Machine Intelligence Research
www.mi-research.net

19(4), August 2022, 331-346
DOI: 10.1007/s11633-022-1336-1

A Novel Attention-based Global and Local Information Fusion Neural Network for Group Recommendation

Song Zhang[1, 2*] Nan Zheng[1,2*] Dan-Li Wang[1,2]

[1]State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

[2]School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: Due to the popularity of group activities in social media, group recommendation becomes increasingly significant. It aims to pursue a list of preferred items for a target group. Most deep learning-based methods on group recommendation have focused on learning group representations from single interaction between groups and users. However, these methods may suffer from data sparsity problem. Except for the interaction between groups and users, there also exist other interactions that may enrich group representation, such as the interaction between groups and items. Such interactions, which take place in the range of a group, form a local view of a certain group. In addition to local information, groups with common interests may also show similar tastes on items. Therefore, group representation can be conducted according to the similarity among groups, which forms a global view of a certain group. In this paper, we propose a novel global and local information fusion neural network (GLIF) model for group recommendation. In GLIF, an attentive neural network (ANN) activates rich interactions among groups, users and items with respect to forming a group′s local representation. Moreover, our model also leverages ANN to obtain a group′s global representation based on the similarity among different groups. Then, it fuses global and local representations based on attention mechanism to form a group′s comprehensive representation. Finally, group recommendation is conducted under neural collaborative filtering (NCF) framework. Extensive experiments on three public datasets demonstrate its superiority over the state-of-the-art methods for group recommendation.

Keywords: Group recommendation, attentive neural network (ANN), global information, local information, recommender system.

Citation: S. Zhang, N. Zheng, D. L. Wang. A novel attention-based global and local information fusion neural network for group recommendation. *Machine Intelligence Research*, vol.19, no.4, pp.331–346, 2022. http://doi.org/10.1007/s11633-022-1336-1

1 Introduction

With the increasing development of social media, group activities are becoming more and more popular[1–4]. For example, a group of people can take part in an activity together on Meetup or see a movie that they are all interested in. Such activities have led to the rapid development of group recommendation[5, 6]. Similar to personalized recommendation[6–8], group recommendation aims to address the information overloading issue for groups. For example, when a group of people is ready to dine out, group recommendation may filter much of the information they are not interested in to help the group focus on the restaurants they may like.

Traditionally, a variety of group recommendation approaches[9, 10] utilized individuals′ explicit profiles to generate group profiles by some predefined strategies; thus, a group could be regarded as a virtual individual, and then personalized recommendation methods could be used to generate group scores on the items. However, these approaches did not take into consideration the distinct preferences of members in a group. In light of this deficiency, some model-based methods[11, 12] have been proposed to model the generative decision process of a group, which considered personal explicit profiles and influences of group members, such as [11, 12].

More recently, there has been an emergence of deep learning-based approaches modelling inherent embeddings (i.e., inherent features) of groups, users and items[13–15]. Due to the sparsity of group-item interactions, representations of groups (or items) become more and more significant for group recommendation under a deep learning framework. For example, some researchers[15–17] modelled members′ embeddings in a group and then aggregated them with dynamic influence weights to obtain the group representation. These approaches work well for groups with stable members and rich interactions but may not fit groups with high sparsity. To enhance group representation, He et al.[18] first leveraged

Research Article
Manuscript received March 2, 2022; accepted May 5, 2022; published online June 17, 2022
Recommended by Associate Editor Zhi-Yuan Liu
Colored figures are available in the online version at https://link.springer.com/journal/11633
*These authors contribute equally to this work

item-user interactions and group-user interactions to represent all members in a group and then indirectly passed item features to represent groups by aggregating members′ representations.

However, there are three challenges faced for group recommendation, which have not been solved well by current state-of-the-art (SOTA) deep learning-based models:

1) (**C1**) There are abundant interactions among objects (i.e., groups, users and items)[15, 19, 20]. For example, in Fig. 1, there are three kinds of interactions (i.e., group-user, group-item and item-user interactions). Most existing group recommendation methods only model one (or two) kind(s) of interactions or indirectly leverage them all. This may lead to two issues: i) They do not get the utmost use out of interaction information; ii) An indirect aggregating strategy may weaken the effects of these objects, since the feature propagating from one to another may lose some important information. For example, some studies first adopt items to represent a user based on item-user interactions and then leverage group-user interactions to pass item inherent embeddings to group representations by aggregating users′. Such an indirect feature propagation process may lose some features appealing to the group. Thus, challenge 1 is how to make full use of these interaction information to represent a group.

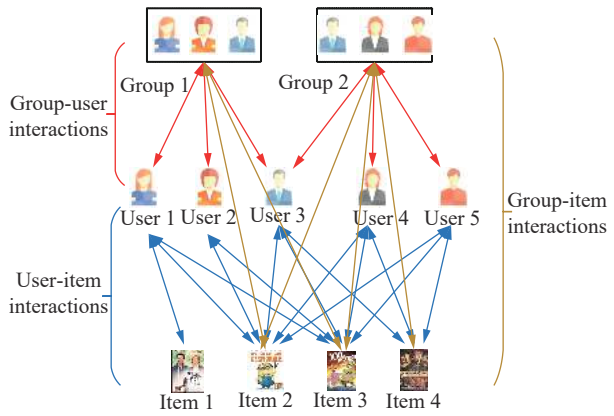


Fig. 1 An example on the interaction graph

2) (**C2**) In personalized recommendation, most methods jointly consider individual interaction records and similarity between users. For example, user-based collaborative filtering methods collaborate similar users to make individual recommendations[6, 21]. This idea can also be generalized to group recommendation. For example, in Fig. 1, Group 1 is similar to Group 2 since they are both interested in Items 2 and 3. Then, it is very possible for Group 1 to like Item 4, which has been selected by Group 2. Thus, challenge 2 is how to utilize information among similar groups to enhance group representation under a deep learning framework.

3) (**C3**) Different types of features may contribute differently to the representation. For example, for groups with a stable topic, items selected by the group may better reflect the tastes of the group since the target group is inclined to choose topic-related items. In contrast, for groups formed by friends, group members′ preferences may better reflect the tastes of the group. However, the direct concatenation operation or average aggregation cannot model such influence well, as it considers different types of embeddings equally important and does not investigate the influences among different types of features. Thus, challenge 3 is how to design a component to unite these different interaction features according to their influences to gain group representation for better group recommendation.

Motivated by the above observations, in this paper, we present a model called GLIF to learn global and local information fusion group representations for group recommendation. In GLIF, we aim to learn two levels of group representations from global information and local information: 1) On the one hand, it learns the global-level group representation by mining features shared over groups; 2) On the other hand, it learns local-level group representation by modelling interactions within a group, such as group-item interactions and group-user interactions. In GLIF, we first design a local information representation module and item representation module, which aim to make full use of three kinds of interactions to obtain representations of items′ and groups′ local features based on attentive neural network (ANN). Next, we design a global information representation module to mine similar groups and generate groups′ global representation. Considering that each similar group has a different impact on the target group, we adopt ANN to aggregate similar groups. Then, to model deep interactions of group local and global representation, we design a global and local information fusion module based on ANN. Finally, in line with previous works, we deploy neural collaborative filtering (NCF)[22] to model interaction features between groups and items and then make predictions. We evaluate GLIF extensively on three real-world datasets. Experimental results show that our model consistently outperforms the state-of-the-art methods.

In general, the main contributions of this paper are summarized as follows.

1) To the best of our knowledge, this is the first work to exploit global information over groups to enhance group representation for group recommendation.

2) Interactions within a group are fully integrated to learn group local representation. GLIF learns item and group embeddings from multiple views and derives the dynamic weight for each view embedding based on ANN.

3) We provide a principled way to exploit both global and local behavior of groups to learn a comprehensive group representation. Meanwhile, we propose a new framework GLIF to perform global and local information via an attentive aggregation module for group recommendation.

4) We conduct extensive experiments on three real

datasets. Our results show that GLIF significantly outperforms the state-of-the-art deep learning-based group recommendation models in terms of hit rate (HR) and normalized discounted cumulative gain (NDCG).

The rest of this paper is organized as follows. Section 2 highlights the related work. Section 3 discusses some preliminaries used in this work. Section 4 describes the overall architecture of GLIF for group recommendation. The experimental results are given in Section 5. Finally, conclusion and future work are given in Section 6.

2 Related work

Group recommendation aims to suggest preferred items to a group of users instead of to an individual user. Categorized by the group recommendation technique, there are two lines of research on group recommendation, namely, memory-based approaches and model-based approaches.

2.1 Memory-based approaches

Memory-based approaches, which adopt a data-independent strategy to aggregate explicit preferences or scores of members, can be characteristically dichotomized into score aggregation and explicit preference aggregation approaches[11].

The key characteristic of score aggregation approaches is to generate group scores on the target item by aggregating the scores of all members in a group obtained from individual recommendation models[23–25]. Some early works adopted predefined score aggregation strategies, including average (AVG)[24, 25], least misery (LM)[26], and maximum satisfaction (MAX)[27]. However, there was no outstanding winner within these predefined strategies[19]. Based on early aggregation approaches, recent works agreed that users′ influence should be taken into consideration when aggregating individuals′ scores. For example, GroupSA[28] proposed a social-attention network, which took social influence and dynamic weighting adjustment into consideration, to model the voting scheme of a group and then assigned weights to members to obtain the final score for group recommendation.

The key characteristic of explicit preference aggregation strategies is to generate group preference by aggregating the explicit profiles (i.e., what he/she likes or dislikes) of group members into a virtual profile. Then, these methods treat a group as a virtual individual and adopt personalized recommendation models to generate items that are preferred to a group. For example, McCarthy et al.[29] proposed a conception named the group preference agent, which could reflect a group of users′ preferences. To obtain this group agent, McCarthy et al.[29] first utilized interaction vectors of users on musical genres (e.g., new music, hot country, and dance) to generate a member′s explicit preference and then aggregated these explicit preference vectors by adding these members′ preference vectors. Yu et al.[30] first gathered all the features of users (e.g., actor, keyword) by relative importance as lexicon vectors and then merged these users′ vectors based on total distance minimization so that group vectors, which were close to most members′ vectors, could be obtained.

2.2 Model-based approaches

Model-based approaches explore and utilize interaction relationships to make group recommendations[11, 12, 31, 32]. For example, Liu et al. [12] agreed that the most influential member could dominate the decision of the group. Yuan et al.[11] proposed a generative model named consensus model (COM) to model the decision process of group activities that considered users′ influence and users′ group behaviors. Note that both of these methods utilized influences among users to model the process of a generative decision of a group, but users′ influences were fixed in different groups.

Distinct from the two methods mentioned above, recently, with the successful application of attentive network[33, 34], several attention-based models have been elaborated to support the influential mechanism for group recommendation[15–18, 28]. In such models, the influence of users is dynamically determined by a specific group or item. Cao et al.[15] first leveraged a neural attention network to learn the relative influence among members in a dynamic way and then utilized group members′ preferences to represent the group preference by considering group-user interactions. Tran et al.[17] first employed a subattention neural network to model user-user interactions in a group and then exploited the remaining members to represent a user′s preference. Finally, they aggregated a group of users by average strategy. He et al.[18] fused users, items and groups′ features from multiple views by attention mechanism to capture representations and then made group recommendations. Firstly, they leveraged item-user interactions to obtain user representation and item representation. Secondly, they adopted group-user interactions to obtain group representation. Finally, they used attention mechanism to aggregate user representation to gain group representation and optimized it under the NCF framework. He et al.[16] adopted a multilayer perceptron (MLP)[35] to model group-user interactions and leveraged item-user interactions to aggregate users′ representations. Guo et al.[36] introduced friend preferences from the social network and further exploited group-level similarity with a hyperedge embedding scheme to learn group representation.

Most existing works usually conduct interaction relations (i.e., item-user, group-user, group-item) for group representation, which can be considered as the local information for a group. However, there also exists global information among similar groups, which may enhance group representation from a global view. To sum up, there are three major differences between the works men-

tioned above and ours: 1) We make full use of three direct interactions (i.e., group-user, group-item and item-user) to represent groups and items instead of using partial or indirect interactions; 2) We conduct group recommendations based on both global and local information instead of just local interactions for groups. 3) Our work introduces an attention mechanism on different levels of group representation to better characterize group preferences.

3 Problem statement and preliminaries

In this section, we present the problem of group recommendation and then introduce two types of neural network models, i.e., type-based ANN and k types-based ANN, that will be used in the paper.

3.1 Notations and problem statement

We use bold capital letters (e.g., $\boldsymbol{X}$) and bold lowercase letters (e.g., $\boldsymbol{x}$) to represent matrices and vectors, respectively. We employ non-bold letters (e.g., x) to denote scalars and non-bold capital letters (e.g., X) to denote sets. If not clarified, all vectors are in column forms.

Let $U = \{u_1, u_2, \cdots, u_n\}$, $V = \{v_1, v_2, \cdots, v_m\}$ and $G = \{g_1, g_2, \cdots, g_s\}$ be the sets of users, items and groups, respectively. The l-th group $g_l \in G$ consists of a set of users, i.e., group members $u_{l,j} \in U$. There are three kinds of observed interaction relations among U, V and G, namely, group-item interactions, item-user interactions and group-item interactions. Item-user interactions denote an item that has been selected or rated by a user. Group-user interactions denote that a user is in a group. Group-item interactions denote that a group has an interaction history with an item. We use $\boldsymbol{A} = [a_{li}]_{s\times m}$ to denote group-item interactions, $\boldsymbol{B} = [b_{ij}]_{m\times n}$ to denote item-user interactions and $\boldsymbol{C} = [c_{lj}]_{s\times n}$ to denote group-user interactions.

Given a target group g_l, the problem of group recommendation is defined as recommending a list of items which users in group g_l should be interested in, which is formally defined as

Input: U, V, G, $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$

Output: One function that generates a score for each group on an item $f : V \to \mathbf{R}$.

3.2 Definition

In this subsection, we present some definitions and concepts used in this paper.

Definition 1 (**Inherent embedding**). Let $\boldsymbol{u} \in \mathbf{R}^d$, $\boldsymbol{v} \in \mathbf{R}^d$ and $\boldsymbol{g} \in \mathbf{R}^d$ be user $u \in U$, item $v \in V$, and group $g \in G$ inherent embeddings, respectively, where d denotes the dimension of a vector. Under deep learning framework, inherent embeddings are learnable parameters that can reflect the latent interests of a user (or group) or the latent features of an item. In our paper, inherent embeddings are initialized by the Xavier strategy[37].

Definition 2 (γ-**neighbor set**) (N_{gl}^{γ})**.** The γ-neighbors set consists of the top γ groups, which are similar to the target group g_l. Firstly, similarities between target group g_l and the rest groups are calculated. Then, the rest groups are ranked by similarities. Finally, top-γ groups $g_{l,1}$, $g_{l,2}$, $\cdots$, $g_{l,\gamma}$ are selected as γ-neighbors set of group g_l, denoted as N_{gl}^{γ}, where γ is a hyperparameter to control the number of neighbors. Note that parameter γ favors the modelling of high semantic information over groups since it is helpless (even noise) for capturing global-level semantic features if beyond the scope of γ.

Definition 3 (Global information). Inspired by homophily theory, which shows that people with high similarity tend to share common tastes[38, 39], we propose the concept of global information among groups with high similarity. According to Definition 2, for each group $g_l \in G$, global information is defined as $\{g_{l,k} | \, g_{l,k} \in N_{gl}^{\gamma}; g_{l,k} \in G\}$.

Definition 4 (Local information). In contrast, local information means learning embeddings from interactions within groups. Suppose we have a target group g_l, each user $u_j \in I_u(g_l)$ and item $v_i \in I_v(g_l)$ is called local information of group g_l. $I_u(g_l)$ and $I_v(g_l)$ are the sets consisting of users and items, respectively, which have interacted with group g_l.

3.3 Attention models: Type-based ANN and K types-based ANN

In this subsection, we present two different attention models to aggregate semantic embeddings in different situations.

Type-based ANN (TANN). TANN aims to aggregate the same type of inherent embeddings to represent the target embedding. Suppose there is a set $I_t(q_j) = \{p_{j,i}^t, (i = 1, 2, \cdots, k)\}$ including k objects of the same type that have interactions with the target object q_j. Denote the t-type interaction as $p_{j,i}^t$. Let $\boldsymbol{p}_{j,i}^t$ and $\boldsymbol{q}_j$ be embedding vectors for $p_{j,i}^t$ and q_j. Then, we employ a neural network named $TANN$ to aggregate content embeddings. Formally, the aggregated t-type content embeddings for q_j are given as follows:

$$\boldsymbol{q}_j^t = TANN(q_j, I_t(q_j)) = \sum_{p_{j,i}^t \in I_t(q_j)} \alpha_i \boldsymbol{p}_{j,i}^t \tag{1}$$

$$\alpha_i = \frac{\exp\left\{MLP\left(\left[\boldsymbol{q}_j \oplus \boldsymbol{p}_{j,i}^t\right]\right)\right\}}{\sum_{p_{j,i}^t \in I_t(q_j)} \exp\left\{MLP\left(\left[\boldsymbol{q}_j \oplus \boldsymbol{p}_{j,i}^t\right]\right)\right\}} \tag{2}$$

where $\oplus$ denotes the operation of concatenation and MLP is multi-layer perceptron that uses the rectified linear

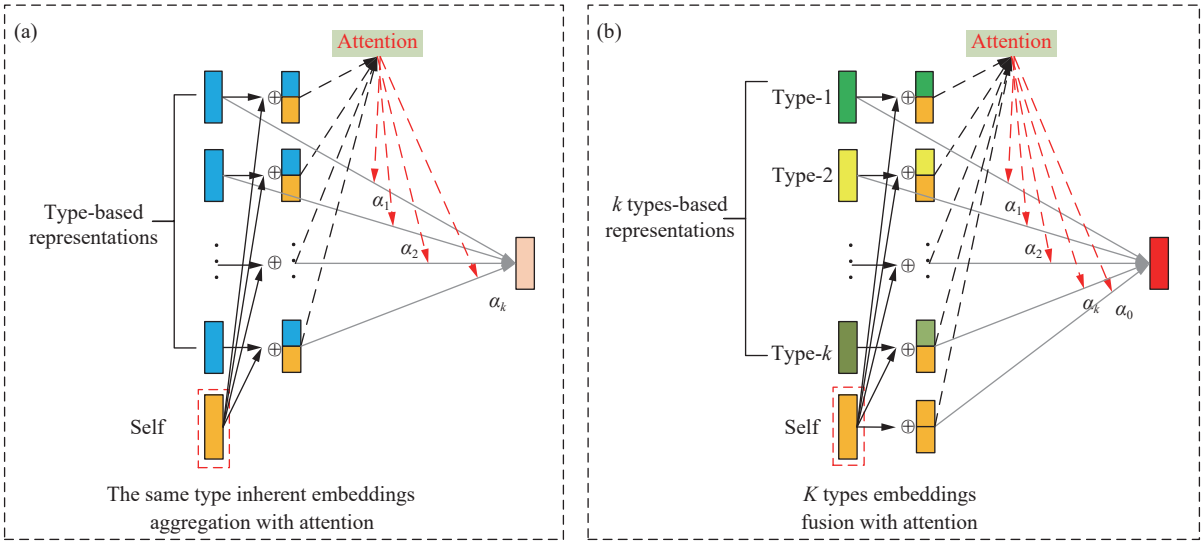


Fig. 2 Overall architectures of two attention models: (a) TANN; (b) KTANN.

unit (ReLU) function as the nonlinear activation function. Fig. 2(a) shows the detailed architecture of the TANN.

***K* types-based ANN (KTANN)**. KTANN aims to combine k types of embeddings to obtain a semantic fusion representation for the target q_j. Suppose we extract k types of features of target q from k types of views. Let $\boldsymbol{q}_j^0$ and $\boldsymbol{q}_j^t (t = 1, 2, \cdots, k)$ be the embeddings of q_j and k types of features, respectively. Then, we use a neural network to fuse k types of features and its self-embedding:

$$KTANN(q_j) = \sum_{t=0}^{k} \beta_t \boldsymbol{q}_j^t \tag{3}$$

$$\beta_i = \frac{\exp\left\{MLP\left(\left[\boldsymbol{q}_j^0 \oplus \boldsymbol{q}_j^t\right]\right)\right\}}{\sum_{t=0}^{k} \exp\left\{MLP\left(\left[\boldsymbol{q}_j^0 \oplus \boldsymbol{q}_j^t\right]\right)\right\}}. \tag{4}$$

Fig. 2(b) shows the detailed architecture of the KTANN.

4 Framework of GLIF

4.1 Overall framework

In this subsection, we formally present a novel global and local information fusion neural network for group recommendation (GLIF). GLIF aims to exploit both global-level and local-level pairwise groups and items for modelling group preference and item feature for group recommendation. Fig. 3 presents the overall architecture of GLIF. It first leverages item-user interactions to represent each item. To obtain a group's local representation, it jointly utilizes group-user interactions and group-item interactions. Next, it uses global information between different similar groups to obtain the group's global representation. Then, it adopts ANN to fuse global and local representations. Finally, it uses fully connected layer (NCF) to model the interaction between group and item and an FCL to predict a score.

4.1.1 Item representation learning module

Different from most previous works that regarded item inherent embedding as final item representation, GLIF models item representation from item-user interactions. Intuitively, if an item has been selected by a user, it should have some features appealing to the user. To understand this, let us consider an example that if an item has been rated by a user who likes comedy films, it may contain some comic features. In light of such an idea, we design a module to learn user-view item representation $\boldsymbol{v}_i^u$ based on a two-layer attention neural network,

$$\boldsymbol{v}_i^u = TANN\left(v_i, I_u\left(v_i\right)\right) = \sum_{u_{i,j} \in I_u(v_i)} \alpha_j \boldsymbol{u}_{i,j} \tag{5}$$

where α_j is the attention weight of $\boldsymbol{u}_{i,j}$, which is calculated in (2), $\boldsymbol{u}_{i,j}$ is the inherent embeddings of user j, and $I_u\left(v_i\right)$ is a set of users who have interactions with item v_i.

Obviously, the inherent and user-view item embeddings (i.e., $\boldsymbol{v}_i$ and $\boldsymbol{v}_i^u$, respectively) imply the complementary features of an item from different perspectives. We design an ANN-based KTANN to fuse these two types of item embeddings as the item representation. Accordingly, item representation for item v_i is finally obtained by

$$\widehat{\boldsymbol{v}}_{\boldsymbol{i}} = KTANN\left(\boldsymbol{v}_i\right) = \beta_1 \boldsymbol{v}_i + \beta_2 \boldsymbol{v}_i^u \tag{6}$$

where β_1 and β_2 are learned from ANN, as described in (4). Fig. 3(a) shows the process of item representation learning.

4.1.2 Group local representation learning module

The key idea of most group recommendation models is to aggregate inherent embeddings from group members to

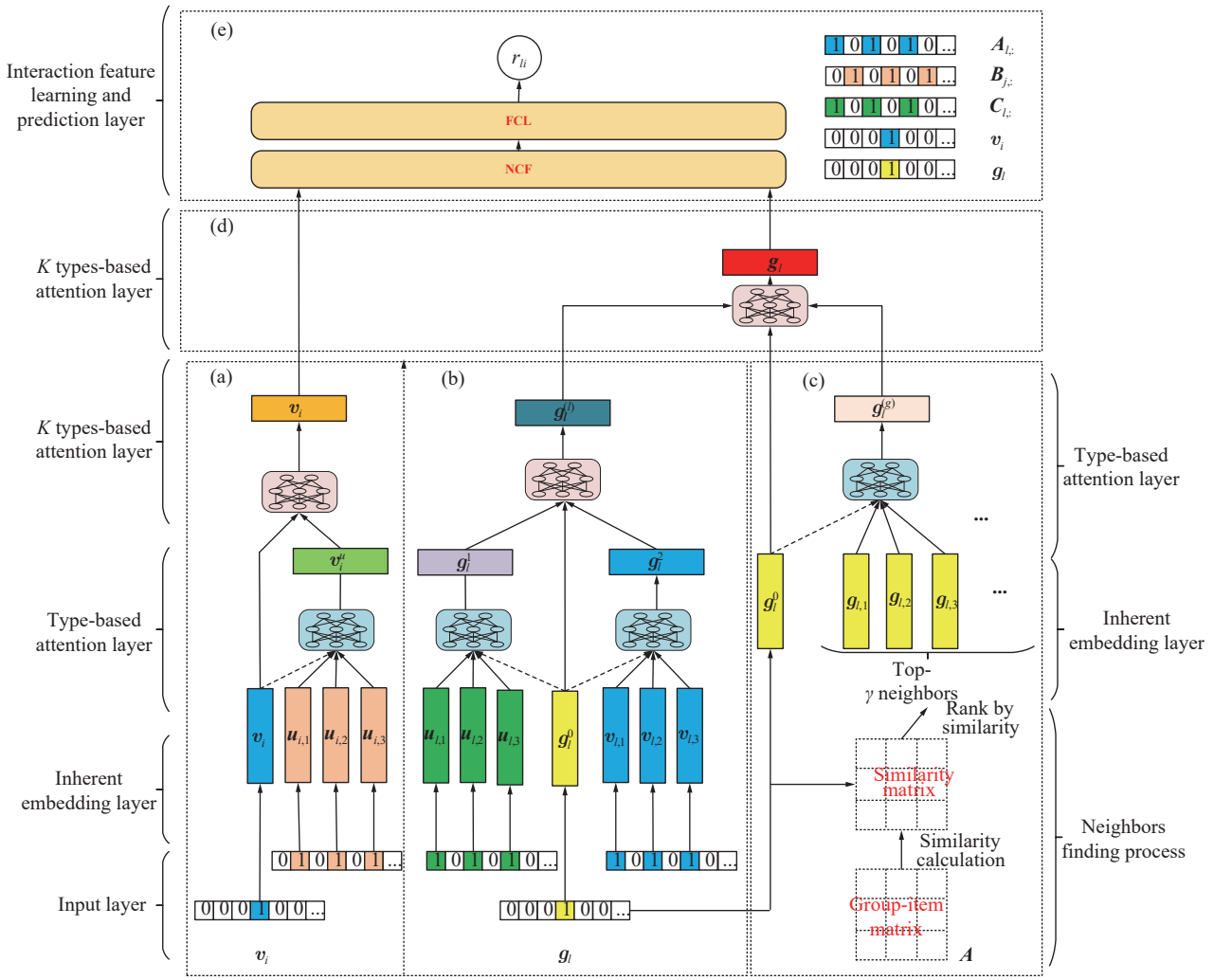


Fig. 3 The overall architecture of GLIF: (a) Item representation learning module; (b) Group local representation learning module; (c) Group global representation learning module; (d) Group global and local representation fusion module; (e) NCF and predicting module. $\boldsymbol{A}_{l,:}$ is the l-th row in group-item interactions $\boldsymbol{A}$. $\boldsymbol{B}_{j,:}$ is the j-th row in item-user interactions $\boldsymbol{B}$. $\boldsymbol{C}_{l,:}$ is the l-th row in group-user interactions $\boldsymbol{C}$. v_i and g_l are one-hot vectors of the item and group, respectively.

obtain the group local representation, regardless of group-item interactions[15–17, 40, 41]. Some works have tried to obtain better user embeddings enhanced by item-user interactions and then aggregated group members' embeddings[18]. These methods may raise two issues: 1) They do not take full use of group-item interactions since items selected and rated by a group naturally reflect this group's tastes. 2) This feature propagation process, (i.e., item→user→group), may lose some features appealing to the group. Considering these two issues, GLIF models a group's local representation from two kinds of interactions, including group-item interactions and group-user interactions. Specifically, we design a local information model based on TANN and KTANN to derive the complementary group's local representation. Firstly, two TANNs are used to aggregate members' inherent preferences and items' inherent features to obtain user-view group embedding and item-view group embedding, respectively. Then, KTANN is leveraged to aggregate three kinds of group embeddings, including group inherent embedding, to derive g_l's local representation $\boldsymbol{g}_l^{(l)}$,

$$\boldsymbol{g}_l^u = TANN\left(g_l^0, I_u\left(g_l\right)\right) = \sum_{u_{l,j} \in I_u(g_l)} \alpha_j^u \boldsymbol{u}_{l,j} \tag{7}$$

$$\boldsymbol{g}_l^v = TANN\left(g_l^0, I_v\left(g_l\right)\right) = \sum_{v_{l,i} \in I_v(g_l)} \alpha_i^v \boldsymbol{v}_{l,i} \tag{8}$$

$$\boldsymbol{g}_l^{(l)} = KTANN\left(g_l\right) = \beta_0 \boldsymbol{g}_l^0 + \beta_1 \boldsymbol{g}_l^u + \beta_2 \boldsymbol{g}_l^v \tag{9}$$

where $\beta_k (j = 0, 1, 2)$ are attention weights of $\boldsymbol{g}_l^0$, $\boldsymbol{g}_l^u$, $\boldsymbol{g}_l^v$, respectively, which are calculated in (4), $\boldsymbol{g}_l^0$ is the inherent embedding of g_l, $\boldsymbol{g}_l^u$ is the user-view embedding of g_l, $\boldsymbol{g}_l^v$ is the item-view embedding of g_l, and α_j^u and α_i^v are weights of $u_{l,j}$ and $v_{l,i}$, respectively, as described in (2). $I_u\left(g_l\right)$ and $I_v\left(g_l\right)$ are sets of users and items that have interacted with g_l.

Fig. 3(b) illustrates the design of the group local representation learning component. With such a soft attention mechanism, a dynamic weight can be assigned to a user by his/her influence (7)[12], and also an unfixed weight can be assigned to an item by degree of group preference on varying items (8). Such attention mechanism may investigate different influences among different types of features. Also, this mechanism is more general, and existing heuristic predefined aggregation can be regarded as its special case. For example, *AVG* (e.g., average strategy) is parallel to allocating a uniform weight to every member; however, *LM* (e.g., least misery) and *MS* (e.g., max satisfaction) mean giving nonzero weights to partial members only. Considering two examples: 1) When a group discusses whether to watch The Lion King, the members who have watched many Disney movies should be more influential and should be assigned a larger weight. 2) A group may prefer to watch comedy movies, so comedy movies should be better for representing the group, and a larger weight should be allocated. As noted previously, history items should contribute more to learning representations of the groups with a stable topic, while the members′ preferences should contribute more in the groups consisting of friends or people with close relationships, so we adopt dynamic weights learned by neural networks to aggregate item-view embeddings and user-view embeddings of g_l (9).

4.1.3 Group global representation learning module

Except for local information, there also exists global information among different groups. Specifically, groups with high similarity tend to have similar tastes. In this module, we capture global information for learning global representations of groups. To obtain group global representation, we first calculate the Pearson′s correlation coefficient (PCC)[42] as similarity and then sample a fixed size neighbor set by selecting top-γ similar groups. Accordingly, we denote the sampled neighbor set of $g_l \in G$ as the γ-neighbor set $N_g^{\gamma}(g_l)$. Then, we employ TANN to aggregate the inherent embeddings of neighbors $g_{l,k} \in N_g^{\gamma}(g_l)$. Formally, the group global embedding $\boldsymbol{g}_l^{(g)}$ for g_l is formulated as follows:

$$\boldsymbol{g}_l^{(g)} = TANN\left(g_l, N_g^{\gamma}(g_l)\right) = \sum\nolimits_{g_{l,k} \in N_g^{\gamma}(g_l)} \alpha_k \boldsymbol{g}_{l,k} \quad (10)$$

where α_k is the attention weight of the k-th neighbor of g_l, as described in (2).

4.1.4 Group global and local representation fusion module

We design a module to fuse global information with local information to enhance group representation. To obtain the final group representation $\widehat{\boldsymbol{g}}_l$, we implement the aggregation function as follows:

$$\widehat{\boldsymbol{g}}_l = KTANN(g_l) = \alpha_0 \boldsymbol{g}_l^0 + \alpha_1 \boldsymbol{g}_l^{(l)} + \alpha_2 \boldsymbol{g}_l^{(g)} \quad (11)$$

where $\alpha_k (k = 0, 1, 2)$ are the attention weights of group inherent embedding $\boldsymbol{g}_l^0$, group local representation $\boldsymbol{g}_l^{(l)}$ and group global representation $\boldsymbol{g}_l^{(g)}$, as described in (4). Fig. 3(d) illustrates the design of the group global and local representation fusion component.

4.1.5 Interaction feature learning and predicting module

NCF is a neural network model for personalized recommendation. Its key idea is to learn an interaction function from history interaction records with inputs of user inherent embeddings and item inherent embeddings[15, 22]. In our paper, NCF customized by [15] is adopted to learn embeddings and interaction function. Formally, given the final item representation $\widehat{\boldsymbol{v}}_i$ and group representation $\widehat{\boldsymbol{g}}_l$, the hidden interaction vector $\boldsymbol{h}_{l,i}$ is computed by

$$\boldsymbol{h}_{l,i} = \boldsymbol{W}^{\mathrm{T}} \sigma\left(\boldsymbol{W}_{\boldsymbol{gv}} [\widehat{\boldsymbol{g}}_l \ \widehat{\boldsymbol{v}}_i] + \boldsymbol{W}_{\boldsymbol{g}} \widehat{\boldsymbol{g}}_l + \boldsymbol{W}_{\boldsymbol{v}} \widehat{\boldsymbol{v}}_i + \boldsymbol{b}\right) \quad (12)$$

where $\boldsymbol{W}^{\mathrm{T}}$, $\boldsymbol{W}_{\boldsymbol{gv}}$, $\boldsymbol{W}_{\boldsymbol{g}}$, $\boldsymbol{W}_{\boldsymbol{v}}$ and $\boldsymbol{b}$ denote weight matrices and bias vector, respectively, $\odot$ denotes element-wise product operation. We simply use ReLU $\sigma(x)$ as a nonlinear activation function.

Finally, we adopt a fully connected layer with sigmoid activation function to map hidden interaction vector $\boldsymbol{h}_{l,i}$ to the score of group g_l on item v_i:

$$r_{li} = Sigmoid\left(\boldsymbol{W}_h^{\mathrm{T}} \boldsymbol{h}_{l,i} + \boldsymbol{b}\right) \quad (13)$$

where $\boldsymbol{W}_h^{\mathrm{T}}$ and b are learnable parameters. Fig. 3(e) shows the design of NCF and predicting components.

4.2 Model optimization

Pairwise learning is adopted to optimize the parameters since we address the top-K group recommendation task. Pairwise learning considers scores of positive instances and negative instances and assumes that the predicted scores for positive instances are near 1 and scores for negative instances are near 0. In line with [15, 18], we employ regression-based pairwise loss:

$$loss = \sum\nolimits_{(l,i,i') \in T} \left(r_{li} - r_{li'} - 1\right)^2 \quad (14)$$

where T denotes the training set, in which each instance is a triplet (l, i, i') meaning that group g_l has interacted with item v_i but not interacted with $v_{i'}$. r_{li} and $r_{li'}$ are scores for group g_l on items v_i and $v_{i'}$, respectively.

4.3 Time complexity

We further analyse the time complexity of the proposed method. The time complexity of the item representation learning module is $O(ed)$, where e is the number of users used to represent an item and d is the embedding dimension. In the group local representation

learning module, the time complexity of the user-view group representation learning component and item-view group representation learning component are $O(nd)$ and $O(cd)$, respectively, where n and c are the group size and the number of items used to represent each group, respectively. In the group global representation learning module, the time complexity of the procedure calculating Pearson's correlation coefficient (PCC) of each paired group is $O(s^2)$, where s is the number of groups. It should be noted that we can calculate all PCCs and find top-γ neighbors for each group first, and it needs no more calculation for the next prediction. Thus, in the prediction stage, its time complexity is $O(1)$. In terms of the component of aggregating neighbors, its time complexity is $O(\gamma d)$, where γ is the neighbor size. The time complexity of the global and local representation fusion module is $O(2d)$. The time complexity of the interaction feature learning and predicting module is $O(d)$. Overall, in the prediction stage, the time complexity of GLIF is $O(ed) + O(nd) + O(cd) + O(\gamma d) + O(2d) + O(d) = O((n + c + e + \gamma)d)$. Please note that the item representation learning module, user-view group representation learning component, item-view group representation learning component and the component of aggregating neighbors can be deployed on different CPUs or GPUs for executing in parallel since their results are independent of each other. Thus, the time complexity of parallel GLIF is $O(\max(n, c, e, \gamma)d)$. In our model, we run it in parallel.

5 Experiments

In this section, we conduct extensive experiments with the aim of answering the following research questions:

1) (**RQ1**) How does GLIF perform VS. state-of-the-art baselines for group recommendation?

2) (**RQ2**) How do different components, e.g., group local representation or group global representation, affect the model performance?

3) (**RQ3**) How do various hyperparameters, e.g., group neighbor size or embedding dimension, impact the model performance?

4) (**RQ4**) How does the attention mechanism work?

5.1 Experimental settings

5.1.1 Datasets

We conduct experiments on three real public datasets: CAMRa2011 is a movie dataset that consists of movie rating records for individual users and households. The dataset that is prepared and available for download by the paper[15]. In line with [15, 16], we transform explicit rating records into positive instances. If a user/group has rated a movie, the score will be set to 1; otherwise, it will be 0. MS and MR are from MovieLens 1M data[1]. It contains one million movie ratings for 4K movies by over 6K users. Following the approach in [17, 24], we extract two datasets from this data: MS and MR, respectively. MS contains groups with high user-to-user similarity. MR contains groups that are formed randomly. For a given group in both cases, if a movie is given 4 stars or above by every member in the group, then the movie is adopted by the group.

Table 1 shows the statistics of the three datasets. Note that from the first dataset to the last dataset, the sparsity of group-item interactions increases.

Table 1 Dataset statistics

Dataset	CAMRa2011	MS	MR
Number of items	7 710	783	447
Number of users	602	4 883	3 431
Number of groups	290	3 000	3 000
Avg. No. of users of a group	2.08	5	5
Avg. No. of interacting items of a group	500.23	12.19	5.17
Avg. No. of interacting groups of an item	18.82	46.71	34.71
Avg. No. of interacting items of a user	193.27	44.14	81.23
Avg. user similarity in a group	0.296	0.495	0.163

5.1.2 Baselines

We evaluate the performance of GLIF by comparing it with the following state-of-the-art models:

1) NCF-Based. Based on three predefined aggregation strategies, we employ SOTA neural recommendation NCF[2][22] to make group recommendations. neural collaborative filtering-least misery (NCF-AVG) adopts average strategy to aggregate the preference scores of individuals in a group, which are generated by NCF, as the group preference score. This aggregation strategy is equivalent to assigning a uniform weight to each user in a group. NCF-LM assumes that the least satisfied member determines the final group decision. Thus, it regards the minimum score of individuals as the group preference score and then optimizes this score. This strategy is equivalent to assigning zero weight to those users whose score is higher than the minimum score. NCF-MS[27] applies the maximum satisfaction strategy to maximize the satisfaction of group members. In our work, the maximum score is treated as the preference of the group.

2) Factorization machines-average (FM-AVG)[9]. Similar to NCF-AVG, FM-AVG takes the average score of all group members as the group preference score and models the interactions by the linear relation of matrix factorization.

3) DeepGroup[3][13]. DeepGroup – A deep neural net-

[1] https://grouplens.org/datasets/movielens/

[2] https://github.com/hexiangnan/neural_collaborative_filtering

[3] https://github.com/sarinasajadi/DeepGroup/tree/main/dataset/sushi

[4] https://github.com/LianHaiMiao/Attentive-Group-Recommendation

work for learning group representations and decision making, learns group representations by employing average aggregator to fuse members' preferences.

4) AGREE[4][15]. Attentive group recommendation (AGREE) directly aggregates the users' inherent embeddings with the group inherent embedding under a standard attention network and adopts a customized NCF to model the interactions between groups and items.

5) MoSAN[17]. Medley of sub-attention networks (MoSAN) directly sums all members' preferences as a group's preference. Each member's preference is fused from the preference of the remaining members in the group by a sub-attention network.

6) GAME[18]. Graphical and attentive multi-view embeddings (GAME) employs graph convolutional network to aggregate first-order neighbors to obtain user/item representations from multiple views and utilizes a soft attention network to learn the item-related influence of members.

5.1.3 Evaluation metrics

The leave-one-out evaluation protocol is adopted to evaluate the performance of GLIF, which has been widely applied to evaluate the performance of the top-K recommendations[15, 43–45]. Since it is too time-consuming to rank all items for each group, we follow the common scheme that randomly selects 100 negative items and ranks the testing item among the 100 items[15, 18, 22, 46, 47]. Following previous works, two widely used evaluation metrics at top-K recommendations are adopted: Hit ratio (HR) and normalized discounted cumulative gain (NDCG)[15, 16, 18, 22, 41] denoted $HR@K$ and $NDCG@K$, respectively. Larger values indicate better performance. We calculate the two metrics for each test instance and report the average. The formulas for the two metrics are defined as follows:

$$HR = \frac{\#hit@k}{N} \tag{15}$$

$$DCG@K = \sum_{i=1}^{K} \frac{2^{rel_i} - 1}{log\,(i+1)} \tag{16}$$

$$NDCG@K = \frac{DCG@K}{IDCG@K} \tag{17}$$

where $\#hit@k$ denotes the number of *hits* in the test set and N is the total number of test cases. $rel_i = 1$ indicates that the item at rank i in the Top-K recommendation list is in the test set; otherwise, $rel_i = 0$. $IDCG$ means the maximum possible DCG through ideal ranking.

5.1.4 Implementation details

GLIF was implemented in PyTorch. For hyperparameter tuning, 6 negative instances were randomly sampled for each group. The embedding layer and other layers were initialized by the Xavier strategy and standard normal strategy, respectively. The embedding size is set as $d = 32$. Root mean square prop (RMSProp) was adopted as the optimizer for all gradient-based methods, where the mini-batch size, learning rate and dropout rate were set as 256, 0.000 1 and 0.2, respectively. We followed scheme[15], which conducted a paired two-sample t test on NDCG@10 based on the 5 experimental results and reported the average results in Table 2.

Table 2 Time complexity and runtime of all models in each iteration (s). n is the group size, d is the embedding dimension, c is the number of items used to represent each user, m is the number of groups used to represent each user, f is the number of users used to represent each item, k is the number of groups used to represent each item, e is the number of users used to represent each item, γ is the neighbor size.

Model	Complexity	Time consuming
NCF-AVG	$O(nd)$	6.32
NCF-LM	$O(nd)$	6.15
NCF-MS	$O(nd)$	6.23
FM-AVG	$O(nd)$	6.87
DeepGroup	$O(nd)$	7.36
AGREE	$O(nd)$	9.45
MoSAN	$O(n^2d)$	12.23
GAME	$O\left((nc + nm + f + k)\,d\right)$	20.32
GLIF	$O\left(\max\left(n, c, e, \gamma\right)d\right)$	11.33

5.2 Experimental results

5.2.1 Overall performance comparison (RQ1)

Effectiveness. We compare our GLIF with some state-of-the-art deep learning-based group recommendation models. The performances of all models are reported in Table 3, where the best results are highlighted in bold, and second-best results are underlined. According to Table 3, 1) GLIF achieves the best performance on three datasets for group recommendation, and its results are significant in most cases. This demonstrates that our proposed GLIF model is effective under attentive neural networks and the positive effect of global and local information fusion solution for group recommendation. 2) Although the similarities of members in a group vary from three datasets, our model can consistently achieve the best performance. This indicates that our model can be adapted to diverse datasets. 3) Among the NCF-based methods, in most cases, NCF-AVG performs better than NCF-LM and NCF-MS. Our explanation is that the aggregation strategies of the latter two only focus on the minority of members, while NCF-AVG can take all members in a group into consideration. Thus, it can describe

Table 3 Overall performance comparison on CAMRa2011, MS and MR datasets. ** for $p < 0.01$ and * for $p < 0.05$

Dataset	Metric	NCF-AVG	NCF-LM	NCF-MS	FM-AVG	Deep Group	AGREE	MoSAN	GAME	GLIF
CAMRa2011	HR@5	0.583 3	0.571 4	0.571 9	0.542 7	0.563 7	0.585 0	0.587 3	0.590 9	**0.591 8**
	HR@10	0.776 5	0.771 3	0.751 2	0.709 5	0.764 5	0.779 3	0.775 1	0.786 4	**0.789 3**
	NDCG@5	0.396 9	0.396 3	0.385 0	0.367 2	0.379 6	0.402 5	0.402 4	0.402 3	**0.403 0**
	NDCG@10	0.462 5	0.458 1	0.444 1	0.420 3	0.433 5	0.466 2	0.463 1	0.467 0	**0.467 3**
	p-value	1.14E–10**	5.06E–11**	1.03E–13**	2.10E–16**	2.38E–7**	8.16E–3**	2.94E–8**	3.63E–2*	–
MS	HR@5	0.591 9	0.633 1	0.644 3	0.643 2	0.652 3	0.659 6	0.664 1	0.659 7	**0.664 3**
	HR@10	0.831 5	0.810 7	0.822 5	0.824 3	0.822 9	0.832 3	0.817 7	0.832 2	**0.835 5**
	NDCG@5	0.473 5	0.459 2	0.466 2	0.472 2	0.479 2	0.473 3	0.470 2	**0.483 8**	0.482 0
	NDCG@10	0.522 1	0.511 9	0.519 8	0.521 0	0.522 5	0.529 4	0.516 3	0.532 5	**0.534 4**
	p-value	1.30E–11**	6.55E–15**	1.86E–13**	3.55E–11**	3.57E–13**	2.46E–9**	3.03E–13**	6.01E–7**	–
MR	HR@5	0.635 2	0.633 2	0.623 5	0.630 3	0.629 6	0.641 0	0.652 1	0.655 5	**0.656 1**
	HR@10	0.784 2	0.784 6	0.778 5	0.783 8	0.781 8	0.790 1	0.797 5	0.793 2	**0.799 3**
	NDCG@5	0.453 2	0.451 8	0.444 3	0.450 5	0.452 3	0.457 6	0.452 3	0.464 1	**0.464 3**
	NDCG@10	0.502 9	0.500 3	0.490 2	0.498 6	0.496 9	0.506 9	0.505 4	0.501 0	**0.510 7**
	p-value	6.92E–10**	2.39E–9**	2.43E–9**	1.55E–10**	2.51E–13**	2.48E–7**	3.81E–9**	1.25E–10**	–

group preference more comprehensively. 4) Compared with predefined aggregation strategies methods, i.e., NCF-AVG, NCF-LM and NCF-MS, ANN-based methods, i.e., AGREE, MoSAN, GAME and GLIF, perform better. This indicates that the traditional score aggregation baseline methods are insufficient to make group recommendations well, while ANN shows remarkable superiority, which dynamically assigns weight to each member by learning from interactions. 5) GAME underperforms GLIF in most cases. The reason is that the feature propagation process (item→user→group) may weaken the effect of items that reflect group preference.

Efficiency. Since all the proposed algorithms rely on stochastic gradient descent, we show the runtime of the methods performed on CAMRa2011 in one iteration in Table 2. In practice, all the algorithms would converge in less than 30 iterations. Compared with NCF-based methods, AGREE, MoSAN GAME and GLIF need more runtime by adding extra attention networks. Among attention-based models, GAME and GLIF need to compute multiple attention weights to further identify different items or groups; thus, they cost more time while performing better than AGREE and MoSAN. Notably, GLIF costs approximately 1.2 times as much as AGREE, while GAME costs approximately 2.2 times as much as AGREE. Given the above analysis, we empirically conclude that our method achieves a good compromise in accuracy and time efficiency.

5.2.2 Study on the components of GLIF (RQ2)

The key characteristics in our proposed GLIF model are: Item representation learning module that captures item features based on item-user interactions; the group local representation learning module that captures group′s local preference based on group-item and group-user interactions; the group global representation learning module that captures group′s global preference based on the similarity among groups; and the global and local representation fusion module that captures group′s comprehensive representation based on the attention mechanism.

Here, we study the components of GLIF by evaluating five variants: 1) GLIF-nGIR without leveraging the group-item interactions to represent groups (i.e., items′ fusion representation by group). 2) GLIF-nIUR without embedding from item-user interactions (i.e., users′ fusion representation by item). 3) GLIF-nGR without considering global information (i.e., global representation). 4) GLIF-nGLF adopting averaging strategy to fuse global and local representations instead of attention neural network. 5) GLIF-nLR without considering local information. 6) GLIF-T that replaced all type-based ANNs as average aggregators. 7) GLIF-K that replaced all K types-based ANNs as average aggregators.

Fig. 4 depicts the recommendation performance of the five variants in comparison to GLIF on NDCG. According to it, GLIF consistently and significantly outperforms the five variants on three datasets. This indicates that all components are beneficial to model group decisions, and combining them contributes to better performance. Specifically,

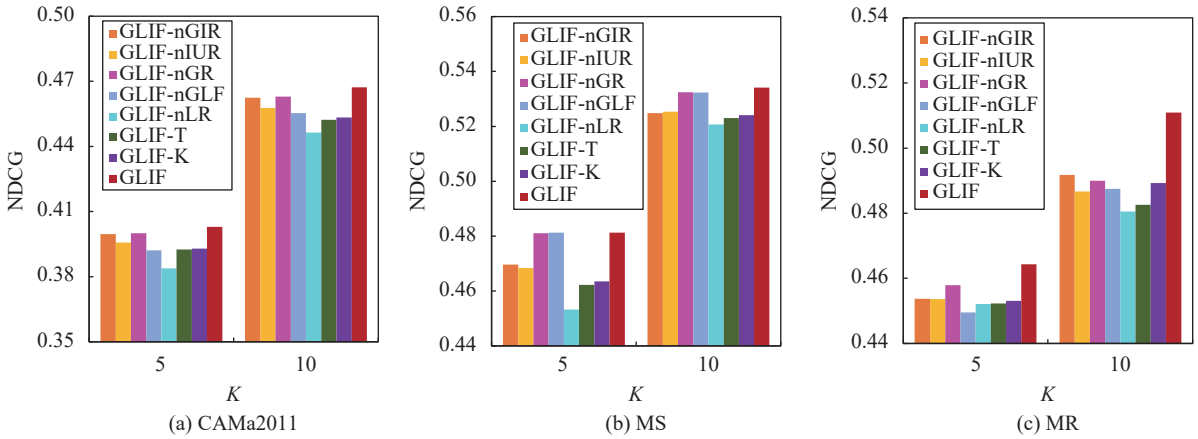


Fig. 4 Performance on the variants of GLIF

Item representation. GLIF-nIUR performs worse than GLIF on all three datasets, which indicates that mining item features from user preferences is useful since user preferences can reflect item features to some extent. Among the three datasets, GLIF-nIUR performs worst (against GLIF) on MR dataset. This may due to that compared with the other two datasets, MR dataset consists of fewer group-item interactions. Thus, item representation learned from item inherent embedding is not sufficient, and user inherent embedding may be a good complement to item representation.

Group's local information representation. GLIF-nLR always performs worse than GLIF. which indicates that it makes sense to derive group local representation based on user-view and item-view group representation and assigning dynamic influence weights to them is meaningful for group recommendation. GLIF-nGIR shows worse performance than GLIF, which indicates that it is useful to compress item inherent embedding rated by group into group representation. It is noticeable that the performances of GLIF-nLR, GLIF-nIUR and GLIF are constantly improving on all three datasets. Our explanation is that they consider one, two and three types of interactions, respectively, which indicates that utilizing more interactions results in more accurate group preference.

Group global information representation. Compared with GLIF, GLIF-nGR performs slightly worse, which demonstrates that utilizing information among groups may enrich group representation for group recommendation. We can also observe that the gaps between GLIF-nGR and GLIF on MR are larger than those on the CAMRa2011 and MS datasets. Our explanation is that with increasing sparsity, just interactions within a group are too crude and can provide very limited latent information to help infer group preference. Global information may become a good complement. Thus, to address the cold-start problem caused by group-item interaction sparsity, it is better to fully exploit sharing features among groups for group recommendation. It can be observed that GLIF-nGR performs better than GLIF-nLR on all three datasets, which indicates that local information is more influential than global information for group preference on our datasets.

Global and local representation fusion. The primary motivation of this component is to learn variable attention weights for group global and local level representation, rather than the commonly used uniform weighting strategy. We can observe that compared with GLIF-nGLF, GLIF performs better on all three datasets, which indicates that it is meaningful to adopt an attention mechanism to fuse global and local representations. We can also observe that GLIF-nGLF performs worse than GLIF-nGR on the MS datasets. One possible explanation is that adopting a simple average strategy may not mine sharing features to reinforce group representation or even weaken it. However, our model adopts the attention mechanism to learn the attentive and dynamic weights of the group's global and local representations, in which a higher weight indicates that the representation is more important; thus, its contribution is more significant for the group's final representation. Therefore, GLIF can always perform better than GLIF-nGR, while GLIF-nGLF cannot.

Type-based ANN and *K* types-based ANN. Compared to GLIF-T and GLIF-K, GLIF performs better, which validates the effectiveness of the type-based ANN and K types-based ANN. One reason is that GLIF-T can identify important objects (i.e., items, users, and groups), and GLIF-K can mine features that are significant to group recommendations by allocating proper weights to each object and different features.

5.2.3 Hyperparameters sensitivity (RQ3)

We conduct experiments to analyse the impacts of two key parameters, e.g., embedding dimension d and size of sampled group neighbors γ, for each group on three datasets. Fig. 5 depicts the impact of embedding dimensions on three datasets, and Fig. 6 depicts the impact of group neighbor sizes on three datasets.

According to Fig. 5, we find that when d varies from 8

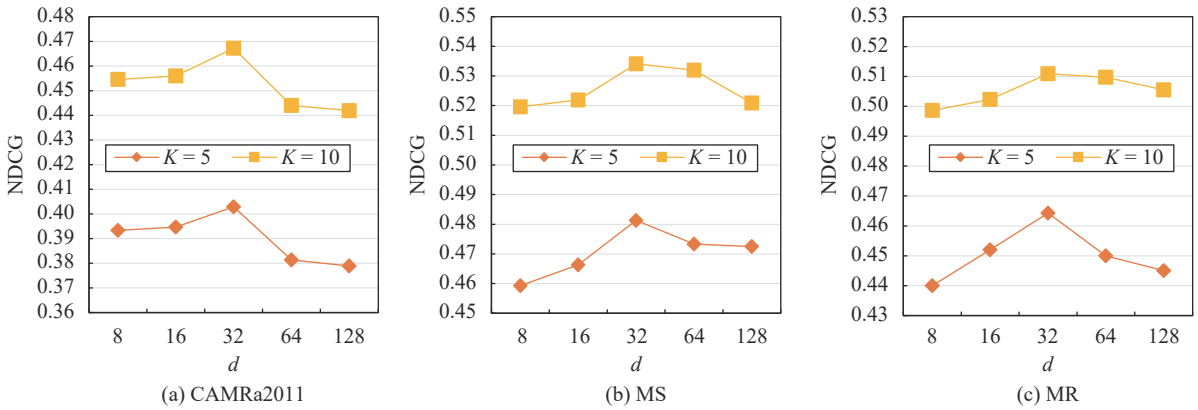


Fig. 5 Impact of embedding dimensions on the three datasets

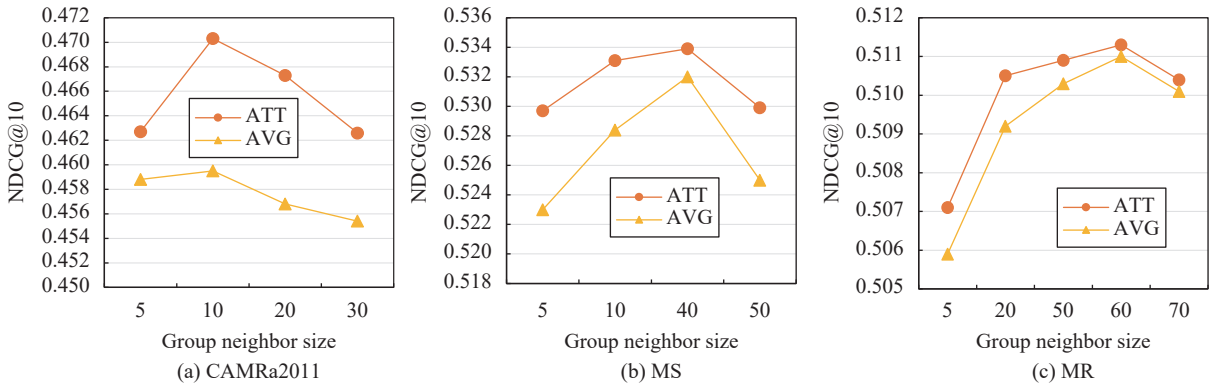


Fig. 6 Impact of group neighbor sizes on three datasets

to 32, NDCGs on the three datasets increase in general since better representations can be learned. However, the performance becomes stable or slightly worse when d further increases. Our explanation is that an oversize embedding dimension leads to an overfitting issue. Therefore, the embedding dimension of GLIF is set to 32.

To investigate the effect of group neighbor size γ on the recommendation performance, group neighbor size was searched in [5, 10, 20, 30, 40, 50, 60, 70]. According to Fig. 6, we can find that: 1) As γ increases, NDCG first rises and then declines. The decline may be due to the decreasing average similarity (i.e., Pearson's correlation coefficient). Low average similarity may be caused by high diversity among groups. In other words, as the neighbor size increases, the preferences of groups become more diverse, and it is more difficult to extract shared tastes from group neighbors. With the diversity of groups increasing, it may involve more noises for calculating group preference. Thus, as γ further increases, increasing noise leads to worse performance. 2) Compared with average-based aggregating strategy, attention-based strategy can always perform better, which demonstrates that utilizing attention mechanism to assign different weights may capture better group representations from global information. 3) Each dataset has different optimal values of γ, which are 10, 40 and 60, respectively.

Note that from dataset CAMRa2011 to MS, then MR, group-item interactions become increasingly sparse. This indicates that the fewer group-item interactions within a group, the more global information GLIF needs to enhance group representation. Therefore, the neighbor sizes are set to 10, 40 and 60 for CAMRa2011, MS and MR, respectively.

Dropout is a very important technique in deep neural networks to avoid overfitting. Fig. 7 shows the NDCG@10 metric on three datasets with different dropout rates. On the three datasets, as the dropout rate increases, the performance increases before reaching its peak. As the dropout rate surpasses 0.2, the performance on all three datasets decreases. Thus, on three datasets, we set the dropout rate to 0.2.

5.2.4 Case studies on attention (RQ4)

Since many existing works mentioned in Section 2 have evaluated attention-based member-aggregating strategy or item-aggregating strategy, we conduct experiments to explore how attention mechanism works on fusing global information (i.e., group neighbors). Thus, we randomly sampled 10 groups for case studies, and each group contained 10 groups.

According to the heatmap depicted in Fig. 8, we have two observations: 1) On the whole, attention weights and PCCs of group neighbors gain similar results, which demonstrates the effectiveness of both methods in finding similar groups. 2) The discrimination of attention

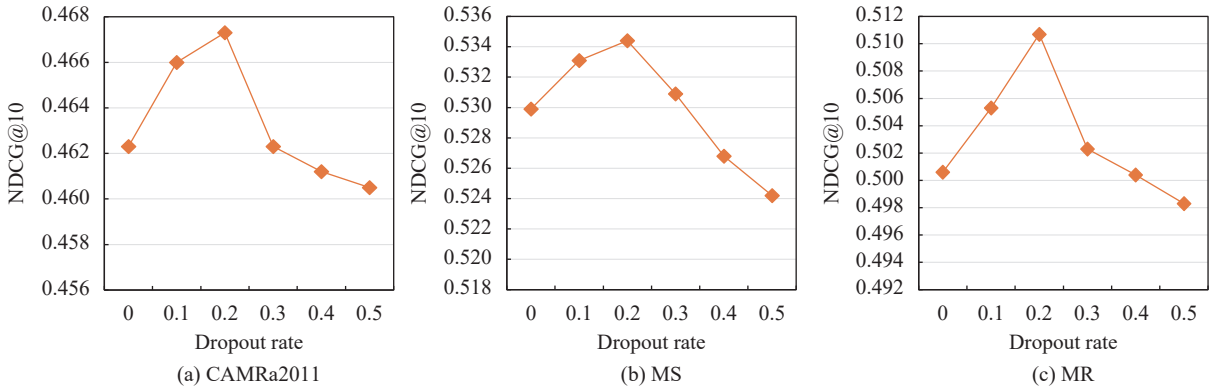


Fig. 7 Impact of dropout rate on three datasets

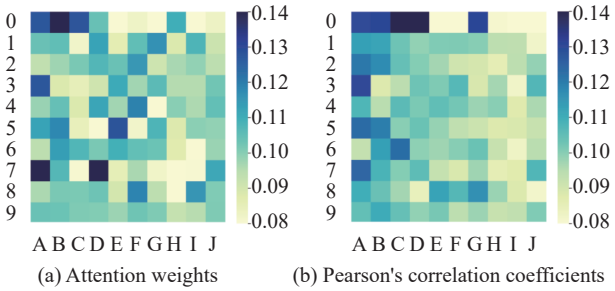


Fig. 8 Visualization for the sampled 10 groups on attention weight and PCC of 10 group neighbors, where the x-axis denotes the relative neighbor-id and the y-axis denotes the group-id. Darker color means larger value.

weights is more obvious than that of PCCs. For example, the neighbors of group 4 have obviously different attention weights but not PCCs′. This is probably because group 4 has few history interactions (76 interaction records), and thus some influential groups (e.g., neighbor G with 1 847 history interactions) may be mistakenly treated as similar groups that have limited contribution to learning the representation of group 4. However, attention can dynamically assign group neighbors′ weights by learning-rich interactions, which shows remarkable superiority.

6 Conclusions and future work

In this paper, we introduce the problems of group recommendation and propose a novel attentive neural network model under a deep learning framework, i.e., GLIF. Specifically, we utilize three kinds of interaction information to generate group local representation and item representation. Furthermore, it learns global-level group representations by finding similar groups and aggregating them under attention mechanism. In addition, considering the effect weights of global and local representation, we adopt ANN to aggregate them. Finally, we model group-item interaction under NCF framework. Extensive experiments on three datasets demonstrate that GLIF can outperform state-of-the-art methods.

In the future, we are interested in exploring social connections for group recommendations since it contains valuable signals on how users trust each other and how users form a group. We may also take contextual information, such as text information and time information, into the model to investigate and improve the explainability of our model as well as achieve online group recommendation.

Acknowledgements

This work was supported by National Natural Science Foundation of China (Nos. 61872363 and 61672507), Natural Foundation of Beijing Municipal Commission of Education, China (No. 21JD0044), National Key Research and Development Program of China (No. 2016YFB 0401202), and the Research and Development Fund of Institute of Automation, Chinese Academy of Sciences, China (No. Y9J2FZ0801).

References

[1] L. Cui, J. Wu, D. C. Pi, P. Zhang, P. Kennedy. Dual implicit mining-based latent friend recommendation. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 5, pp. 1663–1678, 2020. DOI: 10.1109/TSMC.2017.2777889.

[2] S. S. Deng, L. T. Huang, G. D. Xu, X. D. Wu, Z. H. Wu. On deep learning for trust-aware recommendations in social networks. *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 5, pp. 1164–1177, 2017. DOI: 10.1109/TNNLS.2016.2514368.

[3] Z. H. Huang, S. J. E, J. W. Zhang, B. Zhang, Z. L. Ji. Pairwise learning to recommend with both users′ and items′ contextual information. *IET Communications*, vol. 10, no. 16, pp. 2084–2090, 2016. DOI: 10.1049/iet-com.2016.0112.

[4] N. Zheng, S. Y. Song, H. Y. Bao. A temporal-topic model for friend recommendations in Chinese microblogging systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 45, no. 9, pp. 1245–1253, 2015. DOI: 10.1109/TSMC.2015.2391262. (in Chinese)

[5] V. M. Le. Group recommendation techniques for feature modeling and configuration. In *Proceedings of IEEE/ACM*

the 43rd International Conference on Software Engineering: Companion Proceedings, IEEE, Madrid, Spain, pp. 266–268, 2021. DOI: 10.1109/ICSE-Companion52605.2021.00123.

[6] L. V. Nguyen, M. S. Hong, J. J. Jung, B. S. Sohn. Cognitive similarity-based collaborative filtering recommendation system. *Applied Sciences*, vol. 10, no. 12, Article number 4183, 2020. DOI: 10.3390/app10124183.

[7] D. Rafailidis, A. Nanopoulos. Modeling users preference dynamics and side information in recommender systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 46, no. 6, pp. 782–792, 2016. DOI: 10.1109/TSMC.2015.2460691.

[8] L. Hu, J. Cao, G. D. Xu, L. B. Cao, Z. P. Gu, C. Zhu. Personalized recommendation via cross-domain triadic factorization. In *Proceedings of the 22nd International Conference on World Wide Web*, ACM, Rio de Janeiro, Brazil, pp. 595–606, 2013. DOI: 10.1145/2488388.2488441.

[9] S. Rendle. Factorization machines. In *Proceedings of IEEE International Conference on Data Mining*, IEEE, Sydney, Australia, pp. 995–1000, 2010. DOI: 10.1109/ICDM.2010.127.

[10] Y. Koren, R. Bell, C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, vol. 42, no. 8, pp. 30–37, 2009. DOI: 10.1109/MC.2009.263.

[11] Q. Yuan, G. Cong, C. Y. Lin. COM: A generative model for group recommendation. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, New York, USA, pp. 163–172, 2014. DOI: 10.1145/2623330.2623616.

[12] X. J. Liu, Y. Tian, M. Ye, W. C. Lee. Exploring personal impact for group recommendation. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, ACM, Maui, USA, 2012, pp. 674–683. DOI: 10.1145/2396761.2396848.

[13] S. S. Ghaemmaghami, A. Salehi-Abari. DeepGroup: Group recommendation with implicit feedback. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, ACM, Queensland, Australia, pp. 3408–3412, 2021. DOI: 10.1145/3459637.3482081.

[14] Z. H. Huang, X. Xu, H. H. Zhu, M. C. Zhou. An efficient group recommendation model with multiattention-based neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 11, pp. 4461–4474, 2020. DOI: 10.1109/TNNLS.2019.2955567.

[15] D. Cao, X. N. He, L. H. Miao, Y. H. An, C. Yang, R. C. Hong. Attentive group recommendation. In *Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, ACM, Ann Arbor, USA, pp. 645–654, 2018. DOI: 10.1145/3209978.3209998.

[16] Z. X. He, C. Y. Chow, J. D. Zhang, N. Li. GRADI: Towards group recommendation using attentive dual top-down and bottom-up influences. In *Proceedings of IEEE International Conference on Big Data*, IEEE, Los Angeles, USA, pp. 631–636, 2019. DOI: 10.1109/BigData47090.2019.9005686.

[17] L. V. Tran, T. A. N. Pham, Y. Tay, Y. D. Liu, G. Cong, X. L. Li. Interact and decide: Medley of sub-attention networks for effective group recommendation. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, Paris, France, pp. 255–264, 2019. DOI: 10.1145/3331184.3331251.

[18] Z. X. He, C. Y. Chow, J. D. Zhang. GAME: Learning graphical and attentive multi-view embeddings for occasional group recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, China, pp. 649–658, 2020. DOI: 10.1145/3397271.3401064.

[19] A. Said, S. Berkovsky, E. W. De Luca. Group recommendation in context. In *Proceedings of the 2nd Challenge on Context-aware Movie Recommendation*, ACM, Chicago, USA, pp. 2–4, 2011. DOI: 10.1145/2096112.2096113.

[20] L. Hu, S. L. Jian, L. B. Cao, Z. P. Gu, Q. K. Chen, A. Amirbekyan. HERS: Modeling influential contexts with heterogeneous relations for sparse and cold-start recommendation. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*, AAAI, Honolulu, USA, pp. 3830–3837, 2019. DOI: 10.1609/aaai.v33i01.33013830.

[21] C. Y. Yin, L. F. Shi, R. X. Sun, J. Wang. Improved collaborative filtering recommendation algorithm based on differential privacy protection. *The Journal of Supercomputing*, vol. 76, no. 7, pp. 5161–5174, 2020. DOI: 10.1007/s11227-019-02751-7.

[22] X. N. He, L. Z. Liao, H. W. Zhang, L. Q. Nie, X. Hu, T. S. Chua. Neural collaborative filtering. In *Proceedings of the 26th International Conference on World Wide Web*, ACM, Perth, Australia, pp. 173–182, 2017. DOI: 10.1145/3038912.3052569.

[23] J. K. Wang, Y. C. Jiang, J. S. Sun, Y. Z. Liu, X. Liu. Group recommendation based on a bidirectional tensor factorization model. *World Wide Web*, vol. 21, no. 4, pp. 961–984, 2018. DOI: 10.1007/s11280-017-0493-6.

[24] L. Baltrunas, T. Makcinskas, F. Ricci. Group recommendations with rank aggregation and collaborative filtering. In *Proceedings of the 4th ACM Conference on Recommender Systems*, ACM, Barcelona, Spain, pp. 119–126, 2010. DOI: 10.1145/1864708.1864733.

[25] S. Berkovsky, J. Freyne. Group-based recipe recommendations: Analysis of data aggregation strategies. In *Proceedings of the 4th ACM Conference on Recommender Systems*, ACM, Barcelona, Spain, pp. 111–118, 2010. DOI: 10.1145/1864708.1864732.

[26] S. Amer-Yahia, S. B. Roy, A. Chawlat, G. Das, C. Yu. Group recommendation: Semantics and efficiency. *Proceedings of the VLDB Endowment*, vol. 2, no. 1, pp. 754–765, 2009. DOI: 10.14778/1687627.1687713.

[27] L. Boratto, S. Carta. State-of-the-art in group recommendation and new approaches for automatic identification of groups. In *Information Retrieval and Mining in Distributed Environments*, A. Soro, E. Vargiu, G. Armano, G. Paddeu, Eds., Berlin, Germany, Springer, pp. 1–20, 2010.

DOI: 10.1007/978-3-642-16089-9_1.

[28] L. Guo, H. Z. Yin, Q. Y. Wang, B. Cui, Z. Huang, L. Z. Cui. Group recommendation with latent voting mechanism. In *Proceedings of the 36th IEEE International Conference on Data Engineering*, IEEE, Dallas, USA, pp. 121–132, 2020. DOI: 10.1109/ICDE48307.2020.00018.

[29] J. F. McCarthy, T. D. Anagnost. MusicFX: An arbiter of group preferences for computer supported collaborative workouts. In *Proceedings of ACM Conference on Computer Supported Cooperative Work*, ACM, Seattle, USA, pp. 363–372, 1998. DOI: 10.1145/289444.289511.

[30] Z. W. Yu, X. S. Zhou, Y. B. Hao, J. H. Gu. TV program recommendation for multiple viewers based on user profile merging. *User Modeling and User-adapted Interaction*, vol. 16, no. 1, pp. 63–82, 2006. DOI: 10.1007/s11257-006-9005-6.

[31] S. S. Feng, H. X. Zhang, L. Wang, L. Liu, Y. C. Xu. Detecting the latent associations hidden in multi-source information for better group recommendation. *Knowledge-Based Systems*, vol. 171, pp. 56–68, 2019. DOI: 10.1016/j.knosys.2019.02.002.

[32] M. Ye, X. J. Liu, W. C. Lee. Exploring social influence for recommendation: A generative model approach. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, Portland, USA, pp. 671–680, 2012. DOI: 10.1145/2348283.2348373.

[33] L. J. Zhou, J. W. Dang, Z. H. Zhang. Fault classification for on-board equipment of high-speed railway based on attention capsule network. *International Journal of Automation and Computing*, vol. 18, no. 5, pp. 814–825, 2021. DOI: 10.1007/s11633-021-1291-2.

[34] X. Zhang, Q. Yang. Correction to: Transfer hierarchical attention network for generative dialog system. *International Journal of Automation and Computing*, vol. 18, no. 5, Article number 856, 2021. DOI: 10.1007/s11633-020-1223-6.

[35] J. Orbach. Principles of neurodynamics. Perceptrons and the theory of brain mechanisms. *Archives of General Psychiatry*, vol. 7, no. 3, pp. 218–219, 1962. DOI: 10.1001/archpsyc.1962.01720030064010.

[36] L. Guo, H. Z. Yin, T. Chen, X. L. Zhang, K. Zheng. Hierarchical hyperedge embedding-based representation learning for group recommendation. *ACM Transactions on Information Systems*, vol. 40, no. 1, Article number 3, 2021. DOI: 10.1145/3457949.

[37] X. Glorot, Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*, Sardinia, Italy, pp. 249–256, 2010.

[38] M. McPherson, L. Smith-Lovin, J. M. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, vol. 27, pp. 415–444, 2001. DOI: 10.1146/annurev.soc.27.1.415.

[39] H. F. Liu, E. P. Lim, H. W. Lauw, M. T. Le, A. X. Sun, J. Srivastava, Y. A. Kim. Predicting trusts among users of online communities: An epinions case study. In *Proceedings of the 9th ACM Conference on Electronic Commerce*, ACM, Chicago, USA, pp. 310–319, 2008. DOI: 10.1145/1386790.1386838.

[40] L. Hu, J. Cao, G. D. Xu, L. B. Cao, Z. P. Gu, W. Cao. Deep modeling of group preferences for group-based recommendation. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence*, AAAI, Québec City, Canada, pp. 1861–1867, 2014. DOI: 10.5555/2892753.2892811.

[41] H. Z. Yin, Q. Y. Wang, K. Zheng, Z. X. Li, J. L. Yang, X. F. Zhou. Social influence-based group representation learning for group recommendation. In *Proceedings of the 35th IEEE International Conference on Data Engineering*, IEEE, Macao, China, pp. 566–577, 2019. DOI: 10.1109/ICDE.2019.00057.

[42] P. Sedgwick. Pearson's correlation coefficient. *BMJ*, vol. 345, Article number e4483, 2012. DOI: 10.1136/bmj.e4483.

[43] J. Y. Chen, H. W. Zhang, X. N. He, L. Q. Nie, W. Liu, T. S. Chua. Attentive collaborative filtering: Multimedia recommendation with item-and component-level attention. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, Shinjuku, Japan, pp. 335–344, 2017. DOI: 10.1145/3077136.3080797.

[44] S. Rendle, C. Freudenthaler, Z. Gantner, L. Schmidt-Thieme. BPR: Bayesian personalized ranking from implicit feedback. In *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence*, Montreal, Canada, pp. 452–461, 2009.

[45] X. N. He, H. W. Zhang, M. Y. Kan, T. S. Chua. Fast matrix factorization for online recommendation with implicit feedback. In *Proceedings of the 39th International Conference on Research and Development in Information Retrieval*, ACM, Pisa, Italy, pp. 549–558, 2016. DOI: 10.1145/2911451.2911489.

[46] Y. Koren. Factorization meets the neighborhood: A multifaceted collaborative filtering model. In *Proceedings of the 14th International Conference on Knowledge Discovery and Data Mining*, ACM, Las Vegas, USA, pp. 426–434, 2008. DOI: 10.1145/1401890.1401944.

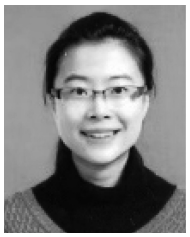
[47] A. M. Elkahky, Y. Song, X. D. He. A multi-view deep learning approach for cross domain user modeling in recommendation systems. In *Proceedings of the 24th International Conference on World Wide Web*, ACM, Florence, Italy, pp. 278–288, 2015. DOI: 10.1145/2736277.2741667.

Song Zhang received the B. Sc. degree in software engineering from Xiamen University, China in 2019. He is currently a master student in computer technology at University of Chinese Academy of Sciences, China.

His research interests include data mining and machine learning.

E-mail: zhangsong2019@ia.ac.cn

ORCID iD: 0000-0001-5396-0003



Nan Zheng received the Ph. D. degree in computer application from Institute of Automation, Chinese Academy of Sciences, China in 2012. She is currently an associate professor at State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, China. She was a visiting scholar at University of California, Berkeley, USA from 2018 to 2019.

Her research interests include data mining and machine learning.

E-mail: nan.zheng@ia.ac.cn
ORCID iD: 0000-0002-8560-9231

Dan-Li Wang received the Ph. D. degree in computer application from Beihang University, China in 1999. She is currently a researcher at State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, China.

Her research interests include complex systems, metasynthesis, group intelligence, human-computer interaction, psychosomatic computation, data mining and machine learning.

E-mail: danli.wang@ia.ac.cn (Corresponding author)
ORCID iD: 0000-0003-0980-5457