

Deep Social Collaborative Filtering

Wenqi Fan

Department of Computer Science
City University of Hong Kong
wenqifan03@gmail.com

Jianping Wang

Department of Computer Science
City University of Hong Kong
jianwang@cityu.edu.hk

Yao Ma

Data Science and Engineering Lab
Michigan State University
maya04@msu.edu

Dawei Yin

JD.com
yindawei@acm.org

Qing Li

Department of Computing
The Hong Kong Polytechnic
University
csqli@comp.polyu.edu.hk

ABSTRACT

Recommender systems are crucial to alleviate the information overload problem in online worlds. Most of the modern recommender systems capture users' preference towards items via their interactions based on collaborative filtering techniques. In addition to the user-item interactions, social networks can also provide useful information to understand users' preference as suggested by the social theories such as homophily and influence. Recently, deep neural networks have been utilized for social recommendations, which facilitate both the user-item interactions and the social network information. However, most of these models cannot take full advantage of the social network information. They only use information from direct neighbors, but distant neighbors can also provide helpful information. Meanwhile, most of these models treat neighbors' information equally without considering the specific recommendations. However, for a specific recommendation case, the information relevant to the specific item would be helpful. Besides, most of these models do not explicitly capture the neighbor's opinions to items for social recommendations, while different opinions could affect the user differently. In this paper, to address the aforementioned challenges, we propose DSCF, a Deep Social Collaborative Filtering framework, which can exploit the social relations with various aspects for recommender systems. Comprehensive experiments on two-real world datasets show the effectiveness of the proposed framework.

CCS CONCEPTS

- Information systems → Social recommendation; Social recommendation;
- Computing methodologies → Neural networks;

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

RecSys '19, September 16–20, 2019, Copenhagen, Denmark
© 2019 Association for Computing Machinery.
ACM ISBN 978-1-4503-6243-6/19/09...\$15.00
<https://doi.org/10.1145/3298689.3347011>

Jiliang Tang

Data Science and Engineering Lab
Michigan State University
tangjili@msu.edu

KEYWORDS

Social Recommendation, Recommender Systems, Social Network, Recurrent Neural Network, Random Walk, Neural Networks

ACM Reference Format:

Wenqi Fan, Yao Ma, Dawei Yin, Jianping Wang, Jiliang Tang, and Qing Li. 2019. Deep Social Collaborative Filtering. In *Thirteenth ACM Conference on Recommender Systems (RecSys '19), September 16–20, 2019, Copenhagen, Denmark*. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3298689.3347011>

1 INTRODUCTION

Recommender systems play a crucial role to alleviate the information overload in the era of information explosion. Collaborative filtering is one of the most popular techniques to build modern recommender systems, which models users' preference towards items by utilizing the history of user-item interactions such as ratings [27]. In addition to the user-item interactions, social relations between users provide another stream of potential information of users' preference. As argued in social theories, people in social networks are influenced by their social connections, which leads to the homophily phenomenon of similar preference in social neighbors [3, 8, 10, 23]. More specifically, information diffuses through social interactions and users tend to acquire and disseminate information through social networks. Thus, social relations can play an important role in describing the preferences of users, which, in turn, can help build good recommender systems. In fact, social relations have been shown to boost the performance of recommender systems [8, 10, 16, 32].

Recent years have witnessed the great success of deep neural networks on various areas such as computer vision (CV) [39], speech recognition [15] and Natural Language Processing (NLP) [17]. It is not surprising that deep neural networks are adopted to enhance recommender systems. Some recent proposed recommender systems facilitate deep neural networks as feature learning tools to extract useful features from auxiliary information such as text description of items [5, 18, 37], audio of music [34, 42] and visual information of images [45], while others [14] try to utilize deep neural networks to capture the non-linearity between user-item interactions. There are some recent works utilizing deep neural networks for social recommendations [6, 8–10, 40]. For example, GraphRec [10] proposes a graph neural networks

framework for social recommendation, which aggregates both user-item interactions information and social interaction information when performing prediction; DASO [8] harnesses the power of adversarial learning to dynamically generate "difficult" negative samples, learn the bidirectional mappings between the social domain and item domain.

Although the aforementioned deep social recommender systems facilitate the social network information to enhance the recommendation performance, they do not fully take advantage of social network information. First, most of them only involve direct neighbors, while information from users that are a few hops away could also be helpful [31, 32]. The reasons are as follows: 1) information is diffusing through the social network and users might be affected by indirect neighbors; and 2) users might refer to distant neighbors (or weak ties) when the direct neighbors cannot share useful information. Therefore, it is desired to consider the distant social relations for recommender systems. Second, most of the aforementioned methods treat neighbors' information equally for all recommendation cases. However, not all information from neighbors are useful when the recommender system is performing a specific recommendation. For example, when predicting whether a user will purchase an iPhone X, the interactions between his/her friends and iPhone X or other iPhone related items might be helpful while the interactions between his/her friends and Nike shoes might not be relevant. Therefore, it is necessary to filter information from neighbors. Finally, most of the deep social recommender systems do not consider the users' opinions towards items, which are usually expressed in the forms of reviews or ratings. It is obvious that bad and good opinions from a user's friends would affect the user's decision in tremendously different ways. Hence, it is desired to carefully consider the opinions of user-item interactions.

While it is of great potential to sufficiently exploit the social network information for recommendations, it faces tremendous challenges. First, the social interactions in distant social relations are complex and it is difficult to properly extract helpful information for recommendations. Second, it is not trivial to select relevant information from neighbors, as they could have interactions with many different items. Finally, it is challenging to capture the user's opinions while modeling the user-item interactions. In this paper, to tackle the aforementioned challenges, we propose a deep social collaborative filtering framework DSCF, which can sufficiently exploit the social network information for recommendations. Our contributions can be summarized as follows:

- We propose a principle way based on deep neural networks to extract helpful information from distant social relations for recommendations;
- We introduce a novel way to capture user's opinions while modeling user-item interactions;
- We propose a deep social collaborative filtering framework which can sufficiently exploit social network information for recommendations; and
- We conduct comprehensive experiments on two real-world datasets to show the effectiveness of the proposed framework.

The remainder of this paper is organized as follows. We introduce the proposed framework in Section 2. In Section 3, we conduct

experiments on two real-work datasets to illustrate the effectiveness of the proposed method. In Section 4, we review work related to our framework. Finally, we conclude our work with future directions in Section 5.

2 THE PROPOSED FRAMEWORK

In this section, we introduce the proposed deep social collaborative filtering framework **DSCF**. As discussed earlier, to exploit social networks for recommendations, we need to (a) consider information from not only direct neighbors but also distant neighbors; (b) select relevant information of each neighbor for recommending a specific item; and (c) capture neighbor's opinions towards items when modeling user-item interactions. An overview of the proposed framework is demonstrated in Figure 1. It consists of four layers – the random walk layer that is designed for addressing challenges (a) and (b), the embedding layer that is designed for solving the challenge (c), the sequence learning layer and the output layer. Next we will give details of each layer.

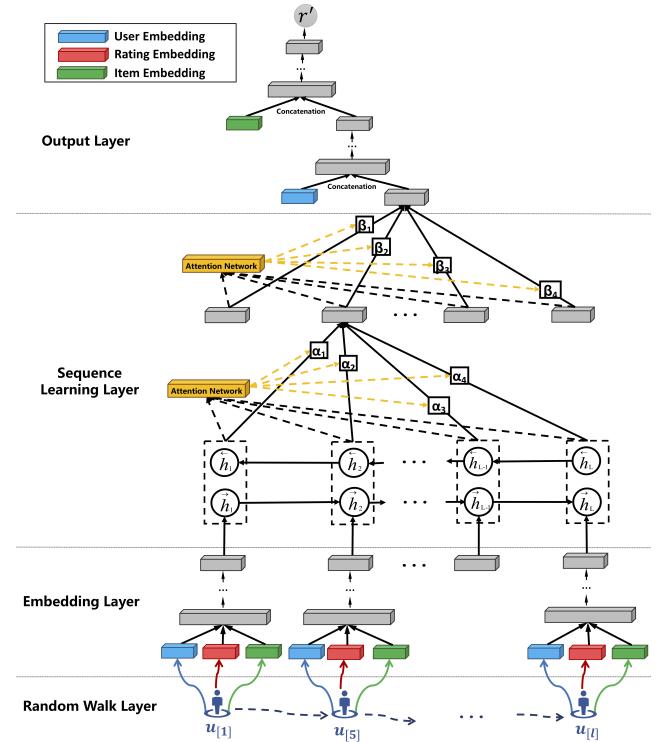


Figure 1: An overview of the proposed framework.

Before introducing the details of each layer, we first introduce definitions and notations that are used throughout the paper. Let $\mathcal{U} = \{u_1, u_2, \dots, u_N\}$ and $\mathcal{V} = \{v_1, v_2, \dots, v_M\}$ denote the sets of users and items respectively, where N is the number of users, and M is the number of items. Let $R \in \mathbb{R}^{N \times M}$ be the rating matrix (or the user-item interaction matrix), where the i, j -th element $r_{i,j}$ is the rating score of item v_j given by user u_i . If the user u_i has not rated the item v_j , then $r_{i,j}$ is set to 0, which means the rating is unknown. The social network between users can be described by a matrix

$T \in \mathbb{R}^{N \times N}$, where $T_{i,j} = 1$ if there is a social relation between user u_i and user u_j , otherwise 0. Given the rating matrix R and the social network T , we aim to predict the unknown ratings in R . As in the traditional collaborative filtering methods, we embed users and items to low-dimensional latent vectors. The embeddings for user u_i and item v_j are denoted as $p_{[i]} \in \mathbb{R}^d$, $q_{[j]} \in \mathbb{R}^d$ respectively, where d is the length of the embedding.

2.1 The Random walk layer: generating item-aware social sequences

In social recommendation, when we try to perform recommendation for a given user u , not only his/her direct neighbors can provide useful information, but also his/her distant neighbors that are within a few hops (or neighbors in his/her local neighborhood) can help. Furthermore, neighbors with different distance to the user u are likely to be of different importance for the recommendation. Thus, it is also necessary to differentiate neighbors of u according to their distance to user u when including them for recommendations. Random walk is a popular tool to explore the local neighborhood of networks [20, 30]. Additionally, random walk explores the neighborhood in the form of node sequences (user sequences) [12, 24], which naturally maintains the order of neighbors according to the distance to the user u . Thus, we can effectively utilize random walk to generate distant user sequences from social networks. More specifically, the user sequence can be generated by a random walk starting from user u and ending after l steps, where l is the length of the random walk. The generated user sequence can be denoted as $S_{(i)}^u = \{u_{[1]}, \dots, u_{[l]}\}$, where the subscript (i) indicates $S_{(i)}^u$ is the i -th user sequence generated for user u as we need to generate multiple user sequences to sufficiently explore the neighborhood of u and $[k]$ means that the user $u_{[k]}$ is the k -th user in the user sequence.

While the user sequences contain the information of neighbors, they are not specified for a given recommendation case, i.e., predicting preference of user u on the item v , as such information is shared by all the recommendation cases involving user u . However, not all information from the neighbors is helpful for recommending the specific item v . Only that information related to this item v would be useful. Thus, we need to select an item related to item v for each user in the generated user sequences and form an item-aware social sequence, denoted as $S_{(i)}^{u,v} = \{(u_{[1]}, v_{[1]}), \dots, (u_{[l]}, v_{[l]})\}$. Note that only the most relevant item is exploited for a specific recommendation case. The reasons are two-fold. First, the most relevant item is most important to affect the decision making of a target item (item v), while other items may not be helpful since they may bring in noise. Second, multiple user sequences are generated by the random walk process to sufficiently explore different relevant items for a specific recommendation case, which, in turn, can help form these item-aware social sequences. More specifically, for each user $u_{[k]}$ in one user sequence, we choose the item $v_{[k]}$ from the set of items that have been interacted with user $u_{[k]}$ as:

$$v_{[k]} = \operatorname{argmax}_{v_h \in \mathcal{V}_{u_{[k]}}} \sim(v_h, v) \quad \text{Fah Liq 11}$$

where $\mathcal{V}_{u_{[k]}}$ denotes the set of items interacted with user $u_{[k]}$ and $\sim(v_h, v)$ is a function to measure the similarity between item v_h and item v . In this paper, we empirically select cosine similarity as

follows

$$\sim(v_h, v) = \frac{\mathbf{x}_h^T \mathbf{x}_v}{|\mathbf{x}_h| |\mathbf{x}_v|}, \quad (2)$$

$$\mathbf{x}_m = f(v_m). \quad (3)$$

where f function is to generate appropriate features \mathbf{x}_m for item v_m . Different features sources, such as the textual descriptions, the visual content of images and the user-item interactions, could be used to represent the items. In this paper, we adopt the user-item interactions to represent the items since the auxiliary information such as textual descriptions and visual content is not available. More specifically, we use the item embeddings learned by NeuMF [14] as the item features to measure similarity between items.

The set of all item-aware social sequences generated for predicting the rating of (u, v) is $\mathcal{S}_{(i)}^{u,v} = \{S_{(i)}^{u,v}\}_{i=1}^H$, where H is the number of social sequences generated for this recommendation case.

The advantages of the item-aware social sequences for predicting interaction between users u and items v are twofolds. First, the social sequences contain not only direct neighbors but also distant neighbors. Second, these sequences are specific for the recommendation from u to v . An illustration example of the process of generating item-aware social sequence is shown in Figure 2. We are predicting the rating of user u_1 to item v_3 (Spider-Man). As shown in figure, starting from source user u_1 , we perform our random walk on the direct neighbors. The random walk is employed to generate possible user sequence, denoted as $S_{(1)}^{u_1} = \{u_{[2]}, u_{[3]}, u_{[6]}, u_{[7]}\}$. For each user in the user sequence, we need to collect the most similar item to v_3 . The generated item-aware social sequence is $S_{(1)}^{u_1, v_3} = \{(u_{[2]}, v_{[3]}), (u_{[3]}, v_{[5]}), (u_{[6]}, v_{[5]}), (u_{[7]}, v_{[3]})\}$. To prevent clutter, here, we suppose that item v_5 (Captain America) is the most similar to the item v_3 (Spider-Man) in our example, and the length of random walk is 4.

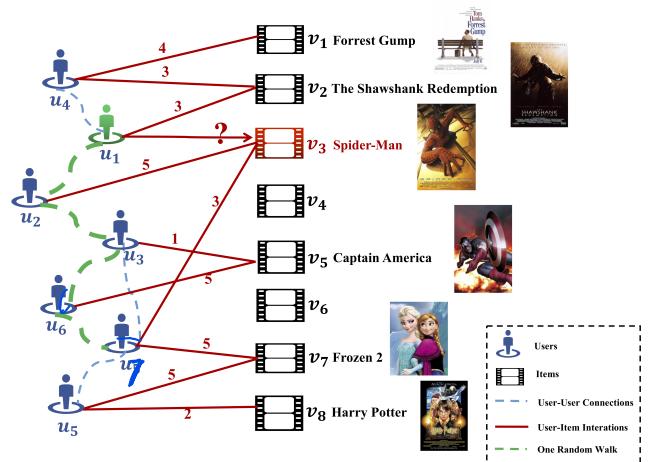


Figure 2: An illustration example of generating item-aware social sequences. Note that the number on the edges of user-item interactions denotes the opinions (or rating score) of users on the items via the interactions. There are 5 different rating levels.

2.2 The embedding layer: modeling user-item interactions

The item-aware sequences consist of user-item interactions from the user's neighbors, hence, we need to first model the user-item interactions. When modeling the user-item interactions, it is important to carefully consider the opinions the users expressed on the interactions. Obviously, bad and good opinions from the user's social neighbors can affect the user's opinion towards the item in tremendously different ways. Thus, we propose to include the user's opinion towards the item when modeling the user-item interaction in the sequence. The opinions are usually expressed in the form of ratings. For example, as shown in Figure 2, both user u_3 and user u_6 interact with the same item v_5 (Captain America); however, user u_6 likes v_5 while user u_3 dislikes v_5 .

To model the ratings, we propose to embed each discrete rating value into a rating embedding vector. Therefore, if there are I different rating levels, there would be I rating embedding vectors. Note that the rating embeddings are also parameters of the framework. The rating embedding of the rating value o is denoted as $\mathbf{r}_{\{o\}} \in \mathbb{R}^d$, with d the embedding length. For an interaction $(u_{[k]}, v_{[k]})$ in the item-aware social sequence $S_{(i)}^{u,v}$, the non-zero rating score of this interaction can be found in the rating matrix \mathbf{R} and let us denote it as $o_{u_{[k]}, v_{[k]}}$. Then the corresponding rating embedding is $\mathbf{r}_{\{o_{u_{[k]}, v_{[k]}}\}}$, which we denote as $\mathbf{r}_{[k]}$ for convenience. The interaction between user and item is highly non-linear, and including the rating information further adds the complexity. Hence, we use a multi-layer perception (MLP) to fuse the interaction information with the rating information. The MLP takes the concatenation of user embedding $\mathbf{p}_{[k]}$, rating embedding $\mathbf{r}_{[k]}$, item embedding $\mathbf{q}_{[k]}$ as input and output the user-item interaction embedding $\mathbf{e}_{[k]}$ of interaction $(u_{[k]}, v_{[k]})$. The procedure can be briefly represented as follows

$$\mathbf{e}_{[k]} = g_{u,r,v}([\mathbf{p}_{[k]}, \mathbf{r}_{[k]}, \mathbf{q}_{[k]}]) \quad (4)$$

where $[\mathbf{p}_{[k]}, \mathbf{r}_{[k]}, \mathbf{q}_{[k]}]$ denotes the concatenation of $\mathbf{p}_{[k]}, \mathbf{r}_{[k]}, \mathbf{q}_{[k]}$.

Following this procedure, we process each sequence $S_{(i)}^{u,v} = \{(u_{[1]}, v_{[1]}), \dots, (u_{[l]}, v_{[l]})\}$ and get a sequence of fused interaction embedding $E_{(i)}^{u,v} = \{\mathbf{e}_{[1]}, \dots, \mathbf{e}_{[l]}\}$. The set of all sequences of fused interaction embedding from neighbors for predicting the rating of (u, v) can be denoted as $\mathcal{E}^{u,v}$.

2.3 The sequence learning layer: learning representation for item-aware social sequences

After generating the item-aware social sequences for (u, v) and transforming each user-item interaction with opinions information in the sequences to fused interaction embedding, we proceed to the sequence learning layer. The sequence learning layer aims to extract features for each sequence and then combine the extracted features of all the sequences to obtain a unified representation, which can be used to predict the rating for (u, v) in the output layer.

As all the neighbors in the sequence would affect the prediction of (u, v) , for distant neighbors, we need to capture the distant social information between them and the user u . Furthermore, in social networks, users influence each other. Hence, we need

to capture the bi-directional influence in the model. Recently, a bi-directional long short-term memory network (Bi-LSTM) based language model [1, 44] has been proposed to capture the long-range bi-directional semantic dependencies between words in sentence in NLP domain. Inspired by these model, we regard the sequence as a sentence and elements in this sequence as "words" and adopt a similar Bi-LSTM model to extract features from the sequence of fused interaction embeddings. The bi-directional LSTM contains the forward LSTM \overrightarrow{LSTM} which reads the sequence $E_{(i)}^{u,v}$ from $\mathbf{e}_{[1]}$ to $\mathbf{e}_{[l]}$, and a backward LSTM \overleftarrow{LSTM} which reads from $\mathbf{e}_{[l]}$ to $\mathbf{e}_{[1]}$.

$$\overrightarrow{\mathbf{h}}_{[k]}^{(i)} = \overrightarrow{LSTM}(E_{(i)}^{u,v}), k \in [1, l], \quad (5)$$

$$\overleftarrow{\mathbf{h}}_{[k]}^{(i)} = \overleftarrow{LSTM}(E_{(i)}^{u,v}), k \in [l, 1]. \quad (6)$$

where $\overrightarrow{\mathbf{h}}_{[k]}^{(i)}$ and $\overleftarrow{\mathbf{h}}_{[k]}^{(i)}$ are hidden states of $\overrightarrow{LSTM}, \overleftarrow{LSTM}$, respectively.

These hidden states, which are corresponding to the neighbors in the sequence, are then combined using an attention mechanism [10, 35, 36] to generate the features $\mathbf{s}_{(i)}^{u,v}$ of the sequence $E_{(i)}^{u,v}$.

$$\mathbf{s}_{(i)}^{u,v} = \sum_{k=1}^l \alpha_k \overrightarrow{\mathbf{h}}_{[k]}^{(i)}. \quad (7)$$

where $\mathbf{h}_{[k]}^{(i)}$ is $[\overrightarrow{\mathbf{h}}_{[k]}^{(i)} \oplus \overleftarrow{\mathbf{h}}_{[k]}^{(i)}]$, the concatenation of $\overrightarrow{\mathbf{h}}_{[k]}^{(i)}$ and $\overleftarrow{\mathbf{h}}_{[k]}^{(i)}$. Specially, we parameterize the attention weight α_k with one-layer network, and extract these user (neighbor)-item interaction embeddings that are important to learn representation for the item-aware social sequence. The normalized importance weight α_k is calculated through a Softmax function follows

$$\alpha_k = \tanh(\mathbf{W}_a \cdot \mathbf{h}_{[k]} + \mathbf{b}_a), \quad (8)$$

$$\alpha_k = \frac{\exp(\mathbf{a}_k^T \mathbf{a}_u)}{\sum_j \exp(\mathbf{a}_j^T \mathbf{a}_u)}. \quad (9)$$

where the neighbor-level context vector \mathbf{a}_u can be seen as a high level representation of a fixed query "what is the informative neighbor-item interaction embedding?" over all the neighbor-item interaction embeddings in the item-aware social sequence. Note that the neighbor-level context vector \mathbf{a}_u is parameters in the framework and needs to be jointly learned during the training process.

We then combine the representations of all the user-item interaction embedding sequences to generate the unified representation of item-aware social sequences for (u, v) as

$$\mathbf{s}^{u,v} = \sum_{i=1}^H \beta_i \mathbf{s}_{(i)}^{u,v} \quad (10)$$

where we adopt an attention mechanism to differentiate the importance weight β_i of item-aware social sequences as follows

$$\mathbf{z}_i = \tanh(\mathbf{W}_z \cdot \mathbf{s}_{(i)}^{u,v} + \mathbf{b}_z), \quad (11)$$

$$\beta_i = \frac{\exp(\mathbf{z}_i^T \mathbf{z}_u)}{\sum_j \exp(\mathbf{z}_j^T \mathbf{z}_u)}. \quad (12)$$

Similar to eq. (9), \mathbf{z}_u can be seen as a high level representation query “which is the informative item-aware social sequence?” over all the social sequences. The reason why we introduce two attentions is that not all user-item interaction with opinions information in one item-aware social sequence contribute equally to the representation of this item-aware social sequence; and not all these sequences contribute equally to the unified representation of item-aware social sequences for (u, v) .

2.4 The output layer: rating prediction

In the output layer, we will design recommendation tasks to learn model parameters. There are various recommendation tasks such as item ranking and rating predation. In this work, we apply the proposed DSCF model for the recommendation task of rating prediction. We finally make the prediction of rating score of the user u to item v . The input of the output layer includes the user embedding $\mathbf{p}_{[u]}$, the item embedding $\mathbf{q}_{[v]}$ and the unified item-aware social representations $\mathbf{s}^{u,v}$ learned in the sequence learning layer. As shown in the output layer in Figure 1, a multi-layer perception (MLP) is first used to combine the user embedding $\mathbf{p}_{[u]}$ and the unified item-aware social representations $\mathbf{s}^{u,v}$. Let us denote this MLP as $f_{u,s}$. Then, another MLP, which can be denoted as $f_{u,v}$, is used to predict the rating score of (u, v) . The prediction procedure, which takes $\mathbf{p}_{[u]}, \mathbf{q}_{[v]}, \mathbf{s}^{u,v}$ as input, can be represented as

$$r'_{u,v} = f_{u,v}([\mathbf{q}_{[v]}, f_{u,s}([\mathbf{p}_{[u]}, \mathbf{s}^{u,v}])]), \quad (13)$$

where $[,]$ denotes the concatenation operation, and $r'_{u,v}$ is the predicted rating from user u to item v .

2.5 Model Training

To estimate parameters of the framework DSCF, we need to specify an objective function to optimize. Since the task we focus on in this work is rating prediction, a commonly used objective function is formulated as,

$$\text{Loss} = \frac{1}{2|O|} \sum_{(u_i, v_j) \in O} (r'_{i,j} - r_{i,j})^2 \quad (14)$$

where O denotes all the observed user-item interactions, $|O|$ is the number of interactions in O , and $r'_{i,j}$ is the predicted rating while $r_{i,j}$ is the ground truth rating assigned by the user u_i on the item v_j .

To optimize the objective function, we adopt the Adaptive Moment Estimation (Adam) [7] as the optimizer in our implementation. We also adopt the dropout strategy [28] to alleviate the overfitting issue in optimizing deep neural network models.

There are three embedding in our model, including item embedding \mathbf{q}_i , user embedding \mathbf{p}_i , and rating embedding \mathbf{r}_o . They are randomly initialized and jointly learned during the training stage. We do not use one-hot vectors to represent each user and item, since the raw features are very large and highly sparse. By embedding high-dimensional sparse features into a low-dimensional latent space, the model can be easy to train [14, 38]. Rating embedding matrix \mathbf{r} depends on the rating scale of the system. For example, for a 5-star rating system, rating embedding matrix \mathbf{r} contains 5 different embedding vectors to denote scores in $\{1, 2, 3, 4, 5\}$.

3 EXPERIMENTS

In this section, we conduct experiments to verify the effectiveness of our model. We first introduce the experimental settings, then discuss the results of the performance comparison of various recommender systems, and finally study the impact of different components in our model.

3.1 Experimental Settings

3.1.1 Datasets. In our experiments, two representative datasets Ciao and Epinions¹ are utilized to verify the effectiveness of our model. They are taken from the product review sites Ciao (www.ciao.co.uk) and Epinions (www.epinions.com). Each site allows users to rate items, and add friends to their Circle of Trust. Therefore, they provide a large amount of rating information and social information. The rating scale is from 1 to 5. We randomly initialize rating embedding with 5 different embedding vectors based on 5 scores in $\{1, 2, 3, 4, 5\}$. The statistics of these two datasets are presented in Table 1.

Table 1: Statistics of the datasets

Dataset	Ciao	Epinions
# of Users	7,317	18,088
# of Items	104,975	261,649
# of Ratings	283,319	764,352
# of Density(Ratings)	0.0368%	0.0161%
# of Social Connections	111,781	355,813
# of Density(Social Relations)	0.2087%	0.1087%

3.1.2 Evaluation Metrics. In order to evaluate the quality of the recommendation algorithms, two popular metrics are adopted to evaluate the predictive accuracy, namely Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) [10]. Smaller values of MAE and RMSE indicate better predictive accuracy. Note that small improvement in RMSE or MAE terms can have a significant impact on the quality of the top-few recommendations [19].

3.1.3 Baselines. To evaluate the performance, we compared DSCF with three groups of methods including traditional recommender systems, traditional social recommender systems, and deep neural network based recommender systems. For each group, we select representative baselines and below we detail them.

- **PMF** [26]: Probabilistic Matrix Factorization utilizes user-item rating matrix only and models latent factors of users and items by Gaussian distributions.
- **SoRec** [21]: Social Recommendation performs co-factorization on the user-item rating matrix and user-user social relations matrix.
- **SoReg** [22]: Social Regularization models social network information as regularization terms to constrain the matrix factorization framework.
- **SocialMF** [16]: It considers the trust information and propagation of trust information into the matrix factorization model for recommender systems.

¹ Both Ciao and Epinions datasets are available at: <https://www.cse.msu.edu/~tangjili/trust.html>

- **TrustMF** [43]: This method adopts matrix factorization technique that maps users into two low-dimensional spaces: trustee space and trustee space, by factorizing trust networks according to the directional property of trust.
- **NeuMF** [14]: This method is a state-of-the-art matrix factorization model with neural network architecture. The original implementation is for recommendation ranking task and we adjust its loss to the squared loss for rating prediction.
- **DeepSoR** [9]: This model employs deep learning to learn representations of each user from social relations, and to integrate them into probabilistic matrix factorization for rating prediction.
- **GCMC+SN** [4]: This model is a state-of-the-art recommender system with graph neural network architecture. In order to incorporate social network information into GCMC, we utilize the *node2vec* [12] to generate user embedding as user side information, instead of using the raw feature social connections ($T \in \mathbb{R}^{n \times n}$) directly. The reason is that the raw feature input vectors is highly sparse and high-dimensional. Using the network embedding techniques can help compress the raw input feature vector to a low-dimensional and dense vector, then the model can be easy to train.

PMF and NeuMF are pure collaborative filtering model without social information for rating prediction, while the others are social recommendations. Besides, we compared DSCF with two state-of-the-art neural network based social recommender systems, i.e., DeepSoR, and CGMC+SN.

3.1.4 Parameter Settings. We implemented our proposed model in Pytorch². For each dataset, we used $x\%$ as a training set to learning parameters, $(1 - x\%)/2$ as a validation set to tune hyper-parameters, and $(1 - x\%)/2$ as a testing set for the final performance comparison, where x was varied as $\{80\%, 60\%\}$ [10]. For the embedding size d , we tested the value of $\{8, 16, 32, 64, 128, 256\}$. The batch size and learning rate were searched in $\{16, 32, 64, 128, 512\}$ and $\{0.0005, 0.001, 0.005, 0.01, 0.05, 0.1\}$, respectively. Moreover, we empirically set the size of the hidden layer the same as the embedding size (the dimension of the latent factor) and the activation function as ReLU. Without special mention, we employed three hidden layers for all the neural components. The early stopping strategy was performed, where we stopped training if the RMSE on validation set increased for 5 successive epochs. The parameters for the baseline algorithms were initialized as suggested in the corresponding papers, and were then carefully tuned to achieve optimal performance.

3.2 Performance Comparison

We first compare the recommendation performance of all methods. Table 2 shows the overall rating prediction error w.r.t. RMSE and MAE among the recommendation methods on Ciao and Epinions datasets, respectively. We have the following findings:

- SoRec, SoReg, SocialMF and TrustMF improve over PMF. All of these methods are based on matrix factorization. SoRec, SoReg, SocialMF and TrustMF leverage both the user-item interactions and social information; while PMF only

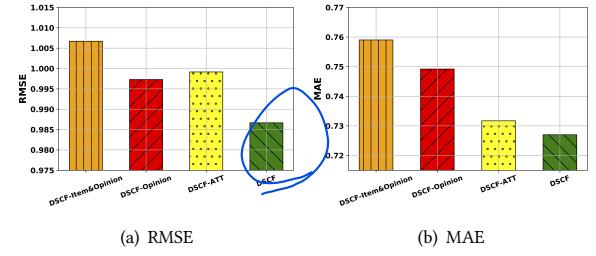


Figure 3: Component Analysis on Ciao dataset. DSCF-* means the component * is removed in DSCF.

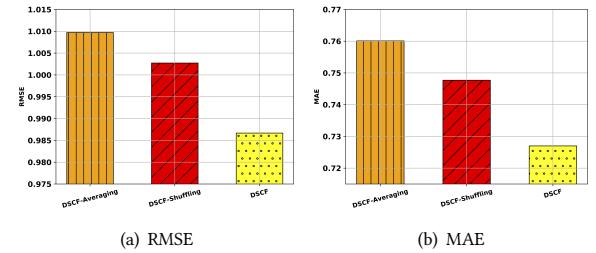


Figure 4: Effect of Bi-LSTM model on Ciao dataset.

utilizes user-item interactions. These improvements show the effectiveness of incorporating social information for recommender systems.

- NeuMF achieves much better performance than PMF. Both of them utilize the user-item interactions only. NeuMF is based on deep architecture; while PMF is a traditional method with shallow architecture. This suggests the power of employing deep architecture on the task of recommendation.
- Two deep models, DeepSoR and GCMC+SN, obtain better performance than SoRec, SoReg, SocialMF, and TrustMF, which are based on matrix factorization with shallow architecture. These improvements further reflect the power of employing deep architecture on the task of recommendation.
- DSCF outperforms NeuMF. This result further supports that social information is complementary to user-item interactions for recommendation.
- Our model DSCF consistently outperforms all the baseline methods. Compared with DeepSoR and GCMC+SN, our model proposes advanced model components to integrate user-item interactions and social information. In addition, our model introduces ways to capture user's opinions while modeling user-item interactions. We will provide further investigations to better understand the contributions of model components to the proposed framework in the following subsection.

3.3 Model Component Analysis

In the previous subsection, we have demonstrated the effectiveness of the proposed framework. To deeply understand DSCF,

²<https://pytorch.org/>

Table 2: Performance comparison of different recommender systems

Training	Metrics	Algorithms								
		PMF	SoRec	SoReg	SocialMF	TrustMF	NeuMF	DeepSoR	GCMC+SN	DSCF
Ciao (60%)	MAE	0.952	0.8489	0.8987	0.8353	0.7681	0.8251	0.7813	0.7697	0.7501
	RMSE	1.1967	1.0738	1.0947	1.0592	1.0543	1.0824	1.0437	1.0221	1.0157
Ciao (80%)	MAE	0.9021	0.8410	0.8611	0.8270	0.7690	0.8062	0.7739	0.7526	0.7270
	RMSE	1.1238	1.0652	1.0848	1.0501	1.0479	1.0617	1.0316	0.9931	0.9867
Epinions (60%)	MAE	1.0211	0.9086	0.9412	0.8965	0.8550	0.9097	0.8520	0.8602	0.8427
	RMSE	1.2739	1.1563	1.1936	1.1410	1.1505	1.1645	1.1135	1.1004	1.0999
Epinions (80%)	MAE	0.9952	0.8961	0.9119	0.8837	0.8410	0.9072	0.8383	0.8590	0.8275
	RMSE	1.2128	1.1437	1.1703	1.1328	1.1395	1.1476	1.0972	1.0711	1.0667

we compare it with three variants, i.e., DSCF-Opinion, DSCF-Item&Opinion, DSCF-ATT, DSCF-Averaging and DSCF-Shuffling, which are defined as follows:

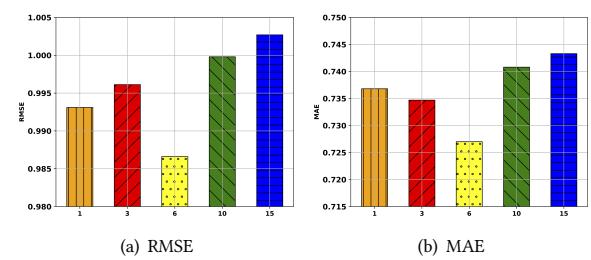
- DSCF-Opinion: This variant uses the item-aware social sequences to represent user's social information; while ignoring the opinions on the user-item interaction.
- DSCF-Item&Opinion: Based on DSCF-Opinion, it further eliminates the associated items in the social sequence.
- DSCF-ATT: This variant is to study the impact of attention mechanisms on learning $s_{(i)}^{u,v}$ and $s^{u,v}$. The attention mechanisms α and β are removed in this variant.
- DSCF-Averaging: This variant replaces Bi-LSTM with averaging the elements in the input of the sequence in the sequence learning layer.
- DSCF-Shuffling: This variant randomly shuffles the order of elements in the sequence in the sequence learning layer.

The variant DSCF-Averaging considers that all users in the sequence have the same influence to the target user; while DSCF-Shuffling assumes that the influence is not related to the distance to the target user. These two variants are designed to understand the benefit of adapting Bi-LSTM to capture the item-aware social sequences.

The results on Ciao are given in Figure 3 and Figure 4. We do not show the results on Epinions since similar observations can be made. From the results, we have the following findings:

Item-aware Social Sequences with Opinions. We now focus on analyzing the effectiveness of opinions on interactions. From the Figure 3, we can see that the performance of DSCF reduces significantly when ignoring the opinions on the user-item interactions in the social sequence (i.e., DSCF-Opinion), which suggests that it is necessary to consider opinions on interactions. In other words, different opinions from a user's friends would affect the user's decision in tremendously different ways.

Item-aware Social Sequences. To recommend a specific item, not all information from users in the sequence is useful; in other words, interactions of these users with related items are more useful. From the results in Figure 3, DSCF-Item&Opinion performs worse than DSCF and DSCF-Opinion. These observations support the importance to generate item-aware sequences. In other words, not all information from neighbors are useful for recommending a

**Figure 5: Performances w.r.t. the length of sequence.**

specific item (e.g., *Spider-man*). Only the information related to this item would be useful (e.g., *Captain America*).

Attention Mechanisms. We conducted experiments to verify the effectiveness of the attention mechanism. From the results in Figure 3, we can observe that DSCF-ATT obtains worse performance than DSCF. The reason is that not all the user (neighbor)-item interactions in one social sequence contribute equally to learn the representation of item-aware social sequence; and not all these item-aware social sequences have the same importance to the unified representation of all item-aware social sequences. These results demonstrate the benefits of the attention mechanisms on learning $s_{(i)}^{u,v}$ and $s^{u,v}$.

Bi-LSTM. Figure 4 presents the effect of Bi-LSTM on Ciao dataset. The performance of both DSCF-Averaging and DSCF-Shuffling reduces significantly. It suggests that the Bi-LSTM component is better to learn representations for item-aware social sequences. The reason is that the social sequence reflects the information diffusion to the target user and the influence to the target user should be heterogeneous and related to the distance.

3.4 Parameter Analysis

There are two important parameters of the proposed framework, i.e., the length of each item-aware social sequence and the number of item-aware social sequences. In this subsection, we investigate the impact of these parameters by examining how the performance changes when varying one parameter and fixing others. Similarly, we only show results on Ciao.

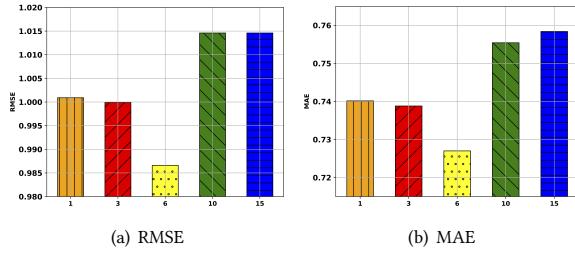


Figure 6: Performance w.r.t the number of sequences.

Effect of the length of sequences *l*. Figure 5 shows the performance with the varied length of sequences on Ciao. If the length of sequence is one, our model boils down to use the direct neighbors. When the length of sequence increases, the performance tends to increase first. This indicates that the direct neighbors cannot sufficiently capture the useful social information and including distant neighbors could help. However, when the length of sequences becomes too large, the performance degrades as we may introduce too many noises with the distant neighbors.

Effect of the number of sequences *H*. Figure 6 shows how the number of sequences affects the performance of recommendations. Generally more sequences can sufficiently explore the neighborhood of users, which can help us understand social information better; however, it is also risky to generate too many since we may introduce noise as well.

4 RELATED WORK

In this section, we briefly review some researches related to our work. Collaborative filtering [11], which captures users' preference towards items utilizing user-item interactions, is the most popular approach to build modern recommender systems. In addition to the user-item interactions, social relations also have potential to help understand users' preference. Many social recommendation methods [13, 21, 25, 31, 33, 41, 46] have shown the effectiveness of including social relations for recommendations. Among them, SoRec [21] co-factorizes the rating matrix (user-item interaction matrix) and the social relation matrix for recommendation by sharing user latent vectors between them. SoDimRec [33] utilizes the heterogeneity of social relations and the weak dependency connections in social networks for recommendation. A comprehensive survey on social recommendations can be found in [32].

Recently, deep neural networks have been adopted to enhance recommender systems [2, 47]. Most of them utilize deep neural networks as feature learning tools to extract features from auxiliary information such as text description of an item [5, 18, 37] and visual information of images [45]. NeuMF [14], is a matrix factorization based deep recommendation method, which uses deep neural networks to explore the non-linearity in user-item interactions. NSCR [40] extends the NeuMF model by utilizing the social network information as a graph regularization, which enforces nearby neighbors to have similar latent vectors. NSCR addresses the task of cross-domain recommendations for ranking metric, and focuses

on how to distill useful signal from an external social network (e.g., Facebook and Twitter) on the cross-domain task, while our model focuses on how to learn the social information from the user-user interaction in the same e-commerce platform, rather than external social network. ARSE [29] proposes the problem of temporal social recommendation for ranking metric, which has dynamic and static part to model the dynamic and static preferences of users. ARSE targets on the dynamic preferences of the recommendation, rather than the social information.

Most related to our task with neural networks includes DLMF [6], GCMC [4], DeepSoR [9], GraphRec [10] and DASO [8]. DeepSoR [9] first represents users using pre-trained node embedding technique, and further utilizes deep neural networks to capture non-linear features in social relations and integrate them into probabilistic matrix factorization. DASO [8] proposes a deep adversarial social recommendation framework, which adopts a bidirectional mapping method to transfer users' information between social domain and item domain using adversarial learning. GraphRec [10] harness the power of graph neural networks (GNNs) techniques to model graph data in social recommendations by aggregating the both user-item interactions information and direct social neighbors. However, these deep social recommendation methods cannot take full advantages of social networks. In this paper, we propose a deep social recommendation framework which can sufficiently exploit the social network information for recommendations.

5 CONCLUSION AND FUTURE WORK

We have presented a Deep Social Collaborative Filtering (DSCE) which can exploit the social information with various aspects for recommendations. Particularly, we propose to utilize the random walk to generate item-aware social sequences, which consider information from not only direct neighbors but also distant neighbors. In addition, we also introduce a novel way to capture neighbors' opinions when modeling user-item interactions. Finally, the Bi-LSTM with attention mechanism is proposed to extract feature for the social sequence. Our experiments reveal that the item-aware sequences and the opinion information play a crucial role in modeling social information. Comprehensive experiments on two real-world datasets show the effectiveness of our model. In this work, we only utilize the user-item interactions to measure the similarity between items, while rich side information may be associated with items, such as the textual description, and the visual content of images. Therefore, incorporating side information would be considered as an interesting future direction.

ACKNOWLEDGMENTS

The work is partly supported by NSFC-Guangdong Joint Fund under project U1501254, Science Technology and Innovation Committee of Shenzhen Municipality Under project JCYJ20170818095109386, and an internal research grant (project no. 9B0V) from the Hong Kong Polytechnic University. Yao Ma and Jiliang Tang are supported by the National Science Foundation (NSF) under grant numbers IIS-1714741, IIS-1715940, IIS-1845081 and CNS-1815636, and a grant from Criteo Faculty Research Award.

REFERENCES

- [1] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural Machine Translation by Jointly Learning to Align and Translate. In *International Conference on Learning Representations (ICLR)*.
- [2] Ting Bai, Ji-Rong Wen, Jun Zhang, and Wayne Xin Zhao. 2017. A neural collaborative filtering model with interaction-based neighborhood. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. ACM, 1979–1982.
- [3] Eytan Bakshy, Itamar Rosenn, Cameron Marlow, and Lada Adamic. 2012. The role of social networks in information diffusion. In *World Wide Web conference*. 519–528.
- [4] Rianne van den Berg, Thomas N Kipf, and Max Welling. 2017. Graph convolutional matrix completion. *arXiv preprint arXiv:1706.02263* (2017).
- [5] Chong Chen, Min Zhang, Yiqun Liu, and Shaoping Ma. 2018. Neural attentional rating regression with review-level explanations. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 1583–1592.
- [6] Shuguang Deng, Longtao Huang, Guadong Xu, Xindong Wu, and Zhaohui Wu. 2017. On deep learning for trust-aware recommendations in social networks. *IEEE transactions on neural networks and learning systems* 28, 5 (2017), 1164–1177.
- [7] Jimmy Ba Diederik P. Kingma. 2015. Adam: A Method for Stochastic Optimization. In *International Conference on Learning Representations (ICLR)*.
- [8] Wenqi Fan, Tyler Derr, Yao Ma, Jianping Wang, Jiliang Tang, and Qing Li. 2019. Deep Adversarial Social Recommendation. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence, IJCAI-19*.
- [9] Wenqi Fan, Qing Li, and Min Cheng. 2018. Deep Modeling of Social Relations for Recommendation. In *AAAI*.
- [10] Wenqi Fan, Yao Ma, Qing Li, Yuan He, Eric Zhao, Jiliang Tang, and Dawei Yin. 2019. Graph Neural Networks for Social Recommendation. In *The World Wide Web Conference*. ACM, 417–426.
- [11] David Goldberg, David Nichols, Brian M Oki, and Douglas Terry. 1992. Using collaborative filtering to weave an information tapestry. *Commun. ACM* 35, 12 (1992), 61–70.
- [12] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 855–864.
- [13] Guibing Guo, Jie Zhang, and Neil Yorke-Smith. 2015. TrustSVD: collaborative filtering with both the explicit and implicit influence of user trust and of item ratings. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.
- [14] 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*. International World Wide Web Conferences Steering Committee, 173–182.
- [15] Geoffrey Hinton, Li Deng, Dong Yu, George E Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N Sainath, et al. 2012. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal processing magazine* 29, 6 (2012), 82–97.
- [16] Mohsen Jamal and Martin Ester. 2010. A matrix factorization technique with trust propagation for recommendation in social networks. In *Proceedings of the fourth ACM conference on Recommender systems*. ACM, 135–142.
- [17] Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. 2014. A convolutional neural network for modelling sentences. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*.
- [18] Donghyun Kim, Chanyoung Park, Jinoh Oh, Sungyoung Lee, and Hwanjo Yu. 2016. Convolutional matrix factorization for document context-aware recommendation. In *Proceedings of the 10th ACM Conference on Recommender Systems*. ACM, 233–240.
- [19] Yehuda Koren. 2008. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 426–434.
- [20] László Lovász et al. [n. d.]. Random walks on graphs: A survey. ([n. d.]).
- [21] Hao Ma, Haixuan Yang, Michael R Lyu, and Irwin King. 2008. Sorec: social recommendation using probabilistic matrix factorization. In *Proceedings of the 17th ACM conference on Information and Knowledge Management*. ACM, 931–940.
- [22] Hao Ma, Dengyong Zhou, Chao Liu, Michael R Lyu, and Irwin King. 2011. Recommender systems with social regularization. In *Proceedings of the fourth ACM international conference on Web Search and Data Mining*. ACM, 287–296.
- [23] Daniel J O'keefe. 2002. *Persuasion: Theory and research*. Vol. 2. Sage.
- [24] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 701–710.
- [25] Sanjay Purushotham, Yan Liu, and C-C Jay Kuo. 2012. Collaborative topic regression with social matrix factorization for recommendation systems. In *Proceedings of the 24th International Conference on Machine Learning*.
- [26] Ruslan Salakhutdinov and Andriy Mnih. 2007. Probabilistic Matrix Factorization. In *21th Conference on Neural Information Processing Systems*, Vol. 1, 2–1.
- [27] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. 2001. Item-based collaborative filtering recommendation algorithms. In *World Wide Web Conference*. ACM, 285–295.
- [28] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research* 15, 1 (2014), 1929–1958.
- [29] Peijie Sun, Le Wu, and Meng Wang. 2018. Attentive Recurrent Social Recommendation. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. ACM, 185–194.
- [30] Bosiljka Tadić. 2003. Exploring complex graphs by random walks. In *AIP Conference Proceedings*, Vol. 661. AIP, 24–27.
- [31] Jiliang Tang, Xia Hu, Huiji Gao, and Huan Liu. 2013. Exploiting local and global social context for recommendation.. In *IJCAI*, Vol. 13. 2712–2718.
- [32] Jiliang Tang, Xia Hu, and Huan Liu. 2013. Social recommendation: a review. *Social Network Analysis and Mining* 3, 4 (2013), 1113–1133.
- [33] Jiliang Tang, Suhang Wang, Xia Hu, Dawei Yin, Yingzhou Bi, Yi Chang, and Huan Liu. 2016. Recommendation with Social Dimensions. In *AAAI*. 251–257.
- [34] Aaron Van den Oord, Sander Dieleman, and Benjamin Schrauwen. 2013. Deep content-based music recommendation. In *Advances in neural information processing systems*. 2643–2651.
- [35] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*. 5998–6008.
- [36] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems* 30. 5998–6008.
- [37] Hao Wang, Naiyan Wang, and Dit-Yan Yeung. 2015. Collaborative deep learning for recommender systems. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1235–1244.
- [38] Jun Wang, Lantao Yu, Weinan Zhang, Yu Gong, Yinghui Xu, Benyou Wang, Peng Zhang, and Dell Zhang. 2017. Irgan: A minimax game for unifying generative and discriminative information retrieval models. In *Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, 515–524.
- [39] Naiyan Wang and Dit-Yan Yeung. 2013. Learning a deep compact image representation for visual tracking. In *Advances in neural information processing systems*. 809–817.
- [40] Xiang Wang, Xiangnan He, Liqiang Nie, and Tat-Seng Chua. 2017. Item silk road: Recommending items from information domains to social users. In *Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, 185–194.
- [41] Xin Wang, Wei Lu, Martin Ester, Can Wang, and Chun Chen. 2016. Social Recommendation with Strong and Weak Ties. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. ACM, 5–14.
- [42] Xinxin Wang and Ye Wang. 2014. Improving content-based and hybrid music recommendation using deep learning. In *Proceedings of the 22nd ACM International Conference on Multimedia*. ACM, 627–636.
- [43] Bo Yang, Yu Lei, Jiming Liu, and Wenjie Li. 2017. Social collaborative filtering by trust. *IEEE transactions on pattern analysis and machine intelligence* 39, 8 (2017), 1633–1647.
- [44] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. Hierarchical attention networks for document classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 1480–1489.
- [45] Lili Zhao, Zhongqi Lu, Sinno Jialin Pan, and Qiang Yang. 2016. Matrix Factorization+ for Movie Recommendation. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*. 3945–3951.
- [46] Tong Zhao, Julian McAuley, and Irwin King. 2014. Leveraging social connections to improve personalized ranking for collaborative filtering. In *Proceedings of the 23rd ACM international conference on conference on information and knowledge management*. ACM, 261–270.
- [47] Yu Zhu, Hao Li, Yikang Liao, Beidou Wang, Ziyu Guan, Haifeng Liu, and Deng Cai. 2017. What to do next: Modeling user behaviors by time-lstm. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*. 3602–3608.