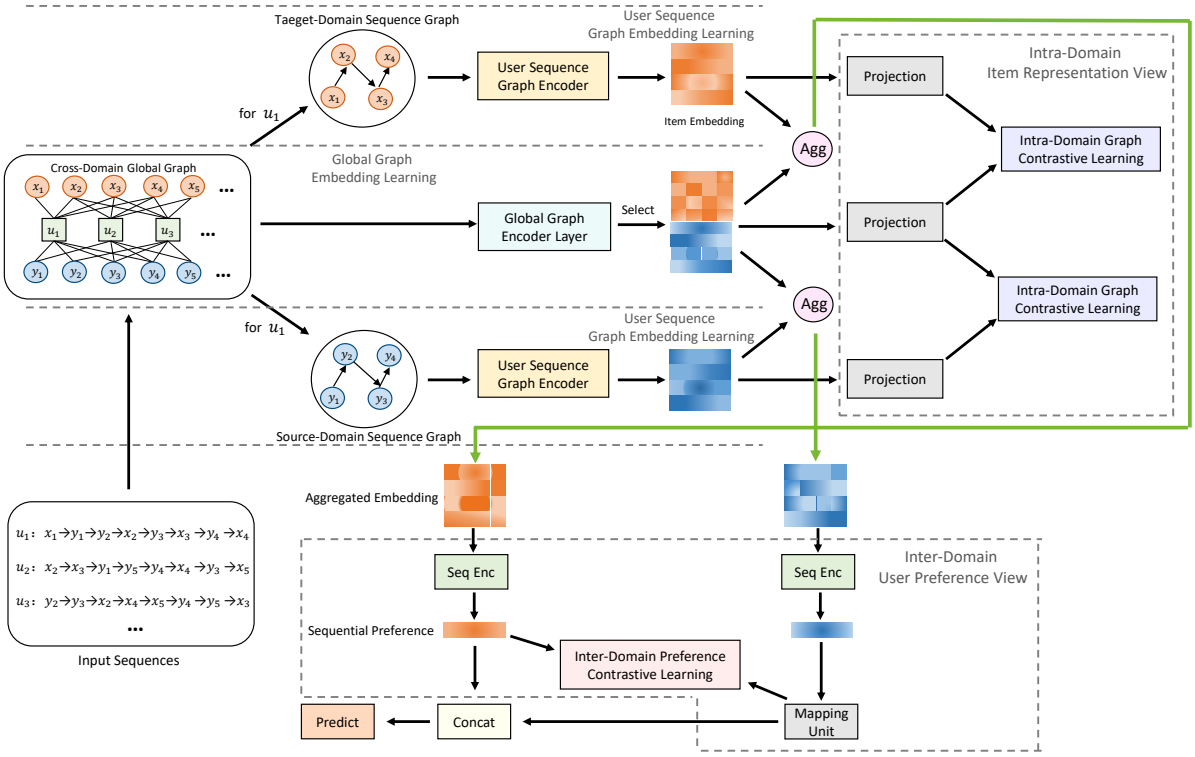


Figure 2: The framework of our proposed multi-view graph contrastive learning (MGCL).

multi-layer perceptron (MLP):

$$\boldsymbol{h}_i^{SP,X} = \boldsymbol{h}_i^{S,X}\boldsymbol{W}^{SP,X} + \boldsymbol{b}^{SP,X}, \tag{2}$$

$$\boldsymbol{h}_i^{GP,X} = \boldsymbol{h}_i^{G,X}\boldsymbol{W}^{GP,X} + \boldsymbol{b}^{GP,X}, \tag{3}$$

where $\boldsymbol{h}_i^{SP,X}, \boldsymbol{h}_i^{GP,X} \in \mathbb{R}^d$ denote the projected item representations, and $\boldsymbol{W}^{SP,X}, \boldsymbol{W}^{GP,X} \in \mathbb{R}^{d\times d}, \boldsymbol{b}^{SP,X}, \boldsymbol{b}^{GP,X} \in \mathbb{R}^d$ are learnable parameters. Likewise, we can get $\boldsymbol{h}_i^{SP,Y}$ and $\boldsymbol{h}_i^{GP,Y}$ in a similar way.

Item representation contrastive learning. Following [2, 33], we regard the graph embedding learned from the sequence graph and the global graph as different aspects of item representations, and adopt some contrastive mechanism to extract the self-supervised signal from different aspects. The learning objective of contrastive learning is to maximize the mutual information between some positive pairs while minimizing the agreement between some negative samples. Specifically, we treat the representations for a same item from different graphs (i.e., $\boldsymbol{h}_i^{SP,X}$ and $\boldsymbol{h}_i^{GP,X}$) as a pair of positive samples. Moreover, we naturally treat the item representations of different users (i.e., $\boldsymbol{h}_i^{SP,X}$ and $\tilde{\boldsymbol{h}}_i^{SP,X}$) as a pair of negative samples. Then, we adopt InfoNCE [24, 35] with a standard binary cross-entropy loss between samples as our learning objective. The loss function can be formalized as follows:

$$\mathcal{L}_t = -\sum_{i=1} \log\sigma(sim(\boldsymbol{h}_i^{SP,X}, \boldsymbol{h}_i^{GP,X})) + \log\sigma(1 - sim(\tilde{\boldsymbol{h}}_i^{SP,X}, \boldsymbol{h}_i^{GP,X})), \tag{4}$$

where $\sigma(\cdot)$ is the sigmoid function and $sim(\cdot)$ is a dot product to measure the similarity between two item representations.

Similarly, we can calculate the item representation contrastive loss in the source domain as follows:

$$\mathcal{L}_s = -\sum_{i=1} \log\sigma(sim(\boldsymbol{h}_i^{SP,Y}, \boldsymbol{h}_i^{GP,Y})) + \log\sigma(1 - sim(\tilde{\boldsymbol{h}}_i^{SP,Y}, \boldsymbol{h}_i^{GP,Y})). \tag{5}$$

3.4 Inter-Domain User Preference View

After obtaining the item representation, a typical method is to adopt a sequence encoder to capture a user's sequential preferences, i.e., to explore the user's current interest and transition patterns from his/her behavior sequence. However, most of the existing methods focus only on a user's sequence in a single domain, and the recommendation performance may be unsatisfactory when the data in the target domain is highly sparse.

Besides, a user's interaction in one domain may influence his/her next interaction in other domains, which means that there are also item transition patterns across sequences from different domains. Therefore, we aim to consider the user preferences in both the target and source domains, and achieve knowledge transfer across domains. Moreover, we apply the contrastive mechanism to the target-domain sequential preference and the source-domain sequential preference to learn the complementary information and derive some self-supervised signals.

3.4.1 Sequential preference learning. In this part, we first aggregate the graph embedding learned via a user sequence graph and a cross-domain global graph. Then we employ sequence encoders to capture the user sequential preferences in the target and source domains.

Graph embedding aggregation. As discussed in Section 3.3, the item embeddings in the user sequence graph contain the sequential information while those in the cross-domain global graph contain the collaborative information. To obtain more suitable item representations, we adaptively aggregate these two by a weighted sum. Specifically, we concatenate the representations of a same item from different graphs, feed it into an MLP and use an activation function to obtain the corresponding weights:

$$g_i^X = \sigma([\boldsymbol{h}_i^{S,X}, \boldsymbol{h}_i^{G,X}]\boldsymbol{W}_g^X + \boldsymbol{b}_g^X), \tag{6}$$

where $[\cdot]$ denotes the concatenation operation, $\boldsymbol{W}_g^X \in \mathbb{R}^{2d\times 1}$ and $\boldsymbol{b}_g^X \in \mathbb{R}$ are learnable parameters. We use the sigmoid function $\sigma(\cdot)$ as the activation function, so that g_i^X is restricted to $(0, 1)$.

The final representation of the item can be calculated by a weighted sum of $\boldsymbol{h}_i^{S,X}$ and $\boldsymbol{h}_i^{G,X}$ as follows:

$$\boldsymbol{h}_i^{A,X} = g_i^X \times \boldsymbol{h}_i^{S,X} + (1 - g_i^X) \times \boldsymbol{h}_i^{G,X}. \tag{7}$$

Similarly, we can obtain the aggregated item representations $\boldsymbol{H}^{\boldsymbol{A},\boldsymbol{Y}} = \{\boldsymbol{h}_1^{A,Y}, \boldsymbol{h}_2^{A,Y}, \ldots, \boldsymbol{h}_{t'}^{A,Y}\}$ from the source domain.

Sequence encoder. Next, we adopt a sequence encoder to model the aggregated item embeddings in order to capture the user's dynamic preferences and sequential dependencies among items. Sequence encoders have been extensively studied in various previous works in sequential recommendation with significant progress, and we do not focus on their design in our work. We utilize the architecture of SASRec [13] model in this paper, which applies a unidirectional Transformer encoder [31] and has been shown as an effective and efficient model in sequential recommendation.

The sequential preferences of the user at the t-th time step in the target domain $\boldsymbol{f}_t^X \in \mathbb{R}^d$ and in the source domain $\boldsymbol{f}_t^Y \in \mathbb{R}^d$ can be formalized as:

$$\boldsymbol{f}_t^X = SeqEnc(\{\boldsymbol{h}_1^{A,X}, \boldsymbol{h}_2^{A,X}, \ldots, \boldsymbol{h}_t^{A,X}\}), \tag{8}$$

$$\boldsymbol{f}_t^Y = SeqEnc(\{\boldsymbol{h}_1^{A,Y}, \boldsymbol{h}_2^{A,Y}, \ldots, \boldsymbol{h}_{t'}^{A,Y}\}). \tag{9}$$

3.4.2 Inter-domain preference contrastive learning. Considering that a same user's preferences in different domains may be similar in a period of time, e.g., in Fig. 1, the user prefers to read suspense novels in the book domain and watch suspense movies in the movie domain, so the items of the suspense category are more likely to attract his attention. Even though the types of items are different, they reflect the same user preference. Therefore, we apply the contrastive mechanism to the target-domain sequential preference and the source-domain sequential preference.

Mapping unit. Since the item characteristics of different domains may be not consistent, to achieve knowledge transfer from the source domain to the target domain, we employ a mapping unit to reconstruct the user's sequential preference in the source domain. Specifically, we feed the source-domain user's preference $\boldsymbol{f}_t^Y$ into an MLP:

$$\boldsymbol{f}_t^{M,Y} = \boldsymbol{f}_t^Y \boldsymbol{W}^{M,Y} + \boldsymbol{b}^{M,Y}, \tag{10}$$

where $\boldsymbol{W}^{M,Y} \in \mathbb{R}^{d\times d}$ and $\boldsymbol{b}^{M,Y} \in \mathbb{R}^d$ are learnable parameters.

Preference contrastive learning. To enable a user's preferences from different domains complement each other and learn the self-supervised signals, we maximize the mutual information between them. Specifically, we treat the sequential preferences of a same user in different domains (i.e., $\boldsymbol{f}_t^X$ and $\boldsymbol{f}_t^{M,Y}$) as a pair of positive samples while the sequential preferences of two different users (i.e., $\boldsymbol{f}_t^X$ and $\tilde{\boldsymbol{f}}_t^X$) as a pair of negative samples. Then, the preference contrastive loss function can be formalized as:

$$\mathcal{L}_p = -\log \sigma(sim(\boldsymbol{f}_t^X, \boldsymbol{f}_t^{M,Y})) + \log \sigma(1 - sim(\boldsymbol{f}_t^{\boldsymbol{X}}, \tilde{\boldsymbol{f}}_t^X)). \tag{11}$$

3.5 Prediction Layer

In the prediction layer, we employ concatenation to aggregate the sequential preferences from the target and source domains in order to adequately utilize the information of different domains.

$$\boldsymbol{f} = \left[\boldsymbol{f}_t^X, \boldsymbol{f}_t^{M,Y}\right], \tag{12}$$

where $\boldsymbol{f} \in \mathbb{R}^{2d}$ denotes the concatenation of the sequential preferences. Then, the concatenation vector is fed into an MLP to obtain the final representation of the user's preferences:

$$\boldsymbol{f}_t = \boldsymbol{f}\boldsymbol{W}^{(f)} + \boldsymbol{b}^{(f)}, \tag{13}$$

where $\boldsymbol{W}^{(f)} \in \mathbb{R}^{2d\times d}$ and $\boldsymbol{b}^{(f)} \in \mathbb{R}^d$ are learnable parameters, and $\boldsymbol{f}_t \in \mathbb{R}^d$ denotes the final representation of the user's preferences. Finally, the prediction score of item i can be calculated as follows:

$$r_{t,i} = \boldsymbol{f}_t(\boldsymbol{x}_i)^T. \tag{14}$$

We adopt the binary cross-entropy loss function as the recommendation loss for our MGCL:

$$\mathcal{L}_r = -\sum_{u\in\mathcal{U}}\sum_{t=1}^{L-1} \delta(x_{t+1})[\log(\sigma(r_{t,x_{t+1}})) + \log(1 - \sigma(r_{t,j}))], \tag{15}$$

where $j \in \mathcal{I}^X \backslash \mathcal{X}^u$ is a sampled negative item and $\sigma(\cdot)$ is the sigmoid function. The indicator function $\delta(x_{t+1}) = 1$ only if x_{t+1} is not a padding item, and 0 otherwise.

Finally, we combine the recommendation loss and the three contrastive losses as follows:

$$\mathcal{L} = \mathcal{L}_r + \alpha\mathcal{L}_t + \beta\mathcal{L}_s + \gamma\mathcal{L}_p, \tag{16}$$

where α, β, γ are the hyper-parameters to control the intensity of the self-supervised tasks.

4 EXPERIMENTS

In this section, we introduce the experimental settings and conduct extensive empirical studies to answer the following five research questions:

(RQ1) How does our MGCL perform compared with the state-of-the-art methods?
(RQ2) Does our MGCL alleviate the data sparsity issue?
(RQ3) What's the influence of various components in our MGCL?
(RQ4) How do the weights of different contrastive losses affect the performance of our MGCL?
(RQ5) Does our MGCL really learn better item representation with the help of contrastive learning compared with other state-of-the-art baselines?

Table 1: Statistical details of the datasets.

Dataset	# Overlapped-Users	# Items	# Interactions	Avg. Length	Density
Movie	10929	60902	462314	42.30	0.07%
CD		94171	348746	31.91	0.03%
Book		242363	615912	56.36	0.02%

4.1 Datasets

We follow [1, 39] and conduct experiments on Amazon[1], which is a review data collected by [23] from the eponymous e-commerce platform. The Amazon data contains overlapped users in multiple domains, which make it suitable for the study of CDSR compared with other datasets commonly used in the community of recommender systems. We choose three datasets with different categories, i.e., "Movie", "CD" and "Book" from the Amazon data. According to the setting in [13, 17], we preprocess the datasets as follows: 1) We suppose that the presence of review, check-in and purchase behaviors are positive feedback (i.e., a user interacted with an item) and use the timestamps to determine the order of the interactions. 2) We only keep the users and items with no fewer than five related interactions. And we discard later duplicated (user, item) pairs. 3) We only keep the sequence of a user who has interactions in all the three domains. 4) We adopt the leave-one-out evaluation by splitting each sequence into three parts, i.e., the last interaction for test, the penultimate interaction for validation and the remaining interactions for training. Table 1 shows the statistical details of the processed datasets.

4.2 Evaluation Metrics

To evaluate the recommendation performance of all models, we adopt two common ranking-based metrics, i.e., HR@10 (hit ratio) and NDCG@10 (normalized discounted cumulative gain), where the former corresponds to recall because there is exactly one preferred item for each user in the test data in our case. In particular, HR@10 refers to the proportion of the ground-truth items appearing in the top-10 recommended lists, while NDCG@10 is sensitive to the exact ranking positions of the items in the lists. Following the common strategy in [8, 13], we sample 100 negative items as candidates to avoid heavy computation on all the (user, item) pairs. These 100 negative items have not been interacted with by the corresponding users and are sampled according to their popularity to ensure that they are informative and representative [17].

4.3 Baselines

To verify the effectiveness of our MGCL, we compare it with thirteen competitive baselines, including one general recommendation method (i.e., BPRMF), one cross-domain recommendation method (i.e., CoNet), eight sequential recommendation methods (i.e., FPMC, GRU4Rec, GRU4Rec+, Caser, GCSAN, SASRec, CL4SRec and CoSeRec) and three cross-domain sequential recommendation methods (i.e., π-net, DA-GCN and CD-SASRec).

- BPRMF [25]. A traditional model which optimizes a matrix factorization task using a pairwise ranking loss.
- CoNet [12]. A neural transfer learning model which enables dual information transfer across domains by developing cross-connection units on MLPs.
- FPMC [26]. A classic method that combines matrix factorization and Markov chains to model the sequential pattern.
- GRU4Rec [11]. An RNN-based method which explores the item dependencies over the sequences by adopting GRUs.
- GRU4Rec+ [10]. An improved model based on GRU4Rec [11] which adopts a new loss function and an additional sampling strategy.
- Caser [30]. A CNN-based model which employs horizontal and vertical convolutional filters to model the sequences.
- GCSAN [38]. A GNN-based model which constructs directed graphs for the sequences and applies gated GNNs to obtain all node vectors involved in the session graphs.
- SASRec [13]. An attention based model that employs the attention mechanism to capture the dynamic preferences. It also works as the sequence encoder in our MGCL.
- CL4SRec [37]. A self-supervised model which adopts three different data augmentation approaches to construct contrastive learning tasks.
- CoSeRec [20]. An improved model based on CL4SRec [37] which introduces two new informative augmentation operators leveraging item correlations to generate contrastive views.
- SASRec-M. To investigate the case where the available data are the same as the cross-domain scenario, we merge the source and target-domain sequences according to the timestamp, and learn a monolithic model based on SASRec [13].
- π-Net [21]. An RNN-based model which devises a cross-domain transfer unit to extract and share the user information across different domains at each timestamp.
- DA-GCN [6]. A GNN-based model which employs graph convolution networks to learn the complicated interaction relationships and the structural information in a cross-domain sequence graph.
- CD-SASRec [1]. An improved method based on SASRec [13] which fuses the source-domain aggregated vector into the target-domain item embedding to transfer information across domains.

4.4 Implementation Details

We implement GRU4Rec[2], Caser[3], SASRec[4], CoSeRec[5] and π-net [6] following the released codes by the authors. For the general setting, the latent dimensionality d is selected from $\{10, 20, 30, 40, 50\}$ and is finally configured as $d = 50$ since we find that these methods usually benefit from a larger value of d on such sparse datasets [13, 30]. The mini-batch size is set to 128, the dropout rate is set to 0.5 and the maximum length of a sequence L is set to 100. For our MGCL, we adopt the Adam optimizer with a learning rate of 0.001. The weights of the contrastive losses $\{\alpha, \beta, \gamma\}$ are set to $\{0.5, 0.5, 0.5\}$. For GRU4Rec+, the negative sampling number is set to 2048. For Caser, the vertical and horizontal filter numbers are set to 4 and 16, respectively. For the methods with Transformer architectures (i.e., SASRec, CD-SASRec and our MGCL), we adopt single-head attention layers and two attention blocks. The depth of the GNN layer k is set to 2 for GCSAN, DA-GCN and our MGCL. For

[1] http://jmcauley.ucsd.edu/data/amazon/

[2] https://github.com/hidasib/GRU4Rec
[3] https://github.com/graytowne/caser_pytorch
[4] https://github.com/kang205/SASRec
[5] https://github.com/YChen1993/CoSeRec
[6] https://github.com/mamuyang/PINet

the shared-account recommendation methods (i.e., π-Net and DA-GCN), the latent user number is set to 1. Other key parameters are configured following the suggestions of the corresponding papers or are tuned on the validation data.

For cross-domain recommendation methods, we only report the performance of the best-performing model with the corresponding source domain (i.e., when the target domain is Movie, we use CD or Book as a source domain to assist in training, and only show the best results). All the models are trained using Tesla V100 PCIe GPU with 32 GB memory. The source codes of our MGCL and datasets are available at https://csse.szu.edu.cn/staff/panwk/publications/MGCL/.

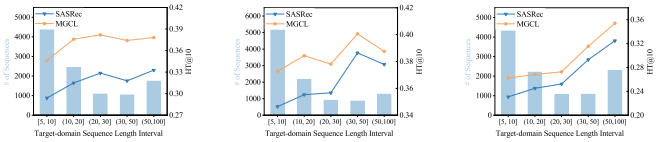
4.5 Overall Performance Comparison (RQ1)

Table 2 illustrates the experimental results of our MGCL and baselines on three datasets. We mark the best result in each column in bold and underline the second-best one.

Firstly, we can observe that our proposed MGCL outperforms all the baselines on all the three datasets, achieving an average improvement of 10.78% on NDCG@10 and 9.41% on HR@10 compared with the strongest baseline, which indicates the superiority of our MGCL in CDSR. Moreover, the sequential recommendation methods outperform the non-sequential recommendation methods, which demonstrates the importance of modeling the sequential information from users' behaviors. And the CDSR methods outperform most single-domain sequential recommendation methods, which indicates that the introduction of cross-domain information is often beneficial to improve the recommendation performance. However, SASRec-M doesn't achieve favorable performance because simply merging the two sequences may introduce more noise. Besides, the attention-based models achieve outstanding performance in both sequential recommendation and CDSR, which demonstrates the capability of the attention mechanism in capturing users' dynamic preferences. In contrast, the GNN-based methods do not achieve better performance than the attention-based methods, probably because the graph structure constructed on such sparse datasets is also sparse and the message propagation between nodes becomes ineffective. Furthermore, among the three datasets, the "Movie" dataset obtains the most significant improvement, which may be due to the fact that it is more tightly related to the other domains, i.e., a user's interaction sequences in the "Book" and "CD" domains are more likely to influence his/her next interaction in the "Movie" domain. Another reason is that the "Movie" dataset has the fewest items records (as is shown in Table 1), for which knowledge transfer from another domain is likely to be more helpful.

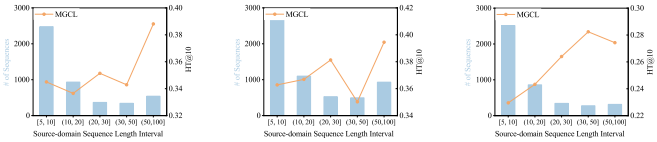
4.6 Performance Analysis w.r.t. Sparsity (RQ2)

In this subsection, we conduct two experiments to verify the effectiveness of introducing cross-domain information to alleviate the data sparsity problem. (1) We divide users into groups based on their behavior sequence length in the target domain, and identify the reasons of improvement by comparing the performance of SASRec and MGCL in different user groups. (2) We divide the users into groups based on their behavior sequence length in the source domain while fixing the target-domain sequence length interval,



(a) Movie (b) CD (c) Book

Figure 3: Performance of SASRec and our MGCL w.r.t. different sequence lengths in the target domain.



(a) Movie (b) CD (c) Book

Figure 4: Performance of SASRec and our MGCL w.r.t. different sequence lengths in the source domain.

and study how the source-domain sequence length affects the recommendation performance. Notice that due to space limitation, we only report the performance on HR@10, and the variation tendency on NDCG@10 is similar.

4.6.1 Performance w.r.t. Target-Domain Sequence Length. According to the users' sequence lengths in the target domain, we divide them into five user groups. Fig. 3 depicts the size of each user group and the corresponding HR@10 performance on the three datasets. It can be observed that the interaction data of most users is sparse in the target domain. The group with the shortest sequence length contains the most users in all the datasets, and the size decreases as the sequence length becomes longer.

Moreover, the maximum improvement of our MGCL against SASRec occurs in the short-sequence interval, ranging from 13.60% to 19.34% on all the datasets. The reason is that the shorter users' sequence lengths indicate that their interaction data is sparser, in which case the traditional single-domain method (i.e., SASRec) struggles to capture users' preferences. In contrast, the introduction of the rich source-domain data can enhance learning users' preferences, and the knowledge transfer across domains are more effective in this situation. It demonstrates that our MGCL is effective in alleviating the data sparsity problem. Furthermore, we also observe that our MGCL achieves better performance than SASRec on most user groups, which confirms the superiority of our MGCL in sequential recommendation.

4.6.2 Performance w.r.t. Source-Domain Sequence Length. To investigate the effect of the source-domain sequence length on performance, we select the interval with the shortest target-domain sequences (i.e., the sparsest data) and then divide the users into groups based on their sequence lengths in the source domain.

As is shown in Fig. 4, similar to the target domain, user group with the shortest source-domain sequences contains the most users in all the datasets, and the sizes decrease as the sequences become longer. Moreover, we can find that the recommendation performance in the target domain generally improves as the sequence length in the source domain increases. This is reasonable since our

Table 2: Recommendation performance of our MGCL and the baselines on three datasets.

Method	Movie		CD		Book	
	NDCG@10	HR@10	NDCG@10	HR@10	NDCG@10	HR@10
BPRMF	0.0597	0.1256	0.0492	0.1142	0.0465	0.1088
CoNet	0.0675	0.1489	0.0756	0.1484	0.0764	0.1819
FPMC	0.0723	0.1697	0.0819	0.1785	0.0695	0.1416
GRU4Rec	0.1017	0.1984	0.1210	0.2247	0.1066	0.2162
GRU4Rec+	0.1133	0.2157	0.1440	0.2536	0.1293	0.2407
Caser	0.1231	0.2243	0.1267	0.2473	0.1163	0.2274
GCSAN	0.1576	0.2889	0.1783	0.3206	0.1291	0.2409
SASRec	0.1740	0.3126	0.1965	0.3539	0.1402	0.2597
CL4SRec	0.1821	0.3179	0.1936	0.3350	0.1409	0.2556
CoSeRec	0.1842	0.3214	0.1968	0.3409	0.1428	0.2631
SASRec-M	0.0961	0.1970	0.0965	0.1921	0.0812	0.1666
π-Net	0.1113	0.2080	0.1265	0.2335	0.1042	0.2101
DA-GCN	0.1736	0.3124	0.1897	0.3458	0.1283	0.2375
CD-SASRec	0.1787	0.3159	0.1995	0.3610	0.1438	0.2677
MGCL	**0.2092**	**0.3693**	**0.2156**	**0.3797**	**0.1542**	**0.2842**

model can identify more users' preference characteristics in the source domain with more interaction data, which are then transferred to the target domain.

4.7 Ablation Study (RQ3)

We conduct an ablation study to evaluate the contribution of different components of our MGCL, and the results are presented in Table 3. Specifically, we separate out the important components of our MGCL and recompose them. We denotes 'SeqEnc' as the sequence encoder (i.e., SASRec). 'U' and 'G' represents user sequence graph embedding learning and global graph embedding learning, respectively. 'S' represents the introduction of the source-domain data, capturing the users' preferences in the source domain and aggregating them with the target-domain preferences for prediction. 'GCL' and 'PCL' denotes the intra-domain graph contrastive learning and the inter-domain preference contrastive learning, respectively. 'Projection' represents a nonlinear projection on the output of each graph encoder layer. From Table 3, we have the following observations:

- Simply attaching a graph encoder (denoted as 'U' or 'G') to a sequence encoder to learn the item representations cannot achieve better results, even though both graph encoders are employed. The reason is that on such sparse datasets, the graph structure constructed only by users' sequences from a single domain is still very sparse and the message propagation between nodes cannot be performed well.
- Aggregating the users' preferences from different domains (denoted as 'S') significantly improves the recommendation performance, which demonstrates that the introduction of the source-domain data and the knowledge transfer across domains enable the model to adequately capture the users' preferences and effectively alleviate the sparsity problem.
- Our model typically benefits from intra-domain graph contrastive learning (denoted as 'GCL'). It indicates that jointly learning static collaborative information and dynamic sequential information through the contrastive learning mechanism and extracting the self-supervised signal is helpful to improve the recommendation performance. Moreover, the performance is further improved by adding a nonlinear projection (denoted as 'Projection'), which reflects the importance of the projection layer in alleviating loss of information caused by contrastive learning.
- Inter-domain preference contrastive learning (denoted as 'PCL') can significantly improve the model performance. It demonstrates the importance of capturing the complementary information of users' preferences in the target and source domains through the contrastive learning mechanism and performing knowledge transfer between different domains. Moreover, we can observe that all the best results are from the composition of all the modules, which confirms their complementarity.

4.8 Influence of Hyper-parameters (RQ4)

In this subsection, we explore the influence of three hyper-parameters (i.e., the weight of target-domain item representation contrastive learning α, the weight of source-domain item representation contrastive learning β and the weight of user preference contrastive learning γ) on the model performance. We vary the value of the hyper-parameters in the range of $\{0, 0.25, 0.5, 0.75, 1\}$ and report the results in Fig. 5.

These three hyper-parameters control the intensity of the corresponding self-supervised tasks. From Fig. 5, we can observe that the recommendation performance improves when the parameter value is increased from 0 to some larger values on all datasets, which demonstrates the effectiveness of the three contrastive learning tasks. Moreover, the best results are achieved in most cases when the parameter value is 0.5, but the performance gradually decreases as the parameter value further increases to 1. This indicates that an excessive focus on the contrastive learning task may hurt the

Table 3: Recommendation performance in ablation studies of our MGCL with different architectures.

Architecture	Book → Movie		Book → CD		Movie → Book	
	NDCG@10	HR@10	NDCG@10	HR@10	NDCG@10	HR@10
SeqEnc	0.1740	0.3126	0.1965	0.3539	0.1402	0.2597
SeqEnc + U	0.1692	0.2977	0.1881	0.3414	0.1333	0.2441
SeqEnc + G	0.1755	0.3146	0.1941	0.3481	0.1348	0.2540
SeqEnc + U + G	0.1738	0.3095	0.2005	0.3545	0.1349	0.2552
SeqEnc + U + G + S	0.1926	0.3381	0.2034	0.3598	0.1473	0.2701
SeqEnc + U + G + S + GCL	0.1933	0.3427	0.1995	0.3581	0.1487	0.2727
SeqEnc + U + G + S + GCL + Projection	0.1998	0.3574	0.2070	0.3687	0.1496	0.2748
SeqEnc + U + G + S + PCL	0.2029	0.3588	0.2095	0.3704	0.1488	0.2743
SeqEnc + U + G + S + GCL + Projection + PCL	**0.2092**	**0.3693**	**0.2156**	**0.3797**	**0.1542**	**0.2842**

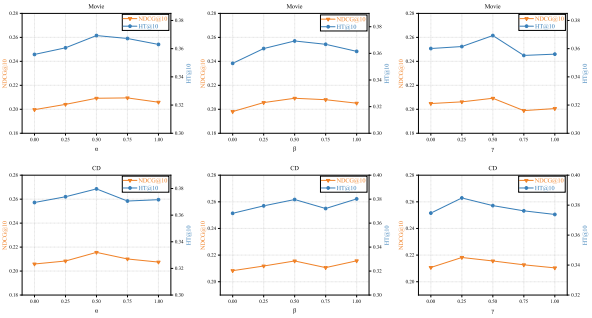


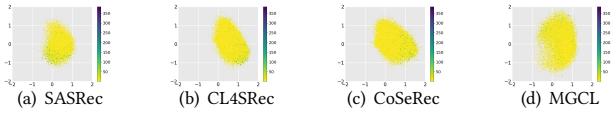
Figure 5: Performance of our MGCL with the weights of different contrastive losses.

recommendation performance, since the supervised signal should assist but not dominate the training process.

4.9 Quality of Item Representations (RQ5)

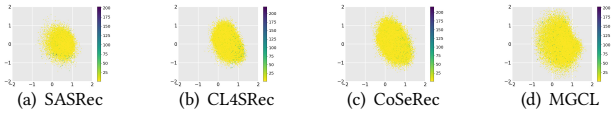
In this subsection, we explore the help of the contrastive learning mechanism used in our MGCL for learning item representations. We project the item embeddings of the trained model into 2D by singular value decomposition (SVD) and visualize the results. The item embedding distributions of SASRec, CL4Srec, CoSeRec and our MGCL on the Movie and CD datasets are shown in Fig. 6 and Fig. 7, respectively. Notice that the shades of color represent the frequency of the item's occurrence.

It can be observed that the distribution of the item embeddings learned by SASRec is in a narrow range and the high frequency items tend to be distributed on the same side. This indicates that the models lack the ability to distinguish different items and cannot model the features of diverse items well [5, 33]. CL4SRec and CoSeRec learn better embedding distributions than SASRec because they introduce some contrastive tasks to assist the learning of item embeddings which maximizes the difference of negative pairs. However, both models are limited by the data of a single domain and cannot learn a broader embedding distribution. It is clear that our MGCL effectively expands the embedding space, and both high and low frequency items can be uniformly distributed among it. We believe that this is because with feature-level contrastive learning across domains, the model is able to obtain more effective information to distinguish different items, and has more capabilities to model diverse features.



(a) SASRec (b) CL4SRec (c) CoSeRec (d) MGCL

Figure 6: Item embeddings on the Movie dataset.



(a) SASRec (b) CL4SRec (c) CoSeRec (d) MGCL

Figure 7: Item embeddings on the CD dataset.

5 CONCLUSIONS AND FUTURE WORK

In this paper, we propose a generic framework named multi-view graph contrastive learning (MGCL) for cross-domain sequential recommendation. Specifically, we tackle the problem from the perspective of an intra-domain item representation view and an inter-domain user preference view. We adopt the item representation contrastive learning to achieve the complementation of the sequential information and the collaborative information. Moreover, we employ preference contrastive learning to enable knowledge transfer of users' preferences from different domains. Extensive empirical studies on three real-world datasets indicate that our MGCL significantly outperforms thirteen competitive baselines. For future works, we aim to apply our MGCL to scenes of cross-domain or cross-organization privacy-aware federated recommendation [18], which can reduce the risk of privacy leakage due to the introduction of rich source-domain data.

ACKNOWLEDGMENTS

We thank the support of National Natural Science Foundation of China No. 62172283, No. 61836005 and No. 62272315.

REFERENCES

[1] Nawaf Alharbi and Doina Caragea. 2022. Cross-Domain Self-Attentive Sequential Recommendations. In *Proceedings of International Conference on Data Science and Applications (ICONDATA '22)*. 601–614.
[2] Philip Bachman, R. Devon Hjelm, and William Buchwalter. 2019. Learning Representations by Maximizing Mutual Information Across Views. In *Advances in Neural Information Processing Systems (NeurIPS '19)*. 15509–15519.
[3] Jiangxia Cao, Xin Cong, Jiawei Sheng, Tingwen Liu, and Bin Wang. 2022. Contrastive Cross-Domain Sequential Recommendation. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management (CIKM '22)*. 138–147.
[4] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey E. Hinton. 2020. A Simple Framework for Contrastive Learning of Visual Representations. In

Proceedings of the 37th International Conference on Machine Learning (ICML '20). 1597–1607.

[5] Jun Gao, Di He, Xu Tan, Tao Qin, Liwei Wang, and Tie-Yan Liu. 2019. Representation Degeneration Problem in Training Natural Language Generation Models. In *Proceedings of the 7th International Conference on Learning Representations (ICLR '19)*.

[6] Lei Guo, Li Tang, Tong Chen, Lei Zhu, Quoc Viet Hung Nguyen, and Hongzhi Yin. 2021. DA-GCN: A Domain-aware Attentive Graph Convolution Network for Shared-account Cross-domain Sequential Recommendation. In *Proceedings of the 13th International Joint Conference on Artificial Intelligence (IJCAI '21)*. 2483–2489.

[7] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yong-Dong Zhang, and Meng Wang. 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 (SIGIR '20)*. 639–648.

[8] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural Collaborative Filtering. In *Proceedings of the 26th International Conference on World Wide Web (WWW '17)*. 173–182.

[9] Yun He, Yin Zhang, Weiwen Liu, and James Caverlee. 2020. Consistency-Aware Recommendation for User-Generated Item List Continuation. In *Proceedings of the 13th International Conference on Web Search and Data Mining (WSDM '20)*. 250–258.

[10] Balázs Hidasi and Alexandros Karatzoglou. 2018. Recurrent Neural Networks with Top-k Gains for Session-based Recommendations. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM '18)*. 843–852.

[11] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. 2016. Session-based Recommendations with Recurrent Neural Networks. In *Proceedings of the 4th International Conference on Learning Representations (ICLR '16)*.

[12] Guangneng Hu, Yu Zhang, and Qiang Yang. 2018. CoNet: Collaborative Cross Networks for Cross-Domain Recommendation. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM '18)*. 667–676.

[13] Wang-Cheng Kang and Julian J. McAuley. 2018. Self-Attentive Sequential Recommendation. In *Proceedings of the 18th IEEE International Conference on Data Mining (ICDM '18)*. 197–206.

[14] Chenyi Lei, Yong Liu, Lingzi Zhang, Guoxin Wang, Haihong Tang, Houqiang Li, and Chunyan Miao. 2021. SEMI: A Sequential Multi-Modal Information Transfer Network for E-Commerce Micro-Video Recommendations. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '21)*. 3161–3171.

[15] Chenglin Li, Mingjun Zhao, Huanming Zhang, Chenyun Yu, Lei Cheng, Guoqiang Shu, Beibei Kong, and Di Niu. 2022. RecGURU: Adversarial Learning of Generalized User Representations for Cross-Domain Recommendation. In *Proceedings of the 15th ACM International Conference on Web Search and Data Mining (WSDM '22)*. 571–581.

[16] Pan Li and Alexander Tuzhilin. 2020. DDTCDR: Deep Dual Transfer Cross Domain Recommendation. In *Proceedings of the 13th International Conference on Web Search and Data Mining (WSDM '20)*. 331–339.

[17] Jing Lin, Weike Pan, and Zhong Ming. 2020. FISSA: Fusing Item Similarity Models with Self-Attention Networks for Sequential Recommendation. In *Proceedings of the 14th ACM Conference on Recommender Systems (RecSys '20)*. 130–139.

[18] Zhaohao Lin, Weike Pan, Qiang Yang, and Zhong Ming. 2023. Recommendation Framework via Fake Marks and Secret Sharing. *ACM Transactions on Information Systems* 41(2) (2023), 40:1–40:37.

[19] Xiao Liu, Fanjin Zhang, Zhenyu Hou, Li Mian, Zhaoyu Wang, Jing Zhang, and Jie Tang. 2023. Self-Supervised Learning: Generative or Contrastive. *IEEE Transactions on Knowledge and Data Engineering* 35, 1 (2023), 857–876.

[20] Zhiwei Liu, Yongjun Chen, Jia Li, Philip S. Yu, Julian J. McAuley, and Caiming Xiong. 2021. Contrastive Self-supervised Sequential Recommendation with Robust Augmentation. *arXiv preprint arXiv:2108.06479* (2021).

[21] Muyang Ma, Pengjie Ren, Yujie Lin, Zhumin Chen, Jun Ma, and Maarten de Rijke. 2019. π-net: A Parallel Information-Sharing Network for Shared-Account Cross-Domain Sequential Recommendations. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '19)*. 685–694.

[22] Tong Man, Huawei Shen, Xiaolong Jin, and Xueqi Cheng. 2017. Cross-Domain Recommendation: An Embedding and Mapping Approach. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI '17)*. 2464–2470.

[23] Julian McAuley, Christopher Targett, Qinfeng Shi, and Anton van den Hengel. 2015. Image-based Recommendations on Styles and Substitutes. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '15)*. 43–52.

[24] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. 2018. Representation Learning with Contrastive Predictive Coding. *arXiv preprint arXiv:1807.03748* (2018).

[25] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian Personalized Ranking from Implicit Feedback. In *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI '09)*. 452–461.

[26] Steffen Rendle, Christoph Freudenthaler, and Lars Schmidt-Thieme. 2010. Factorizing Personalized Markov Chains for Next-basket Recommendation. In *Proceedings of the 19th International Conference on World Wide Web (WWW '10)*. 811–820.

[27] Ajit P Singh and Geoffrey J Gordon. 2008. Relational Learning via Collective Matrix Factorization. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '08)*. 650–658.

[28] Fei Sun, Jun Liu, Jian Wu, Changhua Pei, Xiao Lin, Wenwu Ou, and Peng Jiang. 2019. BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM '19)*. 1441–1450.

[29] Wenchao Sun, Muyang Ma, Pengjie Ren, Yujie Lin, Zhumin Chen, Zhaochun Ren, Jun Ma, and Maarten De Rijke. 2021. Parallel Split-Join Networks for Shared Account Cross-domain Sequential Recommendations. *IEEE Transactions on Knowledge and Data Engineering* (2021).

[30] Jiaxi Tang and Ke Wang. 2018. Personalized top-N Sequential Recommendation via Convolutional Sequence Embedding. In *Proceedings of the 11th ACM International Conference on Web Search and Data Mining (WSDM '18)*. 565–573.

[31] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All You Need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS '17)*. 6000–6010.

[32] Xiao Wang, Nian Liu, Hui Han, and Chuan Shi. 2021. Self-supervised Heterogeneous Graph Neural Network with Co-contrastive Learning. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD '21)*. 1726–1736.

[33] Ziyang Wang, Huoyu Liu, Wei Wei, Yue Hu, Xian-Ling Mao, Shaojian He, Rui Fang, and Dangyang Chen. 2022. Multi-level Contrastive Learning Framework for Sequential Recommendation. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management (CIKM '22)*. 2098–2107.

[34] Shu Wu, Yuyuan Tang, Yanqiao Zhu, Liang Wang, Xing Xie, and Tieniu Tan. 2019. Session-based Recommendation with Graph Neural Networks. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI '19)*. 346–353.

[35] Xin Xia, Hongzhi Yin, Junliang Yu, Qinyong Wang, Lizhen Cui, and Xiangliang Zhang. 2021. Self-supervised Hypergraph Convolutional Networks for Session-based Recommendation. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI '21)*. 4503–4511.

[36] Ruobing Xie, Qi Liu, Liangdong Wang, Shukai Liu, Bo Zhang, and Leyu Lin. 2022. Contrastive Cross-domain Recommendation in Matching. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '22)*. 4226–4236.

[37] Xu Xie, Fei Sun, Zhaoyang Liu, Shiwen Wu, Jinyang Gao, Jiandong Zhang, Bolin Ding, and Bin Cui. 2022. Contrastive Learning for Sequential Recommendation. In *Proceedings of the 38th IEEE International Conference on Data Engineering (ICDE '22)*. 1259–1273.

[38] Chengfeng Xu, Pengpeng Zhao, Yanchi Liu, Victor S. Sheng, Jiajie Xu, Fuzhen Zhuang, Junhua Fang, and Xiaofang Zhou. 2019. Graph Contextualized Self-attention Network for Session-based Recommendation. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI '19)*. 3940–3946.

[39] Xiaolin Zheng, Jiajie Su, Weiming Liu, and Chaochao Chen. 2022. DDGHM: Dual Dynamic Graph with Hybrid Metric Training for Cross-Domain Sequential Recommendation. In *Proceedings of the 30th ACM International Conference on Multimedia (MM '22)*. 471–481.

[40] Feng Zhu, Yan Wang, Chaochao Chen, Jun Zhou, Longfei Li, and Guanfeng Liu. 2021. Cross-Domain Recommendation: Challenges, Progress, and Prospects. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI '21)*. 4721–4728.