Social Recommendation Based on Multi-Auxiliary Information Constrastive Learning

Feng Jiang 1, Yang Cao 2*, Huan Wu 3, Xibin Wang 4, Yuqi Song 5 and Min Gao 2

1 School of Finance and Management, Chongqing Business Vocational College, Chongqing 401331, China  
2 School of Big Data and Software Engineering, Chongqing University, Chongqing 401331, China  
3 College of Environmental Science and Engineering, Tongji University, Shanghai 200092, China  
4 School of Data Science, Guizhou Institute of Technology, Guiyang 550003, China  
5 Department of Computer Science and Engineering, University of South Carolina, Columbia, SC 29201, USA  
* Correspondence: caoyang@cqu.edu.cn

Abstract: Social recommendation can effectively alleviate the problems of data sparseness and the cold start of recommendation systems, attracting widespread attention from researchers and industry. Current social recommendation models use social relations to alleviate the problem of data sparsity and improve recommendation performance. Although self-supervised learning based on user-item interaction can enhance the performance of such models, multi-auxiliary information is neglected in the learning process. Therefore, we propose a model based on self-supervision and multi-auxiliary information using multi-auxiliary information, such as user social relationships and item association relationships, to make recommendations. Specifically, the user social relationship and item association relationship are combined to form a multi-auxiliary information graph. The user-item interaction relationship is also integrated into the same heterogeneous graph so that multiple pieces of information can be spread in the same graph. In addition, we utilize the graph convolution method to learn user and item embeddings, whereby the user embeddings reflect both user-item interaction and user social relationships, and the item embeddings reflect user-item interaction and item association relationships. We also design multi-view self-supervising auxiliary tasks based on the constructed multi-auxiliary views. Signals generated by self-supervised auxiliary tasks can alleviate the problem of data sparsity, further improving user/item embedding quality and recommendation performance. Extensive experiments on two public datasets verify the superiority of the proposed model.

Keywords: social relations; auxiliary information; self-supervised learning; graph convolutional neural network

MSC: 68T01

1. Introduction

The rapid development of the Internet has made life more convenient and produced a large amount of information, causing the problem of information overload. It is difficult for users to select a target product that matches their preferences among hundreds of millions of products. The recommendation system significantly alleviates information overload and improves user experience. However, it is difficult to mitigate the problems of data sparsity and the cold start problem [1–4] in recommendation systems. In recommendation scenarios, user preferences are influenced by the preferences of friends [5,6]. Based on this hypothesis, researchers have integrated users’ social information into the recommendation system as auxiliary information, which can alleviate the problems of data sparseness and cold start problems, thus forming social recommendations.
Some researchers have attempted to make social recommendations based on graph embedding learning of heterogeneous networks [7–9]. Implicit friends with similar preferences do not have explicit links, but they can be indirectly linked based on what they have interacted with. Implicit friends can be used to mine more reliable information from sparse data, specifically learning node embeddings that can accurately express user preferences. Mainstream social recommendation models use heterogeneous networks to describe user social relations and user–item interaction relations and then use the graph embedding learning method to obtain user/item node representation. Node embedding can express node attributes or relationships between nodes (such as user preferences). For example, IF-BPR [8] designed “user–item–user”, “user–user–user”, and other multi-meta paths based on domain knowledge to guide a random walk in a heterogeneous network, thus learning high-quality node embedding. MoHINRec [9] used the meta-path of a variety of motif structures (triangular structures that reflect strong connections between nodes) to guide the random walk and obtain more accurate node embedding. In recent years, graph neural networks (GNNs) have achieved considerable success in node classification and link prediction. Owing to their powerful modeling ability in graph relation, GNNs are also applied in the field of recommendation systems. However, there are three challenges in social recommendation based on GNNs: (1) information is not fully mined from existing data as auxiliary information; (2) users’ social relationships have limited ability to alleviate data sparsity; and (3) information is transmitted independently in the user–item interaction graph, and the user social network graph and node embedding are formed independently, whereas user–item interactions and user social relationships do not affect user preference simultaneously. To tackle these challenges, we design a social recommendation model based on self-supervision and multi-auxiliary information.

The main contributions of this paper are as follows:

- We mine item association relationships, user social relationships, and user–item interaction relationships as auxiliary information to alleviate the problem of data sparsity. Unlike the existing social recommendation models of graph neural networks that independently carry out interactive information and social information dissemination, we design a dissemination mode to make multiple auxiliary information affect the formation of user/item embedding simultaneously.
- We design self-supervised auxiliary tasks for the social recommendation scenario to improve the node embedding quality and alleviate data sparsity, construct several views according to different combinations of auxiliary information, and maximize the mutual information of node embedding under different perspectives based on contrastive self-supervised learning.
- We conduct extensive experiments on two public datasets to demonstrate the effectiveness of the proposed model and analyze the benefits of auxiliary information and self-supervised tasks.

2. Related Work

In this section, we introduce graph neural networks and contrastive self-supervised learning.

2.1. Graph Neural Networks (GNNs)

In recent years, GNNs have attracted increasing attention because, owing to their excellent performance on various tasks. Inspired by the success of other fields, such as node classification and link prediction, the applicability of GNNs in recommendation tasks was investigated. In particular, GCN [10] has driven a large number of graph-based neural network recommendation models, such as GCMC [11], NGCF [12], and LightGCN [13]. The basic idea of these GCN-based models is to improve the representation of the target node by aggregating the node representation of the neighbor [14] to obtain higher-order neighbor information in the user–item interaction graph. In addition to these generic
models, GNNs support other recommendation methods for specific graphs, such as session and social graphs.

GNNs are often used for information transmission in social networks because information is transmitted in social networks the same way as in GNNs. Researchers naturally transplanted GNN into social recommendation work. GraphRec is the first model to introduce GNN into social recommendation [15]. It learns target node embedding by aggregating first-order neighbor information in the user–item interaction graph and the social network graph. User embedding of DiffNet [10] comes from the social network graph and the user–item interaction graph. It carries out more profound information dissemination in the social network through a multi-layer GNN. Moreover, node embedding in the user–item interaction graph only comes from the first-order neighbor. DiffNet uses a multi-layer GNN structure to realize the dynamic propagation of social influence in the social network. Wu et al. [16] proposed a dual-graph attention network for collaborative learning of node embedding influenced by two layers of society. DGRec uses two circulating neural networks to dynamically simulate user behavior and social influence [17]. Yu et al. [18] enhanced social recommendation with adversarial graph convolutional networks to process complicated high-order interactions among users. Later, they [19] improve social recommendation with a multi-channel hypergraph convolutional network to leverage high-order user relations. Huang et al. [20] proposed a knowledge-aware coupled GNN that injects knowledge across items and users into the recommendation. Yang et al. [21] proposed the ConsisRec model to calculate the consistency score between neighbors as the probability of sampling neighbors and further handle the problem of relationship inconsistency through an attention mechanism.

2.2. Contrastive Self-Supervised Learning

Self-supervised learning was first proposed in the field of robotics, whereby training data are automatically generated from data of two or more sensors. The principle of self-supervised learning can be explained a description of complete data based on observation of different aspects or different parts of data. Self-supervised learning enhances data by deforming, intercepting, and disturbing the original data and generates data as pseudo-labels of the original data to make up for the data deficiency. Self-supervised learning can be divided into two types: generative self-supervised learning and contrastive self-supervised learning. In this paper, the comparative self-supervised learning method is used to provide more auxiliary information to alleviate the problem of data sparsity and realize a recommendation algorithm.

Contrast learning is a discriminant method with the aim of making the embeddings of similar samples (positive samples) closer to each other in the representation space and the embeddings of different samples (negative samples) farther apart. This method uses a similarity measure to quantify the distance between two inserts (commonly known as cosine similarity). Studies based on contrastive self-supervision have made significant progress, such as SwAV [22], MoCo [23], and SimCLR [24], and their extensions, with performances comparable to those of related models based on supervised learning.

Pseudo-label construction is an essential strategy for embedded learning based on contrastive self-supervised learning. Positive samples and negative samples are essentially pseudo-labels of data to expand training data. The original samples are based on positive/negative samples for supervised learning. The proposed goal of self-supervised learning is to reduce the cost of manual labeling, so the generation of false labels (the selection of positive/negative samples) is an automatic process.

Based on contrastive self-supervised learning, the model needs to design auxiliary tasks to complete contrastive learning and assist with the main tasks of specific scenes to train the model. The training process is as follows:
• The auxiliary task performs data enhancement based on the original sample. The original and positive/negative samples generate corresponding low-dimensional embedding through the encoder and construct the loss function through contrast learning.

• The main task generates the corresponding low-dimensional embedding of the original sample through the encoder and constructs the principal loss function through the main task. Finally, by comparing the loss function with the main loss function, higher quality embedding can be obtained.

Contrastive self-supervised learning is also applied flexibly in graph embedding learning (graph contrast learning for short). This work mainly constructs self-supervised signals from graph structures of different perspectives to explore higher-quality graph structure embedding [25,26]. Generally, a new perspective can be obtained through random data enhancement of the same graph. Common data enhancement methods include but are not limited to the random deletion of nodes, random deletion of edges, the random transformation of features or attributes, random walk-based pattern graphs, etc. Inspired by such work, some researchers began to apply graph contrast learning to recommendation tasks [27–30]. Zhou et al. [30] designed self-supervised auxiliary tasks, specifically adding random embedding masks for item embedding and randomly skipping given items and sub-sequences to carry out pre-training for sequence recommendation. Yao et al. [29] proposed a two-tower network structure based on DNNs (deep neural networks), on which random feature mask and random feature discarding operations were carried out for self-supervised item recommendations. Ma et al. [28] reconstructed short-term future sequences by observing long-term sequences, which essentially mined more self-supervised signals using feature masks. Wu et al. [27] summarized the above random data enhancement operations (random node/edge deletion and random walk) and integrated them into a recommendation framework based on self-supervised graph learning. Long et al. [31] proposed a heterogeneous graph neural network based on meta-relation and used self-supervised learning to guide the interaction between users and items under different views, incorporating the knowledge information of the item and the high-level semantic relationship between users and items into the user representation. Liu et al. [32] designed two new information augmentation methods. Furthermore, they proposed a contrastive self-supervised learning framework, CoSeRec, for sequence recommendation, which alleviates the problems of data sparsity and noisy interaction issues. Wu et al. [33] proposed a social recommendation model that disentangles the collaborative domain and social domain to learn user representations separately and uses cross-domain contrastive learning to further improve the recommendation performance.

Unlike these models, we use item-association-aware contrastive learning in self-supervised learning. We maximize the mutual information of node embedding in two item-association-aware views in self-supervised learning; view 1 is constructed by the user–item interaction relationship and the item association relationship, and view 2 is formed by the user–item interaction relationship and the user social relationship. Although the model proposed in [33] also uses cross-domain contrastive learning, the domains are the user–item interaction domain and the social relationship domain, not taking item association into consideration. Our model is designed to utilize more auxiliary information to improve recommendation performance, and self-supervised learning can finetune the embeddings from different kinds of auxiliary information. There are two main differences between our model and SEPT [34]. First, we use item-association-aware contrastive learning in self-supervised learning, whereas SEPT does not take item association into consideration. Second, SEPT focuses more on finding positive samples for contrastive learning and uses tri-training to determine which samples are positive; however, our model does not attempt to identify additional positive samples and treat the same node in different views as a positive sample.
3. Problem Analysis

Accurately capturing user preferences is the key to improving the recommendation quality of recommendation systems. The recommendation algorithm is committed to learning more accurate user embedding to improve the recommendation performance. The current social recommendation algorithm based on a graph neural network believes that user preference is jointly determined by items that have interacted historically, as well as the influence of social friends. The process of learning user/item embedding in this kind of algorithm is summarized as follows: In the first step, by default, user–item interaction influence and social influence are spread in an independent scope, and information is spread and embedded in the user–item interaction graph and user social network graph. In the second step, there are two different embeddings of user nodes in the two graphs. The two embeddings representing user–item interaction information and social influence are aggregated to form the final embeddings.

Social recommendations based on graph neural networks are subject to the following problems:

- They carry out user–item interaction information and social information dissemination in two graphs instead of being affected by both simultaneously (more realistic).
- As auxiliary information, social relationship alleviates the problem of data sparsity in the recommendation system, but its role is limited, and more auxiliary information is needed.

4. Methodology

To alleviate the above problems, we propose a social recommendation model based on self-supervised and multi-auxiliary information (SlightGCN). We combine user social relationships, item association relationships, and self-supervised auxiliary signals as auxiliary information to carry out social recommendations. First, the model SlightGCN mines item association relationships based on user–item interaction records and item attributes. Then, a heterogeneous network is constructed, including user social relationships, item association relationships, and user–item interaction relationships. In the heterogeneous network graph containing multi-auxiliary information, information is transmitted and output-embedded through a GCN. The user/item embedding containing rich information is used to recommend the main task. In addition, two heterogeneous networks with different perspectives are constructed according to different combinations of existing auxiliary information, and self-supervised auxiliary tasks are constructed to maximize the mutual information of nodes under different perspectives to obtain higher quality node embedding. Finally, the model is trained through the combination of primary task (recommendation task) loss and auxiliary task (contrastive self-supervised learning).

SlightGCN can be generally divided into three parts: the heterogeneous network construction, main recommendation task, and self-supervision auxiliary task (See Figure 1).
4.1. Heterogeneous Network Construction

First, the association relationship is mined from the user–item interaction relationship and item attribute through the meta-path rule to alleviate the data sparsity problem. Secondly, three types of information (user social relationship, user–item interaction relationship, and item association relationship) are integrated into the same heterogeneous network in the form of edges. The construction process of a heterogeneous network is shown in Figure 2.

![Figure 2. The heterogeneous network construction process.](image)

In the process of information dissemination of the recommendation scenario, the user node is influenced by friends and items that have been interacted with, and the item node is influenced by items that are closely connected and users that have been interacted with to learn user/item embedding of higher quality. To realize this information transmission mode, user–item interaction relationship, user social relationship, and item association relationship need to be integrated into the same heterogeneous network in the form of an edge. The user node receives information from friends and historical interaction records through user–item interaction relationships and user social relationships, and the item node receives information from closely related items and users who have interacted with each other through user–item interaction relationships and item association relationships.

4.2. Main Tasks

A recommendation algorithm is based on the main task of supervised learning to optimize model parameters. The prediction is established on user embedded ($e_u$) and item-embedded ($e_i$). Generally, the inner product of the embedded $e_u^T e_i$ is used to predict user’s ($u$) preference degree ($\hat{y}_{ui}$) for item $i$, as in (1):

$$\hat{y}_{ui} = e_u^T e_i.$$  \hspace{1cm} (1)

Specifically, it assumes the observed interactions as monitoring signals and makes the predicted preference ($\hat{y}_{ui}$) as close as possible to the real preference ($y_{ui}$). It takes the unobserved data as a negative sample. We take the generally used Bayesian personalized recommendation [35] (BPR) as the loss function for the ranking recommendation task. The core idea is that the target user ($u$) prefers interactive item $i$ (observable data) to uninteractive item $j$ (unobserved data). The specific loss function is shown in (2):

$$\mathcal{L}_{\text{main}} = \sum_{(u,i,j) \in \mathcal{D}} -log(\hat{y}_{ui} - \hat{y}_{uj}),$$  \hspace{1cm} (2)

where $\mathcal{D} = \{(u, i, j) | (u, i) \in \mathcal{D}^+, (u, j) \in \mathcal{D}^- \}$ is all training data, $\mathcal{D}^+$ is the set of observable interactions, and $\mathcal{D}^-$ is the set of unobserved interactions.
The basis of the recommendable task is to learn user/item embedding. Heterogeneous networks can output user/item embedding through a graph encoder. Inspired by the excellent graph convolution neural network recommendation model LightGCN [13], feature transformation, nonlinear transformation, and self-connection operation are redundant for the collaborative filtering recommendation model. We integrate user social relations and item associations to realize that user nodes are affected by social friends and interactive items simultaneously, and item nodes are affected by interactive users and closely connected items simultaneously.

In graph convolution mode, simple weighted aggregation is sampled in the convolution process and feature transformation. The specific convolution operation is shown in (3):

\[ e_u^{(k+1)} = \frac{1}{|\mathcal{N}_u||\mathcal{N}_p|} e_p^{(k)} \quad e_i^{(k+1)} = \frac{1}{|\mathcal{N}_i||\mathcal{N}_q|} e_q^{(k)}, \]

where \( e_u^{(k+1)} \) and \( e_i^{(k+1)} \) represent the embedment of user \( u \) and item \( i \) at the \((k + 1)\) layer, respectively; and the \((k + 1)\) layer embedment is aggregated from the \( k \) layer embedment. Taking the formation of \( e_u^{(k+1)} \) as an example, user \( u \)'s neighbor, \( \mathcal{N}_u \), contains the user's neighbor and item neighbor, and \( e_u^{(k+1)} \) is aggregated by embedding the \( k \) layer of user \( u \)'s neighbor, \( \mathcal{N}_u \). \( e_i^{(k+1)} \) is aggregated by embedding the \( k \) layer of item \( i \)'s neighbor, \( \mathcal{N}_i \).

After the multi-layer convolution, the final node embedding is obtained by average aggregation of each layer embedding of the node, as shown in (4):

\[ e_u = \frac{e_u^{(0)} + e_u^{(1)} + \cdots + e_u^{(k+1)}}{k+1}, \quad e_i = \frac{e_i^{(0)} + e_i^{(1)} + \cdots + e_i^{(k+1)}}{k+1}. \]

In the graph matrix representation of the convolution process, the heterogeneous network adjacency matrix \( A \in \mathbb{R}^{(N+M)\times(N+M)} \) is composed of the user–item interaction matrix \( R \) and its transpose \( R^T \), the user social matrix \( S \), and the item association matrix \( T \). The specific expression of adjacency matrix \( A \) is shown in (5):

\[ A = \begin{pmatrix} S & R \\ R^T & T \end{pmatrix}. \]

User embedding and item embedding constitute the 0-layer embedding matrix, \( E^{(0)} \in \mathbb{R}^{(N+M)\times d} \), where \( d \) is the embedding dimension, and \( E^{(0)} \) is randomly initialized. The matrix form of the convolution operation is shown in (6),

\[ E^{(k+1)} = (D^{-1/2}AD^{-1/2})E^{(k)}, \]

where \( D \in \mathbb{R}^{(N+M)\times(N+M)} \) is a diagonal matrix, and the value of \( D_{ii} \) is the number of non-zero elements in the \( i \)th row of matrix \( A \).

Finally, the embedding matrix \( E \in \mathbb{R}^{(N+M)\times d} \) is evenly aggregated by each embedding matrix, as shown in (7),

\[ E = \frac{E^{(0)} + E^{(1)} + \cdots + E^{(k)}}{k+1}. \]

### 4.3. Self-Supervised Auxiliary Tasks

The main view composed of three types of relationships (user–item interaction relationship, user social relationship, and item association relationship) introduced in the previous section helps to recommend the embedding of main task learning nodes. When a designer models a character, it is necessary to observe the character information from different aspects and combine the information from multiple perspectives to build a more accurate character model. In the recommendation scenario, the recommendation algorithm needs to model the user and the item before recommending the user to build a more accurate portrait of the person/item (user/item embedding). The graph neural network
plays the role of a model and can generate more accurate node embedding by combining node information from different perspectives.

Based on the main view, we remove the user’s social relationship and item association relationship and generate auxiliary view 1 and auxiliary view 2, respectively. In the process of information transmission, the two auxiliary views correspond to the adjacency matrices $A_1$ and $A_2$, respectively, as shown in (8). The convolution of the two auxiliary views is to generate two groups of different embedded nodes ($E_1$ and $E_2$), as shown in (9).

$$A_1 = \begin{pmatrix} 0 & R \\ R^T & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} S & R \\ R^T & 0 \end{pmatrix}, \quad (8)$$

$$E_1^{(k+1)} = \left(D_1^{-1/2}A_1D_1^{-1/2}\right)E_1^{(k)}, \quad E_2^{(k+1)} = \left(D_2^{-1/2}A_2D_2^{-1/2}\right)E_2^{(k)}. \quad (9)$$

Contrastive self-supervised learning is used to maximize the mutual information of node embedding in different views to learn node embedding of higher quality. Specifically, the node embedding of the same node in different views is a positive sample pair (i.e., $\{(e'_u, e''_v) | u \in U\}$), and the node embedding of different nodes in different views is a negative sample pair (i.e., $\{(e'_u, e''_v) | u, v \in U, u \neq v\}$). Auxiliary tasks make similar nodes as similar as possible, whereas different sample nodes are embedded as far away as possible. The self-supervised loss function is formed according to the comparative loss, InfoNCE, as shown in (10):

$$L_{ssl} = \sum_{u \in U} -\log \frac{\exp(s(e'_u,e''_v)/\tau)}{\sum_{v \in U} \exp(s(e'_u,e''_v)/\tau)} , \quad (10)$$

where the function $s(\cdot)$ is cosine similarity, which is used to measure the distance between two embeddings, and the hyperparameter $\tau$ is the temperature index, which can reduce or amplify the effect of distance between nodes. Finally, as shown in (11), the loss function ($L_{ssl}$) of the self-supervision auxiliary task is the sum of the loss function ($L_{ssl}^{user}$) of the user node and the loss function ($L_{ssl}^{item}$) of the item node.

To improve the recommendation performance, the main recommendation task and self-supervised auxiliary task are combined for training. The loss function of the combined training is shown in (12), where $\lambda_1$ and $\lambda_2$ are the hyperparameters, and $\theta$ is the model parameter.

$$L = L_{main} + \lambda_1 L_{ssl} + \lambda_2 \|\theta\|^2. \quad (12)$$

5. Experiments

In this study, extensive experiments were carried out on two public datasets to verify the following points: (1) the advantages of SlightGCN in terms of recommendation performance, (2) that SlightGCN effectively alleviates data sparsity and cold start problems; (3) that multi-auxiliary information (users’ social relationship, item association relationship, and self-supervised learning signal) plays an essential role in improving the recommendation performance, and (4) the influence of hyperparameters on SlightGCN.

5.1. The Datasets

We conducted extensive experiments on two public datasets: DoubanMovie and DoubanBook. Detailed information on the datasets is shown in Table 1, including the number of users, the number of items, the number of user–item interactions, the number of user social relationships, and the density of user–item interactions.
Table 1. Dataset statistics.

<table>
<thead>
<tr>
<th>The Dataset</th>
<th>Number of Users</th>
<th>Number of Items</th>
<th>Number of User–Item Interactions</th>
<th>Number of Social Connections</th>
<th>Density of Interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>DoubanMovie</td>
<td>13,367</td>
<td>12,677</td>
<td>1,067,278</td>
<td>4085</td>
<td>0.63%</td>
</tr>
<tr>
<td>DoubanBook</td>
<td>13,024</td>
<td>22,347</td>
<td>792,062</td>
<td>169,150</td>
<td>0.27%</td>
</tr>
</tbody>
</table>

DoubanMovie is a movie dataset of the Douban platform, which contains 13,367 users’ 1,067,278 viewing behaviors on 12,677 movies and 4085 users’ social friend relationships. We select two crucial attributes of film type and director in DoubanMovie to excavate the social relations of objects. DoubanBook is a book dataset of the Douban platform, which contains 792,062 interactions between 13,024 users and 22,347 books and 169,150 social friend relationships between users. Two important attributes of DoubanBook (publisher and author) are selected to mine the social relations. In this study, the dataset is divided into a training set, a validation set, and a test set (70%, 10%, and 20%, respectively). The model was cross-verified ten times, and the average value was taken as the result.

5.2. Baselines and Metrics

To verify the recommendation performance of SlightGCN, we select six baseline algorithms, as shown below:

- **BPR**: Bayesian personalized ranking [35], a classical sorting recommendation algorithm. Based on user–item interaction information, it is assumed that the target user prefers interactive items rather than non-interactive items.

- **SBPR**: A classic social recommendation algorithm [36] that integrates social relations to optimize the item preference priority of target users based on BPR.

- **DiffNet**: A social recommendation algorithm based on a graph neural network [37] that simulates the social influence of dynamic propagation in user social networks.

- **LightGCN**: A recommendation algorithm based on a graph convolutional neural network [13] that learns node embedding in a simple convolution mode suitable for collaborative filtering.

- **SGL**: A recommendation algorithm based on self-supervised learning and a graph convolution neural network [27] that creates multiple views by randomly changing the graph structure to improve the embedding quality.

- **SEPT**: A graph convolution neural network recommendation algorithm based on self-supervised learning and collaborative training [34] that finds more suitable positive/negative samples for self-supervised learning through collaborative training to learn more accurate embedding.

Evaluation metrics are precision, recall, $F_1$, and NDCG (normalized impairment cumulative gain).

\[
Prec@k = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{k}, \quad (13)
\]

\[
Rec@k = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{|T(u)|}, \quad (14)
\]

\[
F_1 = \frac{2 \times Prec \times Rec}{Prec + Rec}, \quad (15)
\]

\[
NDCG@k = \frac{dCG}{dCG^*}, \quad (16)
\]

\[
DCG = \sum_{i=1}^{k} \frac{2^{|rel_i|} - 1}{\log_2(i + 1)} \quad (17)
\]
where \( u \) is a user; \( U \) is the set of all users; \( R(u) \) denotes recommended items for user \( u \); \( T(u) \) is user \( u \)'s real liked items; \( rel \) represents item \( i \)'s relevance score, which can be pre-defined; and \( iDCG \) is the DCG of the ideal order for the recommendation set.

In the experiments, the depth of LightGCN, SGL, and SlightGCN for information propagation is three, and that of SEPT is two because SEPT can achieve the best performance in that case [34].

5.3. Results

5.3.1. Overall Comparison

We compare the proposed model, SlightGCN, with six baselines on DoubanMovie and DoubanBook. The advantages and disadvantages of each model are shown by comparing the four metrics in top-10 and top-20 recommendations. Based on the experimental results in Tables 2 and 3 (the bold numbers mean the best performance), the following conclusions can be drawn:

1. SlightGCN’s recommendation performance is significantly better than that of the other five baseline models on the two datasets. Specifically, the evaluation metrics of SlightGCN on DoubanMovie and DoubanBook improved by 1.84–4.26% and 2.08–3.30%, respectively, compared with the suboptimal model. In addition, the interactive data density of the DoubanMovie dataset is higher than that of DoubanBook. Thus, the recommendation performance of all models on the DoubanMovie dataset is significantly better than that of DoubanBook.

2. The results show that SBPR is superior to BPR in some indicators, indicating that SBPR can alleviate data sparsity to a certain extent by integrating users’ social relationships based on BPR. SBPR’s simple approach of integrating users’ social relationships by directly optimizing the order of items does not accurately simulate the impact of social relationships on users’ preferences. DiffNet enables the influence of friends’ preferences on users to propagate dynamically in social networks through multi-layer graph neural network simulation. The performance of DiffNet is better than that of SBPR owing to the ability of the graph neural network to capture graph relation and the dynamic propagation of social influence. LightGCN’s simple convolution mode makes it more suitable for collaborative filtering, and its recommendation performance is significantly higher than that of the previous recommendation model. Based on LightGCN as a graph encoder, collaborative training is used to find more suitable positive and negative samples for self-supervised learning, and SEPT improves the recommendation performance based on LightGCN. Using LightGCN as the graph encoder and designing the self-supervised auxiliary task based on random graph structural disturbance, the SGL recommendation performance was considerably improved.

3. SlightGCN performs best on both datasets. SlightGCN has the following advantages over other models. First, it not only uses user-item interaction relationships and user-social relationships but also mines item association relationships from existing information to form multi-auxiliary information. Second, a more appropriate convolution mode is designed so that users are affected by both user-item interaction relationships and user-social relationships, and objects are affected by both user-item interaction relationships and item association relationships. Third, multi-view self-supervised auxiliary tasks are designed to learn more accurate user/item embedding.

<table>
<thead>
<tr>
<th></th>
<th>BPR</th>
<th>SBPR</th>
<th>DiffNet</th>
<th>SEPT</th>
<th>LightGCN</th>
<th>SGL</th>
<th>SlightGCN</th>
<th>Ascension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prec@10</td>
<td>0.1243</td>
<td>0.1175</td>
<td>0.1311</td>
<td>0.1617</td>
<td>0.1602</td>
<td>0.1642</td>
<td>0.1712</td>
<td>4.26%</td>
</tr>
<tr>
<td>Prec@20</td>
<td>0.1049</td>
<td>0.0994</td>
<td>0.1119</td>
<td>0.1343</td>
<td>0.1331</td>
<td>0.1352</td>
<td>0.1407</td>
<td>4.07%</td>
</tr>
<tr>
<td>Rec@10</td>
<td>0.0784</td>
<td>0.0903</td>
<td>0.1012</td>
<td>0.1218</td>
<td>0.1186</td>
<td>0.1252</td>
<td>0.1281</td>
<td>2.32%</td>
</tr>
</tbody>
</table>
Table 3. Comparison of recommendation performances on DoubanBook.

<table>
<thead>
<tr>
<th></th>
<th>BPR</th>
<th>SBPR</th>
<th>DiffNet</th>
<th>SEPT</th>
<th>LightGCN</th>
<th>SGL</th>
<th>SlightGCN</th>
<th>Ascension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prec@10</td>
<td>0.0715</td>
<td>0.0664</td>
<td>0.0727</td>
<td>0.0942</td>
<td>0.0876</td>
<td>0.1061</td>
<td>0.1096</td>
<td>3.30%</td>
</tr>
<tr>
<td>Prec@20</td>
<td>0.0561</td>
<td>0.0539</td>
<td>0.0583</td>
<td>0.0743</td>
<td>0.0702</td>
<td>0.0826</td>
<td>0.0850</td>
<td>2.91%</td>
</tr>
<tr>
<td>Rec@10</td>
<td>0.0696</td>
<td>0.0725</td>
<td>0.0830</td>
<td>0.1008</td>
<td>0.0909</td>
<td>0.1079</td>
<td>0.1104</td>
<td>2.32%</td>
</tr>
<tr>
<td>Rec@20</td>
<td>0.1027</td>
<td>0.1093</td>
<td>0.1239</td>
<td>0.1480</td>
<td>0.1372</td>
<td>0.1587</td>
<td>0.1620</td>
<td>2.08%</td>
</tr>
<tr>
<td>F1@10</td>
<td>0.0705</td>
<td>0.0693</td>
<td>0.0775</td>
<td>0.0974</td>
<td>0.0892</td>
<td>0.1070</td>
<td>0.1100</td>
<td>2.80%</td>
</tr>
<tr>
<td>F1@20</td>
<td>0.0726</td>
<td>0.0722</td>
<td>0.0793</td>
<td>0.0989</td>
<td>0.0929</td>
<td>0.1086</td>
<td>0.1115</td>
<td>2.67%</td>
</tr>
<tr>
<td>NDCG@10</td>
<td>0.0956</td>
<td>0.0921</td>
<td>0.1019</td>
<td>0.1320</td>
<td>0.1195</td>
<td>0.1496</td>
<td>0.1541</td>
<td>3.01%</td>
</tr>
<tr>
<td>NDCG@20</td>
<td>0.0980</td>
<td>0.0980</td>
<td>0.1090</td>
<td>0.1371</td>
<td>0.1261</td>
<td>0.1532</td>
<td>0.1577</td>
<td>2.94%</td>
</tr>
</tbody>
</table>

5.3.2. Cold-Start User Experiment

Cold-start users have few or no interactions, and extremely sparse data make it difficult to generate recommend for such users. In this section, users with fewer than 10 interactions are selected as cold-start users; the recommendation results for these users are shown in Figures 3 and 4.

Figure 3. Comparison of the cold-start user recommendation performances on DoubanMovie.

Figure 4. Comparison of the cold-start user recommendation performances on DoubanBook.
The recommendation performance of SBPR is better than that of BPR, indicating that users’ social relationships alleviate the problem of data sparsity to some extent. For the other four recommendation models based on GNN, the recommendation performance increases in the order of DiffNet, LightGCN, SEPT, SGL, and SlightGCN. LightGCN, which only uses user–item interaction relationships, achieves a better recommendation performance than DiffNet, which uses the social relationship as auxiliary information. Its simple and efficient convolution structure is more suitable for collaborative filtering. SEPT uses collaborative training to find positive and negative samples for self-supervised learning. SGL is based on the structure of a random disturbance graph for data enhancement, and the generated self-supervised signal alleviates data sparsity and improves the embedding quality. The proposed SlightGCN model incorporates multiple pieces of auxiliary information, such as user social relationship and item association relationship, as well as various perspectives of self-supervised auxiliary task construction, all of which alleviate the problem of data sparsity to a certain extent. In addition, user/item embedding is influenced by multiple relationships simultaneously, improving the capture of user preferences.

5.3.3. The Benefit of Self-Supervised Auxiliary Tasks

To verify that user social relationship, item association relationship, and self-supervised auxiliary tasks can help improve the recommendation performance of SlightGCN, we designed an ablation experiment as follows.

To verify the effectiveness of self-supervised auxiliary tasks, the self-supervised auxiliary tasks were removed to form the UI model variant. In addition, to verify the benefits of user social relationship and item association relationship to improve the model performance, user–item interaction relationship and user social relationship are reserved to form the U model variant, and user–item interaction relationship and item association relationship are reserved to form the I model variant. The experimental results of SlightGCN and its three deformation models on the two datasets are shown in Tables 4 and 5 (the bold numbers mean the best performance).

**Table 4.** SlightGCN ablation experimental results on DoubanMovie.

<table>
<thead>
<tr>
<th>Metrics</th>
<th>LightGCN</th>
<th>Variant-U</th>
<th>Variant-I</th>
<th>Variant-UI</th>
<th>SlightGCN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prec@10</td>
<td>0.1602</td>
<td>0.1635</td>
<td>0.1630</td>
<td>0.1637</td>
<td>0.1712</td>
</tr>
<tr>
<td>Prec@20</td>
<td>0.1331</td>
<td>0.1348</td>
<td>0.1343</td>
<td>0.1350</td>
<td>0.1407</td>
</tr>
<tr>
<td>Rec@10</td>
<td>0.1186</td>
<td>0.1217</td>
<td>0.1223</td>
<td>0.1226</td>
<td>0.1281</td>
</tr>
<tr>
<td>Rec@20</td>
<td>0.1833</td>
<td>0.1876</td>
<td>0.1869</td>
<td>0.1881</td>
<td>0.1931</td>
</tr>
<tr>
<td>F1@10</td>
<td>0.1363</td>
<td>0.1395</td>
<td>0.1397</td>
<td>0.1402</td>
<td>0.1465</td>
</tr>
<tr>
<td>F1@20</td>
<td>0.1542</td>
<td>0.1569</td>
<td>0.1563</td>
<td>0.1572</td>
<td>0.1628</td>
</tr>
<tr>
<td>NDCG@10</td>
<td>0.2034</td>
<td>0.2091</td>
<td>0.2082</td>
<td>0.2095</td>
<td>0.2166</td>
</tr>
<tr>
<td>NDCG@20</td>
<td>0.2043</td>
<td>0.2101</td>
<td>0.2096</td>
<td>0.2104</td>
<td>0.2161</td>
</tr>
</tbody>
</table>

**Table 5.** SlightGCN ablation experimental results on DoubanBook.

<table>
<thead>
<tr>
<th>Evaluation Index</th>
<th>LightGCN</th>
<th>Variant U</th>
<th>Variant I</th>
<th>Variant UI</th>
<th>SlightGCN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prec@10</td>
<td>0.0876</td>
<td>0.0890</td>
<td>0.0893</td>
<td>0.0897</td>
<td>0.1096</td>
</tr>
<tr>
<td>Prec@20</td>
<td>0.0702</td>
<td>0.0712</td>
<td>0.0719</td>
<td>0.0726</td>
<td>0.0850</td>
</tr>
<tr>
<td>Rec@10</td>
<td>0.0909</td>
<td>0.0927</td>
<td>0.0934</td>
<td>0.0941</td>
<td>0.1104</td>
</tr>
<tr>
<td>Rec@20</td>
<td>0.1372</td>
<td>0.1386</td>
<td>0.1390</td>
<td>0.1389</td>
<td>0.1620</td>
</tr>
<tr>
<td>F1@10</td>
<td>0.0892</td>
<td>0.0908</td>
<td>0.0913</td>
<td>0.0918</td>
<td>0.1100</td>
</tr>
<tr>
<td>F1@20</td>
<td>0.0929</td>
<td>0.0941</td>
<td>0.0948</td>
<td>0.0954</td>
<td>0.1115</td>
</tr>
<tr>
<td>NDCG@10</td>
<td>0.1195</td>
<td>0.1223</td>
<td>0.1226</td>
<td>0.1238</td>
<td>0.1541</td>
</tr>
<tr>
<td>NDCG@20</td>
<td>0.1261</td>
<td>0.1285</td>
<td>0.1290</td>
<td>0.1296</td>
<td>0.1577</td>
</tr>
</tbody>
</table>
Compared with LightGCN, variant UI, variant U, and variant I, the three deformation models all improved to a certain extent, indicating that user social relationship and item association relationship can improve the recommendation performance by alleviating data sparsity. However, there is no significant difference in the recommendation performance between variant UI containing two auxiliary relationships and variant U and variant I, containing only one auxiliary relationship, indicating that the user social relationship and item association relationship cannot be well-integrated. SlightGCN removes the user social relationship and the item social relationship to construct different auxiliary views. Contrastive self-supervised learning maximizes the mutual information of the two views and promotes the integration of user social relationships and item association relationships. SlightGCN’s recommendation performance is superior to that of variant UI, which does not include self-supervised assistance tasks.

5.3.4. Parameter Sensitive Analysis

In the training process of the model, it is constrained by the main recommendation task and the self-supervised auxiliary task. The self-supervised task assists the recommendation task to guide the updating direction of model parameters. In the loss function of the joint training, the hyperparameter $\lambda_1$ is the weight coefficient of the loss of the self-supervised auxiliary task, and $\lambda_1$ can control the influence of the self-supervised auxiliary task on the overall recommendation performance of the model. The training process of the parameter $\lambda_1$ ambassador model is more dependent on the information of the self-supervised auxiliary task and less dependent on the information of the self-supervised auxiliary task. The experiments of hyperparameter $\lambda_1$ on dataset DoubanMovie and dataset DoubanBook are shown in Figures 5 and 6, respectively.

![Figure 5. Parameter $\lambda_1$ sensitivity analysis on DoubanMovie.](image-url)
Figure 6. Parameter $\lambda_1$ sensitivity analysis on DoubanBook.

In DoubanMovie, the overall trend of the values of the four metrics is consistent, increasing first and then decreasing with increased $\lambda_1$, with the recommendation performance reaching its peak at $\lambda_1 = 0.009$. In DoubanBook, the overall variation trend of the four metrics is consistent, increasing with increased $\lambda_1$, with the recommendation performance reaching its peak at $\lambda_1 = 0.05$. Two conclusions can be drawn from observations:

- The optimal $\lambda_1 = 0.009$ on DoubanMovie is far less than the optimal $\lambda_1 = 0.05$ on DoubanBook, indicating that the model relies much more on self-supervised auxiliary tasks on DoubanBook than on DoubanMovie. The reasons for the above phenomenon are as follows: the user–item interaction density on DoubanBook is lower than that on DoubanMovie. To learn higher-quality node embedding and improve recommendation performance, the model in the DoubanBook environment needs to obtain more information from self-supervised auxiliary tasks.
- The improvement of the four metrics in DoubanBook with sparse data is much more significant than that in DoubanMovie, indicating the effectiveness of self-supervised auxiliary tasks in alleviating the data sparsity problem.

6. Conclusions

In this paper, we propose a social recommendation model based on self-supervision and multi-auxiliary information (SlightGCN). The social relationships between users and the related relationships between items are mined from user–item interaction data and item attribute information. The relationships between items and users can be used as multiple auxiliary information to alleviate data sparsity. The user–item interaction relationship, user social relationship, and item associate relationship are integrated into the same heterogeneous network in the form of an edge. A heterogeneous network is used as the input of a graph convolutional neural network to carry out the main recommendation task, and an appropriate convolutional mode is designed; user embedding is simultaneously affected by user social relationship and user–item interaction relationship, and item embedding is simultaneously affected by item association relationship and user–item interaction relationship. Auxiliary views are constructed according to different combinations of information, and self-supervised auxiliary tasks of multi-auxiliary views are designed to improve the embedding quality of nodes and recommendation performance. Extensive experiments were conducted on two public datasets to verify the superiority of SlightGCN in terms of recommendation performance, the effectiveness of multi-auxiliary information, and self-supervised auxiliary tasks.

Many data argumentation methods for comparative self-supervised learning are available in the GNN-based method, such as randomly removing edges or nodes, randomly adding edges or nodes, removing/adding edges or nodes on purpose, and even rewiring the graph. Further research is required with deep analysis of the effectiveness of these data argumentation methods within the framework the proposed SlightGCN model.

Author Contributions: Conceptualization, F.J. and Y.C.; methodology, M.G. and Y.C.; software, Y.C.; validation, H.W. and X.W.; formal analysis, Y.S.; writing—original draft preparation, F.J.; writing—review and editing, Y.S.; supervision, M.G. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Natural Science Foundation (72161005), the Research Program of Chongqing Technology Innovation and Application Development (cstc2020kjcsx-pxm1304), Chongqing Technology Innovation and Application Development Project (cstc2020scx-lyjsA0010), and Scientific and Technological Research Program of Chongqing Municipal Education Commission (KJZD-K202204402).

Conflicts of Interest: The authors declare no conflict of interest.

References


