Deep pairwise learning for user preferences via dual graph attention model in location-based social networks

Weihua Gong (whgong@zjut.edu.cn)
Zhejiang University of Technology
Kechen Zheng
Zhejiang University of Technology
Shubin Zhang
Zhejiang University of Technology
Ping Hu
Zhejiang University of Technology

Research Article

Keywords: Graph neural network, Location based social networks, Dual graph attentions, User preferences

Posted Date: September 15th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2043370/v1

Deep pairwise learning for user preferences via dual graph attention model in location-based social networks

Weihua Gong[1] • Kechen Zheng[1] • Shubin Zhang[1] • Ping Hu[1]

Abstract—The rapid growth of location-based social networks (LBSN) has generated massive multi-mode entities such as users, locations and social topics, as well as multi-relationships among them. Learning deep user preferences over POIs and social topics in LBSN has attracted more and more attentions. However, most of the previous studies have failed to incorporate the severely sparse and multi-dimensional relational data of heterogeneous LBSN, which plays a vital role to enhance the prediction performance. Toward this challenge, in this paper, we propose a unified pairwise relationship prediction framework based on graph neural networks combined with dual attention mechanism model, dubbed GE2AT, which has fully considered to incorporate three types of user behavioral relations to better capture user preference over different entities. Specifically, we first model LBSN as one user social graph only including social relations, and other two bipartite graphs involved user-location check-in relations and user-social topics following relations respectively. Then, we propose dual graph attention networks to jointly learn deep latent representations of different entities by aggregating features from neighbors with different importance weights. After that, we adopt two MLPs to output the pairwise preferences of users over POIs and social topics. In addition, to alleviate data sparsity problem, on the basis of MLPs output results, we further leverage social relationships as social-aware influence for similar user preferences so as to enhance the pairwise prediction performance. Extensive experiments on two real-world datasets have demonstrated that our proposed framework GE2AT significantly outperform the state-of-the-art baselines on both prediction tasks for user preferences on POIs and social topics.

Keywords—Graph neural network• Location based social networks• Dual graph attentions• User preferences

1 INTRODUCTION

As a newly emerging type of heterogeneous information networks, LBSN has become a research hotspot in recent years. We present an illustrative example for LBSN in Fig.1, which consists of three entity layers, e.g., geographical space layer, user layer and social media layer, containing with numerous users, locations and media topics or tags respectively, wherein there also exist massive multi-types of relations among them, namely user-user social relations, user-location check-in relations, and user-social topics following relationships. Different from traditional social networks, LBSN provides users with multi-dimensional information and services around their current locations, thus enriching both on-line and off-line daily lives. For example, the users in popular LBSNs such as Foursquare, Yelp and Facebook, can not only post and share their tips or daily photos with friends like social networks, but also visit or check-in the recommended points of interest (POIs), e.g., restaurants, venues, scenic spots, etc. However, how to exploit user-centric multi-relational data in heterogeneous LBSN, generating offline POIs and online social topics recommendation is still a great challenge at present.

✉*Weihua Gong*
whgong@zjut.edu.cn

Kechen Zheng
kechengzhen@zjut.edu.cn

Shubin Zhang
zhangshubing@zjut.edu.cn

Ping Hu
pinghu@zjut.edu.cn

[1] *School of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China*

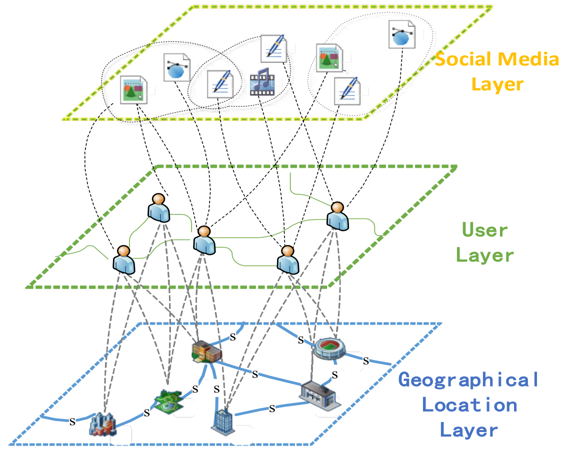


Fig.1. Multilayer model of LBSN with multi-dimensional relations among multi-mode entities (e.g., users, POIs and social topics) such as social relations between users, check-in relations between users and POIs, and following relationships between users and social topics.

As known to all, LBSN is presently facing more serious sparsity problem especially for multi-dimensional relations than traditional social networks, which becomes a great obstacle for the performance improvement of recommendation tasks. The reasons are twofold, on one hand, as for location check-in data, LBSN usually has millions of POIs, while only small number of locations such as scenic spots, restaurants and so on, are most frequently visited by many users, and the number of visited locations by a user usually takes up a very small portion in all available POIs because of the spatial locality of users' daily activities. For instance, the density of check-in data (e.g., in Foursquare or Yelp) is less than 0.1%. On the other hand, as for users' social links, social networks generated by users' friendships generally have small world phenomenon and scale-free characteristics. These laws show that the social links of users obey the power law distribution. In other words, only a few users have more social friends, while many other users only have very few social friends.

Considering these two types of sparse data in LBSN, many previous traditional studies have focused on learning latent features representation based on low-dimensional space embedding via matrix factorization (e.g., GeoMF[1], GraRep[2], and SSTPMF[3]) or random walk based approaches(e.g., DeepWalk[4], LINE[5], and CLoRW[6]). However, these studies have several common limitations. First, the computation complexity will increase exponentially with the dimensionality increasing of factorization or random walks, thus they are not suitable for large-scale graphs or networks. Second, we notice that all these approaches belong to shallow models, which are employed to learn the linear latent features of relational entities in LBSN, instead of real non-linear features among them. Therefore, there still need more efforts to model the deep non-linear embeddings for heterogeneous entities in LBSN, i.e., user preferences on social topics and POIs.

Inspired by recent advancements in deep learning, graph neural networks (GNN) [7-9] have brought a new idea for tackling above problems. The core idea behind GNN is to learn representation by means of iteratively aggregating feature information from local graph neighbors using neural networks. As expected, GNN-based models have demonstrated powerful representation capability on homogeneous graphs. For example, Hamilton *et al.*[8] introduce a framework called GraphSAGE, which uses neural networks, e.g., LSTM, to aggregate neighbors' feature information. Despite of achieving great success, GNN-based methods are insufficient to identify varying importance of different neighbors or modalities, hindering the information propagation and resulting in suboptimal representations. To make up this defect, attention mechanism[10] is furtherly introduced to learn different importance for nodes and their neighbors. For example, Wu *et al.*[11] propose DANSER with dual graph attention networks for user-based and item-based embeddings to improve recommendation performance. Besides, Yan *et al.*[12] employ self-attention mechanism to integrate multiple metapath-based network embeddings for personalized recommendation. Overall, most existing GNN-based models have not fully considered for fusing different types of nodes and relations in heterogeneous graph. Thereby, the challenge of how to jointly capture the complex non-linear representations for various entities with multi-typed relations especially in LBSN is still far from to be resolved.

To address above issues, we propose a unified pairwise relationship prediction framework based on dual graph attention networks, called GE2AT, to learn deep representations of user preferences in LBSN. In our framework, we consider to combine three types of user centric relations in LBSN such as user's check-in relations to locations, following relations on social topics, and social friendships. Specifically, we first model two types of bipartite graphs for pairwise entities, i.e., users and locations, users and social topics. Then, we introduce the attention mechanism into two bipartite graphs to generate dual graph attention networks so as to jointly learn the representations of user preferences, where one is modeled the attentive weights from user-location bipartite graph, and the other is modeled from user-social topics bipartite graph. Finally, we adopt multilayer perceptron (MLP) to predict two types of pairwise relationships, i.e., user-to-POIs relations and user-to-social topics relations. Meanwhile, we further enhance the prediction performance by fully taking users' social influence into account, being combined with the output of MLPs.

The contributions of this paper can be summarized as follows.

1) We propose a unified pairwise relationship prediction framework called GE2AT, to learn non-linear latent representations of heterogeneous entities by utilizing GNN combined with dual graph attentions in heterogeneous LBSN.

2) We further exploit the social influence to improve the prediction performance on the basis of the output of MLPs, which makes the users with social relations have similar preferences on both POIs and social topics as much as possible.

3) In our framework, we consider to jointly capture user preferences by fully incorporating three types of user behavioral relations in LBSN, namely the check-in relations between users and locations, the following relations between users and social topics, and the social relationships among users.

4) We conduct extensive experiments on two real-world datasets. The experimental results show that our proposed framework GE2AT significantly outperforms several state-of-the-art methods on both prediction tasks of user preferences on POIs and social topics.

The remainder of this paper is organized as follows. In Section 2, the related work is reviewed. Section 3 formulates the problem, and then the details of our proposed framework are presented in Section 4. The experimental results are presented in Section 5. Finally, Section 6 concludes this paper and outlines our future work.

2 RELATED WORK

In this section, we review related studies in the field of LBSN, namely POI recommendation and neural network-based recommendation.

2.1 POI Recommendation

So far, POI recommendation has become one popular focal point in LBSN. Most prior studies employ the matrix factorization (MF) based collaborative filtering methods to alleviate data sparsity problem and provide better POI recommendation performance. For example, Lian *et al.*[1] propose a geographical weighted matrix factorization model, dubbed GeoMF, that integrated geographical influence by modeling users' activity regions and the influence propagation on geographical space. Similarly, Griesner *et al.*[13] also present an approach integrating temporal and geographic influences into matrix factorization. However, these models have ignored to consider fusing context information. To combine more influential factors, a ranking-based matrix factorization model is proposed by Li *et al.*[14] for learning users' preference rankings for locations, fusing with the geographical influence scores of neighboring locations and the temporal influence. Zhang *et al.*[15] utilize a fused matrix factorization framework to take into account the social, multi-tag and geographic influences to enhance POI recommendation performance. Lyu *et al.*[16] adopt weighted matrix factorization method with consideration of multi-information for travel location recommendation. Besides, there are other alternative studies such as Davtalab *et al.* [3] and Qian *et al.*[17] integrate user similarity, social influence and POI similarity into a unified probabilistic matrix factorization(PMF) framework for POI recommendation.

In particular, to extend matrix model to high-order space, newly tensor factorization approaches are exploited to incorporate more types of influential factors such as geographical influence, temporal influence, social influence and content influence. For example, Yao *et al.*[18] present a collaborative filtering POI recommendation method based on nonnegative tensor factorization and fuse users' social relations as regularization terms of the factorization to improve the recommendation accuracy. Yang *et al*.[19] design a personalized context-aware location recommendation model STAP, which applies nonnegative tensor factorization to collaboratively infer users' spatial temporal activity preference. Zheng *et al*.[20] leverage ranking-based collective tensor factorization techniques to recommend both POIs and activities. Similarly, for another variant of POI recommendation, Zhao *et al.*[21] propose a ranking-based pairwise tensor factorization framework for successive POI recommendation. He *et al*.[22] further present a tensor-based latent model considering the influence of user's latent behavior patterns, which is determined by the contextual temporal and categorical information.

In summary, we argue that most existing MF-based models leveraging the dot product as the similarity measure are insufficient to capture the real non-linear latent features of multi-mode entities in LBSN and further alleviate the data sparsity issue due to their limited generalization.

2.2 Neural Network Based Recommendation

Recently, there are a bunch of methods using neural networks for various recommendations in LBSN, which may bring new opportunities to promote better recommending performance. One research line is to utilize Recurrent Neural Networks (RNN) to cope with sequential and temporal mobility data in LBSN, some researchers adopt the RNN model to derive a deep representation for the check-in visiting trajectory for next POI prediction. For instance, Liu *et al.*[23] propose a RNN-based architecture to model spatial temporal contexts so as to get more accurate location prediction. Liao *et al.*[24] present multi-task context aware RNN to leverage the spatial activity topic for both activity and location prediction, which considers to integrate the temporal context information and sequential pattern. More recently, several variants of RNN have been investigated to overcome its inherent limitations. For example, Zhao *et al.*[25] extend RNN model and introduce an innovative gate mechanism into LSTM architecture to learn user' sequential visiting behaviors for the next POI recommendation. Huang *et al.*[26] develop a RNN-based network architecture called ST-LSTM for next POI recommendation, to jointly model the spatiotemporal contextual information derived from LBSNs. Similarly, Yang *et al.*[27] also present a novel neural network model to provide both location recommendation and friend recommendation, which employed the RNN and Gated Recurrent Unit (GRU) models to characterize short-term and long-term sequential contexts separately. Moreover, Doan *et al.*[28] design a deep long short term memory RNN model with attention mechanism to capture both the sequential, and temporal/spatial characteristics into its learned representations for successive POI recommendation. Recently, Wang *et al.*[29] propose a novel next-item recommendation model which leverages interval and duration gated LSTM to accurately capture users' long-term and short-term preferences. Although the RNN and its variants are quite effective in sequential modeling problems, they lack the ability of fusing multi-dimensional relations in heterogeneous LBSN.

Apart from above relevance, another research line is to adopt Graph Neural Network (GNN) to learn meaningful representations for graph-structured data. Different from RNNs, GNN based models can be straightforwardly applied to model a set of nodes and their relationships, where there isn't a natural order of nodes in the graph. Recent advances on GNN frameworks have demonstrated superior performance for traditional recommending tasks. Some state-of-the-art methods such as Fan *et al.*[30] propose a graph network model GraphRec+ to learn better user and item representations with attention mechanism for social recommendations, while this model only incorporates the social graph into recommendation, ignoring other associated rich side information on users as well as items. Wang *et al.*[9] also develop a new recommendation framework based on GNN, called NGCF, which exploits the user item graph structure by propagating embeddings on it. As for other variants of GNN, Wei *et al.*[31] design a Multimodal Graph Convolution Network (MMGCN) framework built upon the message-passing idea of GNN for personalized micro-video recommendation, which could yield modal-specific representations of users and micro-videos to better capture user preferences. Ying *et al.*[32] introduce a random walk based GCN algorithm called PinSage for a variety of recommendation tasks, which combines efficient random walks and graph convolutions to generate embeddings of nodes that incorporates both graph structure as well as node feature information. Moreover, Liu *et al.*[33] design a GNN-based model PIGAT to extract user-item interaction representations, wherein the attention mechanism is introduced to capture two types of the importance of each interacted user or item to improve the prediction performance of content-aware recommendation. Recently, Du *et al.*[34] propose a new unified framework based on hierarchical multi-head self-attention network, which fuses both long and short-term preferences of users for better next item recommendation. While regarding to the recommendation of LBSN, so far only a few studies have paid attention to this promising area, e.g., Zhang *et al.*[35] propose an integrated framework named Context Graph Attention (CGA) that could model heterogeneous context information with the attention mechanism, where model CGA used two context-aware attention networks to learn the different influence of friends and neighboring POIs.

To sum up, due to the heterogeneity of LBSN, most existing GNN based methods have the limitations of either dropping node attributes, discarding rich multi-relations or only utilizing specific meta-paths. There is still big room for improvement by exploiting more comprehensively the information embedded in heterogeneous graphs. Thereby, in this paper, we propose an innovative GNN-based framework with dual attention mechanism to learn the latent representations of heterogeneous LBSN for better recommendation performance.

3 PROBLEM DEFINITION AND PRELIMINARIES

In this section, we first introduce the definitions and notations used in this paper, and then present the problem formulation to be addressed in LBSN.

3.1 Notations

The key notations and their descriptions used in this paper are summarized in Table 1.

Table 1. Notations of the parameters used in our model.

Symbol	Description
U	Set of users
P	Set of locations or POIs
R	Set of social topics
$\mathbf{S}$	User social relation matrix
$\mathbf{u}_i$	User embedding vector
$\mathbf{u}_i^*$	Aggregated user embedding vector
$\mathbf{p}_j$	POI embedding vector
$\mathbf{r}_k$	Social topic embedding vector
e_{ij}	Attentive weight for an edge between vertex i and j
$\mathbf{h}_{\mathrm{N}(i)}$	Aggregated neighborhood representation of vertex i
$N_{(i)}$, $M_{(i)}$	The first-order neighboring sets interacted with vertex *i*
α_{ij}, β_{jk}, η_{ij}, μ_{kj}	Normalized attention weights between pair vertices
$D_{(o)}$	The adjacent set of social friends connected with user o
y_{ij}, y'_{ij}	The output results of MLPs

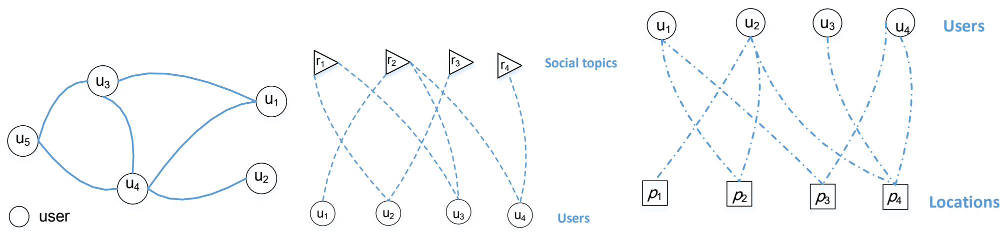
z_{oj}, z'_{oj}	Social-aware prediction of user preferences

For simplicity, in the light of Fig.1, the heterogeneous LBSN intuitively consists of three layers involved with multi-types of entities and relations. Aiming to handle graph structure data easily by our proposed model, we further detach heterogeneous LBSN into three individual graphs according to the relationship types as shown in Fig.2, such as user social links graph, user-social topics bipartite graph, and user-locations bipartite graph. Some fundamental definitions are defined as follows.

Definition 1. ***Social links graph*****:** is an undirected graph that describes the social links of users, denoted as $Gs=(U,V)$, where U is set of users, and $V=\{(u_i,u_j)\,|\,u_i,u_j\in U\}$ is the set of edges with social friendships among users. Here, this graph structure can be easily expressed as adjacent matrix by $\mathbf{s}\in\mathbb{R}^{n\times n}$, if u_i and u_j are social friends, then s_{ij}=1, otherwise it equals 0.

Definition 2. ***User-social topics graph*****:** is a bipartite graph which indicates the following relations between users and social topics. This graph is denoted as $\mathrm{Gt}=(U\cup \mathrm{R},\mathrm{Er})$, where U and R are the sets of users and social topics respectively, and an edge e=$(u_{\mathrm{i}},r_{\mathrm{j}})\in \mathrm{Er}$ represents that user u_{i} follows or interests specific social topic r_{j}.

Definition 3. ***User-locations graph*****:** is another bipartite graph which indicates the relations between users and locations or POIs from their check-in history. This graph is denoted as $\mathrm{Gp}=(U\cup P,\mathrm{Ep})$, where U denotes the same user set, and P is the set of locations respectively, and an edge e=$(u_{\mathrm{i}},p_{\mathrm{j}})\in \mathrm{Ep}$ indicates that user u_{i} has checked in specific location p_{j}.



(a) social links graph (b) User-social topics bipartite graph (c) User-location bipartite graph

Fig.2. A toy example of LBSN model decomposed into one social links graph and two bipartite graphs, note that all users in the three graphs refer to the same users in LBSN. (a) social links graph only involving social relations between users (b) user-social topics graph is one bipartite graph contained with following relations between users and social topics (c) user-locations graph is another bipartite graph contained with check-in relations between users and POIs or locations

3.2 Problem Formulation

In the real world, we observe that the users with social or friendship relations have mutual influences to each other's interests or preferences, in other words, similar social users may have similar interests or preferences. Under this premise, given user social links graph, user-location bipartite graph, and user-social topic bipartite graph, we aim to exploit these input graph data to generate high-quality embeddings of users, POIs, and social topics. Specifically, the further problem we study in heterogeneous LBSN is how to predict the pairwise values for user-centric preferences from the following two aspects.

Task1: Pairwise learning for user preferences on POIs in bottom layer of LBSN

Given three types of relational graph data in LBSN as illustrated in Fig.2, our goal is to output the predicting pairwise results for a target user to preferred POIs.

Task2: Pairwise learning for user preferences on social topics in upper layer of LBSN

Likewise, the other goal is to output the predicting pairwise results for a target user to preferred topics.

After obtaining the predicted pairwise vector values, we can use them to score all POIs and topics for target users, and select the top-ranked POIs and topics as the recommendation results for them.

4 MODEL FRAMEWORK

4.1 Architecture Overview

In this section, we formally present a unified pairwise prediction framework, called GE2AT, which is based on graph neural network with dual attention mechanism. The overview framework of GE2AT model is illustrated in Fig.3, consisting of four components: (1) input layer that offers the initialization of three types of user-centric graph data, i.e., social links graph, user-location bipartite graph, and user-social topic bipartite graph;(2) embedding layer learns the representations by graph neural network with dual attention mechanism;(3) hidden layer performs latent features modeling through parallel MLPs, to fuse pairwise interaction information for user multi-behaviors;(4) prediction layer outputs the predicting pairwise values of user preferences.

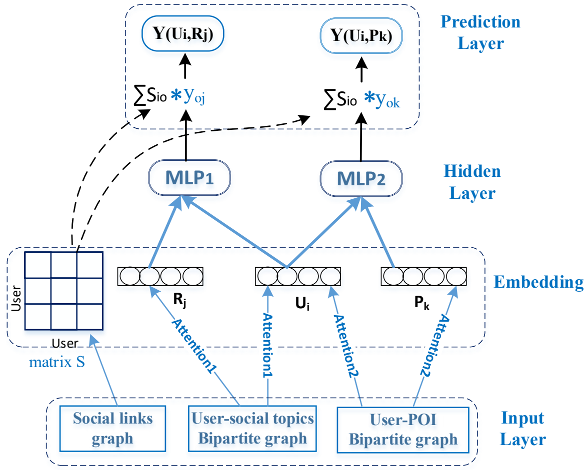


Fig.3.Pairwise prediction framework combined with social influence and graph embeddings with dual attention mechanism. In the framework, Attention1 represents one type of attention from user-social topics bipartite graph. Attention2 represents the other type of attention from user-POI bipartite graph. Social links graph as in Fig2.(a), user-social topics bipartite graph as in Fig2.(b), and user-locations bipartite graph as in Fig2.(c).

4.2 Embedding Layer

In our framework, the embedding layer aims to learn the embeddings of various entities in LBSN through Graph Attention Networks (GAT). To this end, it first requires to model three types of graph structure data from the bottom input layer respectively. More specifically, as for the social links graph, we utilize the adjacent matrix $\mathbf{S} \in \mathbb{R}^{n \times n}$ to express the users' social embeddings, where the row vector $\mathbf{S}_i$ indicates the relation links of some user u_i. As we argue that different social users have different contributions to predict the output results of user's preference on the top prediction layer, where the contribution weights are dependent on his social neighbors or friends. Thereby, we take social-aware influences into account and further discuss their effects for user preferences in *subsection 4.4.*

Next, as for the latter two bipartite graphs, e.g., user-POIs bipartite graph and user-social topics bipartite graph are employed to generate interactive embeddings. To be specific, we utilize dual GATs to combine all the feature representations by aggregating the propagated information from neighbors. Firstly, regarding user-POIs bipartite graph as one GAT with considerations of the check-ins relations between users and POIs, denoted as UL-GAT, we use an embedding vector $\mathbf{u}_i \in \mathbb{R}^d$ to denote a user i and another embedding vector $\mathbf{p}_j \in \mathbb{R}^d$ to represent a POI j, where d is the length of embedding vector. To model user and POI embeddings, let $N_{(i)}$ and $M_{(j)}$ be the first-hop neighbor sets of user i and POI j respectively. We perform embedding propagation between the connected user and POI pairs and aggregate their neighbors combined with attentive weights. Hence, we define the following functions as:

$$\mathbf{u}_i^{(l)} = \mathrm{g}(\mathbf{h}_{\mathrm{N}_{(i)}}^{(l-1)}, \mathbf{u}_i^{(l-1)}) \tag{1}$$

$$\mathbf{p}_j^{(l)} = \mathrm{g}(\mathbf{h}_{M_{(j)}}^{(l-1)}, \mathbf{p}_j^{(l-1)}) \tag{2}$$

where $\mathrm{g}(\cdot)$ is the aggregation function to be specified, l is the embedding propagation layers, a user (or POI) is capable of receiving the messages propagated from its l-hop neighbors. $\mathbf{h}_{\mathrm{N}_{(i)}}^{(l-1)}$ denotes the implicit neighbor-based representations of user i. Likewise, $\mathbf{h}_{M_{(j)}}^{(l-1)}$ denotes the implicit neighbor-based representations of POI j. Specifically, the classical aggregation function for $\mathrm{g}(\cdot)$ is usually adopted LeakyReLU(.) as the nonlinear activation function. In such aggregation, we assume that different neighbors would have the different contributions to the representations of users or POIs, namely, user i is influenced unequally by the attentive weights of his neighbors. Formally, we rewrite above two formulations as

$$\mathbf{u}_i^{(l)} = \mathrm{LeakyReLU}(\mathbf{W}^{(l-1)} \cdot \sum_{j \in N_{(i)}} \alpha_{ij}^{(l-1)} \mathbf{p}_{ij}^{(l-1)} + \mathbf{u}_i^{(l-1)}) \tag{3}$$

$$\mathbf{p}_j^{(l)} = \mathrm{LeakyReLU}(\mathbf{W}'^{(l-1)} \cdot \sum_{k \in M_{(j)}} \beta_{jk}^{(l-1)} \mathbf{u}_{jk}^{(l-1)} + \mathbf{p}_j^{(l-1)}) \tag{4}$$

where $\mathbf{W}$ and $\mathbf{W}'$ are the trainable transformation matrices, and α_{ij} denotes the user-based attention weight interacted with different neighboring POIs of user *i*, β_{jk} denotes the POI-based attention weight interacted with different neighboring users of POI *j*. Both $\mathbf{u}_i^{(l\text{-}1)}$ and $\mathbf{p}_j^{(l\text{-}1)}$ are the representations after (*l*-1) propagation steps, storing the information from their (*l*-1)-hop neighbors.

Hereafter, we introduce the self-attention method to obtain the above attentive weights in Eq.(3) and Eq.(4), which reflects the affinity between the interactive user and POI pairs. For simplicity, we adopt a two-layer neural network called as the attention network to learn the varying importance for each edge with pair of vertices, taking the attention weight of an edge e_{ij} defined as follows

$$e_{ij} = \mathbf{W}_2^{\mathrm{T}} \cdot \vartheta(\mathbf{W}_1 \cdot (\mathbf{W}_s \mathbf{u}_i \oplus \mathbf{W}_s \mathbf{p}_j) + \mathbf{b}_1) + \mathbf{b}_2 \tag{5}$$

where $\vartheta(\cdot)$ is the nonlinear activation function, and $\oplus$ denotes the concatenation of two vectors. The parameters $\mathbf{W}_1$ and $\mathbf{W}_2$ are weight matrices of the attention network, $\mathbf{b}_1$ and $\mathbf{b}_2$ are bias vectors of the attention network, while $\mathbf{W}_s$ is the shared transformation matrix that maps the users and POIs features to the same comparable space. It is worth noting that e_{ij} indicates the importance of an edge with pair vertices is symmetric, i.e., user *i* to POI *j* and vice versa.

More formally, the final user-based or POI-based attention weights are obtained by employing the *softmax* function to normalize above self-attention weights across all neighboring edges, which is formulated as follows:

$$\alpha_{ij} = \mathrm{softmax}(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in \mathrm{N}_{(i)}} \exp(e_{ik})} \tag{6}$$

$$\beta_{jk} = \mathrm{softmax}(e_{jk}) = \frac{\exp(e_{jk})}{\sum_{t \in \mathrm{M}_{(j)}} \exp(e_{jt})} \tag{7}$$

Secondly, we consider the user-social topics bipartite graph as another GAT with considerations of the following relations between users and social topics, denoting as UT-GAT. In order to better learn the embeddings for users and social topics, let $\tilde{\mathbf{u}}_t \in \mathbb{R}^d$ and $\mathbf{r}_k \in \mathbb{R}^d$ denote the embedding vectors of arbitrary user and social topic in UT-GAT respectively, where *d* is the length of embedding vector. Likewise, we then proceed to propagate and aggregate the embeddings of interacted users and social topics, as well as their neighbors. Analogously, we define the following functions as:

$$\tilde{\mathbf{u}}_t^{(l)} = \mathrm{LeakyReLU}(\mathbf{W}^{(l-1)} \cdot \sum_{j \in N_{(t)}} \eta_{tj}^{(l-1)} \mathbf{r}_{tj}^{(l\text{-}1)} + \tilde{\mathbf{u}}_t^{(l\text{-}1)}) \tag{8}$$

$$\mathbf{r}_k^{(l)} = \mathrm{LeakyReLU}(\mathbf{W}'^{(l-1)} \cdot \sum_{j \in M_{(k)}} \mu_{kj}^{(l-1)} \tilde{\mathbf{u}}_{kj}^{(l\text{-}1)} + \mathbf{r}_k^{(l\text{-}1)}) \tag{9}$$

where $N_{(t)}$ and $M_{(k)}$ represent the first-hop neighbor sets of user t and social topic *k* respectively, η_{tj} denotes the neighbor-based attention weight of user *t*, μ_{kj} denotes the neighbor-based attention weight of social topic *k*. Analogously, to obtain these attentive weights, we also employ the *softmax* function to normalize the self-attention weights across all neighboring edges of specific user or social topic as following.

$$\eta_{tj} = \mathrm{softmax}(e_{tj}) = \frac{\exp(e_{tj})}{\sum_{k \in \mathrm{N}_{(t)}} \exp(e_{tk})} \tag{10}$$

$$\mu_{kj} = \mathrm{softmax}(e_{kj}) = \frac{\exp(e_{kj})}{\sum_{i \in \mathrm{M}_{(k)}} \exp(e_{ki})} \tag{11}$$

Note that as shown in Eq.(3) and Eq.(8), all users in UL-GAT actually refer to the same users in UT-GAT, thus we consider to combine the both user embeddings from UL-GAT and UT-GAT into one unified representation. To this end, we utilize the information being propagated from different neighbors along both bipartite graph structures to update the representations of nodes. Namely for each user u_i, let h_{Ni} and h_{Mi} denote his two types of neighbor representations, i.e., locations and social topics, which are recursively propagated from different modalities. As illustrated in Fig.4, in order to encode local graph structural information into the deep user representations, we perform embedding propagation from the first-order to high-order neighbors, and then non-linearly aggregate different types of neighborhood representations. Technically, the aggregation function is employed not only to transform location-based and social topic-based embedding space to the same latent space, but also to make different kinds of neighbors have varying contributions to the aggregated user embeddings.

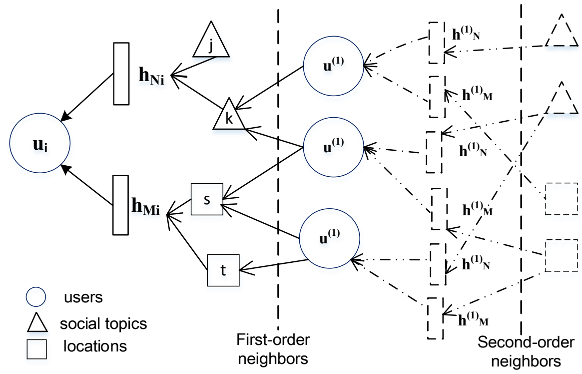


Fig.4. Aggregating two types of user neighbors from UL-GAT and UT-GAT. Given a sampled user u_i , h_{Ni} denotes his one type of neighbors' representations from bipartite graph UT-GAT, and h_{Mi} is his the other type of neighbors' representations from bipartite graph UL-GAT.

Thereby, the user embeddings of Eq.(3) and Eq.(8) can be merged into the aggregated user embedding formally defined as

$$\mathbf{u}_i^{*(l)} = \mathrm{Aggre}(\mathbf{h}_{\mathrm{N}_{(i)}}^{(l)}, \mathbf{h}_{\mathrm{M}_{(i)}}^{(l)}) = \mathrm{Aggre}(\mathbf{u}_i^{(l)}, \tilde{\mathbf{u}}_i^{(l)})$$
$$= \mathrm{LeakyReLU}(\mathbf{W}^{(l-1)} \cdot \sum_{j \in N_{(i)}} (\alpha_{ij}^{(l-1)} \mathbf{p}_{ij}^{(l-1)} + \eta_{ij}^{(l-1)} \mathbf{r}_{ij}^{(l-1)}) + \mathbf{u}*_i^{(l-1)}) \tag{12}$$

In this way, the embedding layer will finally yield three types of interactive embeddings for users, POIs and social topics corresponding to the Eq(4), Eq(9) and Eq(12) respectively.

4.3 Hidden Layer

This layer aims to leverage two parallel MLPs to predict the initial pairwise relationships such as users to POIs, as well as users to social topics. By default, each standard MLP is a stack of fully connected layers, which enables the model to learn the high-order latent interaction features among multiple entities in a hierarchical and non-linear manner. As shown in Fig.3, given two MLPs denoted as $\mathrm{MLP1}(\mathbf{u}_i^*, \mathbf{r}_k)$ and $\mathrm{MLP2}(\mathbf{u}_i^*, \mathbf{p}_j)$, are used to output two types of pairwise relationships for user preferences respectively. More specifically, to predict the pairwise values for users and POIs, we use one MLP, namely MLP2 to concatenate the original user embedding $\mathbf{u}_i^*$ and POI embedding $\mathbf{p}_j$, and the concatenation vector is generated as $[\mathbf{u}_i^* \oplus \mathbf{p}_j]$, then we feed it into multiple hidden layers of MLP for pairwise prediction as

$$\begin{aligned}
\mathbf{h}_1 &= [\mathbf{u}_i^* \oplus \mathbf{p}_j] \\
\mathbf{h}_2 &= \sigma(\mathbf{W}_2 \cdot \mathbf{h}_1 + b_2) \\
&\cdots \\
\mathbf{h}_{l'} &= \sigma(\mathbf{W}_{l'} \cdot \mathbf{h}_{l'-1} + b_{l'}) \\
y_{ij} &= \phi(w^{\mathrm{T}} \cdot \mathbf{h}_{l'})
\end{aligned} \tag{13}$$

where $\sigma(\cdot)$ is the activation function, we empirically adopt the rectified linear unit $\mathrm{ReLU}(x)$. l' is the index of a hidden layer, $\mathbf{W}$ and $\mathbf{b}$ denote the weight matrix and bias vector, and y_{ij} is the initial predicted value for user preference of user $\mathbf{u}_i^*$ to POI $\mathbf{p}_j$. We use sigmoid function $\phi(x) = 1/(1+e^{-x})$ to output the final prediction value.

Analogously, while predicting the pairwise values for users and social topics, we use another MLP, namely MLP1 to concatenate the input embeddings of user and social topic vectors, denoted as $[\mathbf{u}_i^* \oplus \mathbf{r}_k]$, and then feed it into multiple hidden layers for pairwise prediction as

$$\mathbf{h}'_1 = [\mathbf{u}_i^* \oplus \mathbf{r}_k]$$

$$\mathbf{h}'_2 = \sigma(\mathbf{W}'_2 \cdot \mathbf{h}'_1 + b'_2)$$

$$\cdots$$
$$\mathbf{h}'_{l'} = \sigma(\mathbf{W}'_{l'} \cdot \mathbf{h}'_{l'-1} + b'_{l'}) \tag{14}$$
$$y'_{ik} = \phi(w'^{\mathrm{T}} \cdot \mathbf{h}'_{l'})$$

where $\sigma(\cdot)$ uses the same function as above. $\mathbf{W}'$ and $\mathbf{b}'$ also denote the weight matrix and bias vector, and y'_{ik} is the initial predicted value for user preference of user $\mathbf{u}_i^*$ to social topic $\mathbf{r}_k$.

4.4 Prediction Layer

After obtaining two types of initial pairwise predictions from the hidden layer, the last prediction layer proceeds to assemble these prediction results by imposing social influence of neighbors on them. As is well known that a user's preference is similar to or influenced by his/her directly connected social friends according to sociology theory. Thereby, we propose to infer the social-aware based user preferences for the target user which is indirectly dependent on his social neighbors. Towards this end, on the basis of the output predictions of $\mathbf{y}_{ij}$ and $\mathbf{y}'_{ik}$ as stated from Eq.(13) and Eq.(14), we further take the social influence of user neighbors into consideration, assume z_{oj} and z'_{ok} denote the preferences of a target user u_o over POI p_j and social topic r_k respectively. Then, the formulized social-aware preferences of this user determined by his social neighbors are given as the following.

$$\mathbf{z}_{oj} = \sum_{i \in D_{(o)}} \mathbf{S}_{oi} * \mathbf{y}_{ij} \tag{15}$$

$$\mathbf{z}'_{ok} = \sum_{i \in D_{(o)}} \mathbf{S}_{oi} * \mathbf{y}'_{ik} \tag{16}$$

where $D_{(o)}$ denotes the friendship set of target user u_o with social links, and $\mathbf{S}$ represents the social links matrix. $\mathbf{y}_{ij}$ is the prediction result of neighboring user u_i to POI p_j, likewise $\mathbf{y}'_{ik}$ is the prediction result of neighboring user u_i to social topic r_k. The intuition behind Eq.(15) and Eq.(16) reflects that similar social users are prone to have the same preferences over POIs and social topics.

4.5 Model Training

To estimate model parameters of our GE2AT, we need to specify an objective function to optimize. As our model outputs two types of pairwise prediction results for user's preferences, i.e., a target user to POIs, and to social topics, we consider to jointly train the model by optimizing the sum of two loss functions as follows.

$$Loss = \lambda_1 \cdot \mathrm{J}_{(\mathrm{u,p})} + (1-\lambda_1) \cdot \mathrm{J}_{(\mathrm{u,r})} \tag{17}$$

where $\mathrm{J}_{(\mathrm{u,p})}$ denotes the predicting loss of user's preferences over POIs, and $\mathrm{J}_{(\mathrm{u,r})}$ denotes the predicting loss of user's preferences over social topics. While λ_1 is the weight factor to control the trade-off among the two objectives. For simplicity, we would like to fix this parameter λ_1=0.5 in our experiments, that means the same loss weight of user's preferences on POIs and social topics. Moreover, both $\mathrm{J}_{(\mathrm{u,p})}$ and $\mathrm{J}_{(\mathrm{u,r})}$ are corresponding to the commonly used log-loss function which is formally defined by

$$\begin{aligned} \mathrm{J} = \log p(\mathrm{X}^+ \cup \mathrm{X}^- \mid \Theta_e, \Theta_m) &= -\sum_{(i,j)\in \mathrm{X}^+} \log y'_{ij} - \sum_{(i,j)\in \mathrm{X}^-} \log(1-y'_{ij}) \\ &= -\sum_{(i,j)\in \mathrm{X}^+ \cup \mathrm{X}^-} y_{ij} \log y'_{ij} + (1-y_{ij}) \log(1-y'_{ij}) \end{aligned} \tag{18}$$

In above equation, X^+ is a set of positive samples from the dataset, and X^- is the set of negative samples. Θ_e denotes the parameters of the embedding layer for two attention networks included in Eq.(3)~(12), and Θ_m represents the parameters of hidden layer for MLPs included in Eq.(13) and (14). In addition, y_{ij} and y'_{ij} are the actual and prediction of preference of target user u_i over POI or social topic p_j. We would like to minimize the discrepancy of the pairwise loss function so as to obtain the optimized model parameters.

To optimize the combined loss function, we adopt the mini-batch Adam[37] optimizer to train our model by adjusting the learning rate and parameters until the overall loss converges or is sufficiently small. When we perform mini-batch training, where each mini-batch contains both user-POIs and user-social topics graph data. There are three types of embeddings in our model, including user embedding, location embedding, and social topic embedding. In our implementation, the model parameters for embeddings in dual GATs and initial hidden states in MLPs are randomly initialized with Gaussian distribution, where the mean and the standard deviation is 0 and 0.1, and the output layer's parameters are set to follow uniform distribution. More precisely, we set batch size to 512, embedding size to 256, and the number of neighborhood propagation layers to 3. Besides, the MLP layer is set to contain 3 layers with hidden size 256-128-64-1, respectively. In order to avoid additional over-fitting when training the model, the dropout strategy has been applied to our MLPs. The dropout is to randomly drop some neurons with the probability tuned in {0.0,0.1,0.2,0.3}.

5 PERFORMANCE EVALUATION

In this section, we conduct intensive experiments to evaluate the effectiveness of our proposed framework GE2AT in comparison with the state-of-the-art baseline models. All the models are implemented by Python3.6 running on Windows 10 with Intel Xeon4210 CPU, 64GB RAM and GTX1080 GPU.

5.1 Experimental Settings

1) Datasets

We choose two publicly available LBSN datasets, i.e., Foursquare(NYC) and Yelp. Before conducting the experiments, we need to preprocess these datasets in order to filter out some abnormal outlier data such as inactive users and unpopular POIs. For example, as for Foursquare(NYC), we only remove those users as well as their social relations who post less than 5 tips or reviews, and locations which have fewer than 2 check-in records. Meanwhile regarding Yelp, we also eliminate those users and their social relations with less than 5 reviews, and those locations which have no more than 50 check-in records. After that, the filtered datasets comprise the number of users, locations, check-ins, social links and reviews. The statistics of these datasets are shown in Table2.

Table 2. Basic parameters of two LBSN datasets

Datasets	#Users	#Locations	#Check-ins	#Links	#Reviews	#Check-in Density
Foursquare(NYC)	4,811	27,303	59,027	44,397	113,254	4.49E-4
Yelp	16,184	22,005	515,563	314,341	454,086	1.44E-3

As observed from Table 2, the check-in density of user-POI matrix in Foursquare(NYC) and Yelp is 4.49E-4 and 1.44E-3 respectively, which reflects that both datasets are very sparse. Whereas in terms of social relations, the average degree of linked users is about 9.2 for Foursquare(NYC) and 19.4 for Yelp, that indicates the social interactions in these two datasets are relatively close.

2) Baseline methods

To evaluate the performance of our model, we compare our proposed **GE2AT** with the following baselines.

GeoMF[1]: it is based on weighted matrix factorization for POI recommendation, as a variant of matrix factorization based model, which also considers check-ins as an implicit feedback and incorporates geographical influence by modeling users' activity regions and predefined POIs' influence areas.

NeuMF[37]: this model is proposed neural based CF model which integrates matrix factorization(MF) and MLP into a unified model, which uses multiple hidden layers above the element-wise and concatenation of user and location embeddings to capture their nonlinear feature interactions. In this work, we employ three-layered architecture.

PACE[38]: this method is a deep neural architecture based on user preference and context embedding with representation methods. It is a general semi-supervised learning framework that jointly models social influence and user trajectory behavior to predict both user preference over POIs and various context associated with users and POIs. In this work, we set the two configuration parameters of PACE, i.e., R capacity to 3 and Q number of hidden layers to 3.

NGCF[9]: Neural Graph Collaborative Filtering is the most recent item-based recommendation model built upon graph convolutional networks. In this work, we employ NGCF on user-POI check-ins graph by using two graph convolution layers.

CGA[35]: this model is the state-of-the-art attentive collaborative filtering method for POI recommendation in LBSNs, which applies two context-aware attention networks to learn the influence weights of different friends and neighboring POIs respectively.

These baselines could be categorized into three groups: the first belongs to conventional MF based collaborative filtering model without consideration of social network information, the subsequent two models are neural network-based prediction models, and the last two belong to GNN-based models for recommendation.

3) Evaluation Metrics

For performance comparison with baselines, we apply three widely used metrics for evaluating the prediction performance, i.e., Pre@k(precision), Rec@k(recall), and NDCG@k (normalized discounted cumulative gain). These metrics reflect different aspects of the results: precision and recall measure the number of correct prediction results, while NDCG consider the rank of the predictions, and NDCG gives a higher score to the hit items that are ranked higher in the ranking list. To be fair, we repeatedly conduct our experiments 10 times and take the average values of metrics into comparison. For each dataset, we randomly select

80% data as the training set, another 10% as the validation set used for tuning the parameters, and the remaining 10% as testing data.

5.2 Performance Comparison and Analysis

1）Prediction performance of user's preferences over POIs

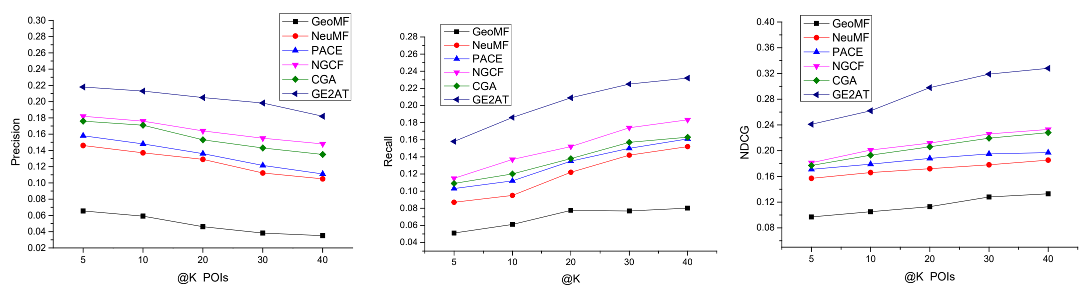
We compare the overall performance of POI predictions for the existing state-of-the-art models on Foursquare(NYC) and Yelp. Specifically, the comparison results of Pre@K, Rec@K and NDCG@K with varying top-k values on the two datasets are reported in Fig.5 and Fig.6, respectively. From the results, we have the following observations:

(1) General MF-based method of GeoMF has the worst performance on two datasets, this indicates that the inner product is insufficient to capture the complex relations between users and locations. Besides, although GeoMF has integrated geographical influence, it did not take other useful information such as social relations into account. Nevertheless, GeoMF is unsuitable for highly sparse user check-in data, which leads to the three types of metrics very low in all cases.

(2) When comparing the neural network based models of NeuMF and PACE, we observe that both of the baseline models consistently outperform GeoMF much better in all metrics, demonstrating the importance of nonlinear feature interactions between user and location embeddings. However, NeuMF achieves slightly lower prediction performance than PACE. The reason might be that PACE shows the extra effectiveness of exploiting user and POI context graphs, while NeuMF fails to explicitly explore the high-order connectivity in embedding learning.

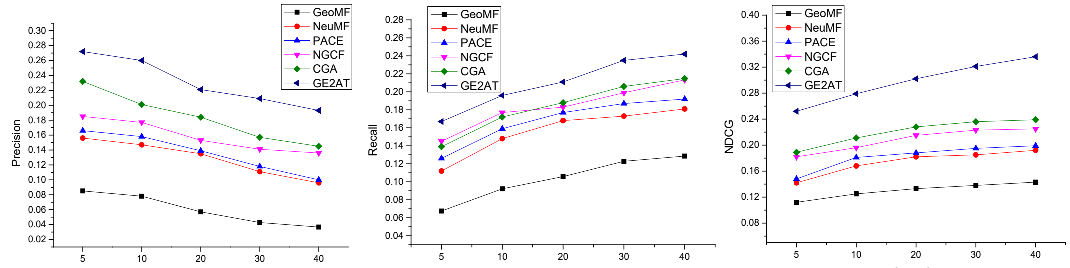
(3) Among the graph neural based models, both NGCF and CGA yield the second best performance compared with above neural network based models on two datasets. By comparison, NGCF slightly outperforms CGA on Foursquare(NYC) dataset, whereas it performs less well than CGA on Yelp dataset in most cases. In particular, compared with PACE, NGCF improves relatively the average performance by 23.1% on precision, 15.3% on recall and 13% on NDCG as shown in Fig.5 on Foursquare(NYC) dataset, respectively. However, CGA exhibits the average performance improvement by 36% on precision, 9.5% on recall and 21.4% on NDCG as shown in Fig.6 on Yelp dataset, respectively. The results demonstrate the benefits of exploiting graph neural networks for POI prediction. We conjecture that the superior performance of these two models are due to two reasons. One is that leveraging neural attentive mechanism just like CGA results in better prediction quality when the user-POI interaction data is less sparse such as Yelp dataset. The other reason is that exploiting the high-order connectivity or neighbors such as NGCF greatly facilitates the representation learning to infer user preferences, as the deeper interaction information can be effectively captured.

(4) Our proposed GE2AT method significantly outperforms all other methods across both datasets. Concretely, by comparing to the strongest method, we observe that GE2AT constantly yields the average increases of precision by 23.3%, recall by 33.3% and NDCG by 37.2% on Foursquare(NYC), and similarly the average improvements on Yelp are about 26% on precision, 10% on recall and 34.9% on NDCG. These comparison results for POI prediction indicate that our GE2AT model benefits from combining the merits of graph neural representations learning from high-order connectivity and dual attentive mechanism to better improve prediction performance. In other words, this happens because GE2AT has fully integrated the advantages of both NGCF and CGA methods. Furthermore, owing to incorporating social influence information, GE2AT could not only effectively alleviate the data sparsity problem, but also consistently achieve the best performance than those utilizing only user check-in relational data in our experiments.



(a)Precision@K (b)Recall@K (c)NDCG@K

Fig.5. POIs prediction performance on Foursquare(NYC) dataset



(a)Precision@K (b)Recall@K (c)NDCG@K

Fig.6. POIs prediction performance on Yelp dataset

2）Prediction performance of user's preferences over social topics

Regarding social topics prediction, we first need to extract the semantic topics from all the reviews or tips associated with users and locations. Similar to many works, we use the popular Gensim tool (https://radimrehurek.com/gensim/) to learn the embedding of each topic with traditional Word2vec model. The number of social topics is tuned to 120 on both datasets.

The performance comparisons of Pre@K, Rec@K and NDCG@K with K setting to [5,10,15,20,25] on the two datasets are shown in Fig.7 and Fig.8, respectively. Note that GeoMF is unsuitable for social topics prediction thus excluded from the competitors. The cause is that GeoMF is only applied for POI recommendation by leveraging the weighted user-POI check-ins matrix as well as geographical influence. From the comparison results, we have the following findings:

(1) Both neural network based and graph neural network based methods have demonstrated good performance for social topics prediction on two datasets, especially they almost have close prediction performance regarding precision. For example, NeuMF and PACE have close precision to NGCF and CGA on Foursquare(NYC) and Yelp. While in terms of the results of recall and NDCG, we observe that NeuMF and PACE obviously underperform the latter two methods, the relative average reduction on Foursquare(NYC) and Yelp is by 11% and 25% on recall metric, and 18% and 17% on NDCG metric, respectively. These results imply that both neural network embedding and graph neural embedding are effective and can accurately capture the complex non-linear embeddings of users and social topics when existing richer interaction data between them, unlike the very sparse interaction between users and POIs. Whereas we further highlight that the latter graph neural based methods still have the stronger overall prediction power than the former two methods such as NeuMF and PACE.

(2) Among all the compared baselines, our GE2AT performs the most superior performance for social topics prediction in all cases, and constantly yields at least 13% relative improvements on Foursquare(NYC) and more than 30% improvements on Yelp when compared with the strongest baseline. Such improvements could be attributed to the following two reasons.

—Our dual attentive graph networks effectively models better user preferences, thus improving prediction quality.

—Incorporating social influence information is helpful to alleviate the data sparsity issue in a large extent. The comparison results have verified that social influence is very important to learn user preferences and boost the prediction performance.

(3) To sum up, comparing the results on Foursquare(NYC) and Yelp datasets, our GE2AT model achieves the best performance for both types of prediction tasks in all cases. On this basis, some meaningful conclusions could be summarized as follows. Firstly, extending graph neural embeddings with dual attention mechanism by differentiating importance weights for neighboring entities is the most effective way to capture the best user preferences in heterogeneous LBSN. Secondly, we exploit neural networks through multiple MLPs to make the deep pairwise prediction of user preferences over POIs and social topics with extra consideration of the social influence to alleviate data sparsity problem. Thirdly, we fully take into account incorporating multi-dimensional relation data of LBSN to boost the prediction performance. For instance, GE2AT model has effectively fused three types of heterogeneous relationships such as user social relations, user-to-location check-ins relations and user-to-social topics following relations.

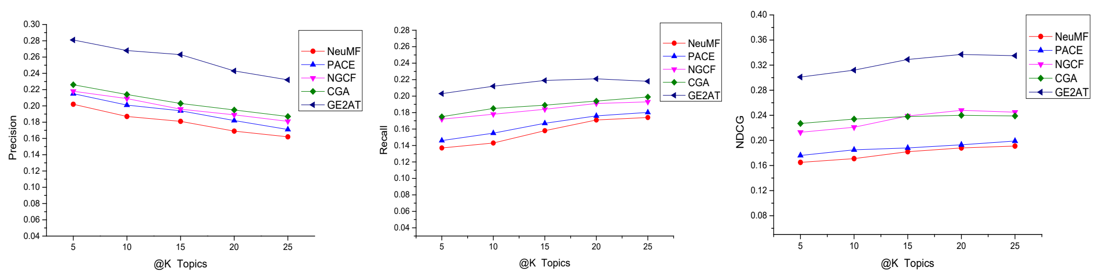


Fig.7. Social topics prediction results on Foursquare(NYC) dataset

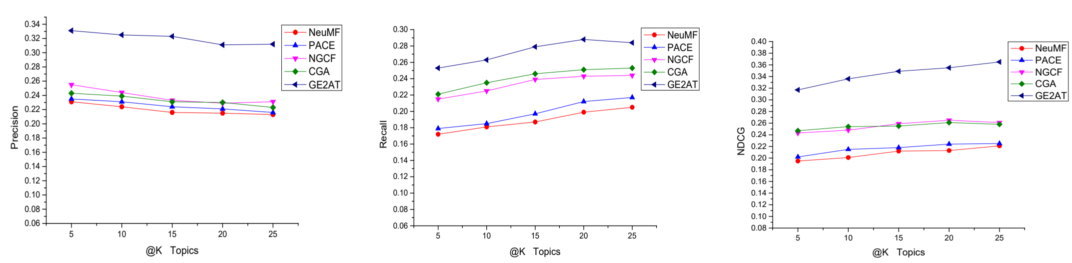


Fig.8. Social topics prediction results on Yelp dataset

3）Parameter analysis

We further conduct experiments to analyze the effects of different model parameters on the performance of GE2AT. Here, we mainly focus on two important components in our framework, one is the parameters of dual graph attention networks in the embedding layer, and the other is the parameters of MLPs in the hidden layer.

To investigate the effects of dual graph attention networks, we evaluate different impacts caused by the two key parameters of this module, including the number of sampled neighbors N and the influence of the propagation layer numbers l. By varying the parameters, we test the performance of GE2AT for predicting top-10 POIs and social topics. The results on NDCG@10 are shown in Fig.9 and Fig.10, from which we have the following findings:

(1) When the number of sampled neighbors varies from 10 to 60 on two datasets for two prediction tasks, the metric of NDCG@10 increases steadily at first as a proper number of neighbors is considered. It demonstrates that integrating the local neighboring information can enhance the node representation. However, when the size of neighbors exceeds a certain value, the prediction performance decreases slightly. Obviously, we can find that the best neighbor size ranges from 30 to 40 for both prediction tasks. This occurs because when the neighbors are not enough, the embedding layer is unable to capture better representations of user preferences hidden in dual graph attention networks. While increasing redundant neighbors may be easily involved more noisy or uncorrelated neighbors to the representation learning, the performance improvement will be hindered because of them.

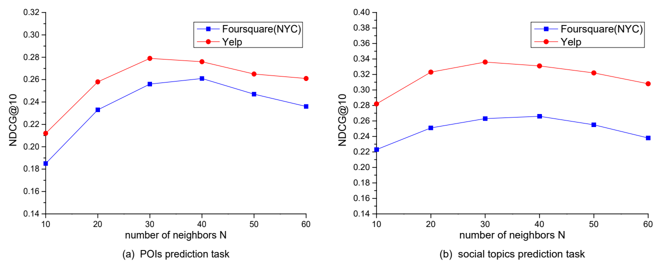


Fig.9. Impact of number of sampled neighbors

(2) When varying the number of the propagation layers (i.e., parameter l), the NDCG@10 performance achieves the best results under 4 propagation layers for POIs prediction task, and 3 propagation layers for social topics prediction task in dual graph attention networks, respectively, which shows that the increasing of propagation layers does not correspondingly lead to better performance. The marginal improvement on the two datasets verifies that conducting 3 propagation layers is sufficient to capture the deep non-linear representations of different entities. This seems to indicate that the discrimination of the nodes is decreasing as the number of layers increases. Thus, we argue that increasing propagation layers makes the high-order neighbors more similar, and further learns more similar node representations.

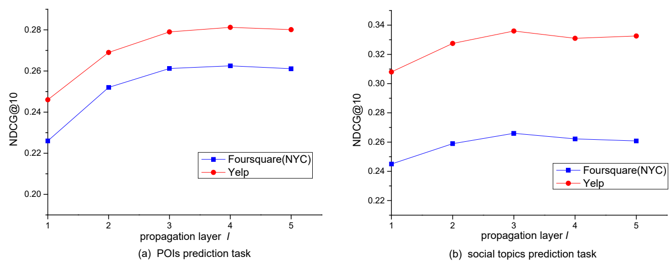


Fig.10. Impact of propagation layers

Next, to evaluate the impacts of the hidden layers in the MLPs, we vary the number of hidden layers (i.e., parameter l') from 1 to 5, but the input embedding size is fixed to 256. As shown in Fig.11, we note that the best architecture of MLP is implemented with 3 hidden layers, while further stacking more hidden layers will lead to slight overfitting issue across two datasets. The results of two prediction tasks have relatively stable performance, and tend to exhibit the lower gaps below 11% variation of NDCG for POIs prediction and 9% variation for social topics prediction among all different layer structures. This verifies the effectiveness of using MLPs to model user preferences over POIs and social topics. In summary, we find that the prediction performance of GE2AT clearly depends on the number of neighbors and the embedding layers of dual graph attention networks, but not much beneficial from the number of hidden layers of MLPs.

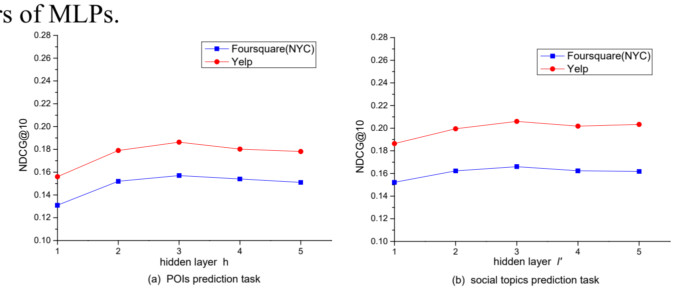


Fig.11. Impact of the number of hidden layers in MLPs

6 Conclusion

In this work, we explore how to jointly learn the deep user preferences by fusing multi-types of relations in heterogeneous LBSN. Towards this end, we have proposed a novel unified pairwise relationship prediction framework known as GE2AT, which utilizes graph neural networks combined with dual graph attentions. Our model has considered to fully incorporate three types of user behavioral relations in LBSN, i.e., social relations between users, user-location check-in relations and user-social topics following relations. More specifically, GE2AT focuses on graph embeddings of different entities by leveraging dual attention mechanism, and it can adaptively capture better non-linear representations of user preferences via aggregating information from neighbors. Furthermore, GE2AT also incorporates the social influence to further improve the pairwise prediction performance, by leveraging social relations to shape similar user preferences. Extensive experiments on Foursquare(NYC) and Yelp datasets have demonstrated the effectiveness of our proposed GE2AT, which achieves very significant improvements of prediction performance in terms of three commonly-used metrics compared to several state-of-the-art baselines. This, in turn, indicates a promising direction that modeling and optimizing prediction tasks by graph representation learning, especially for massive sparse and multi-dimensional data in heterogeneous information networks.

In future, we will extend our work by considering more information such as temporal and sequential behaviors of users to capture the dynamic evolution of user preferences in LBSN, so as to gain better prediction performance and scalability.

ACKNOWLEDGMENTS

This work was partly supported by the Natural Science Foundation of Zhejiang Province under the Grant No. LY13F020026, LQ18F020008, the National Natural Science Foundation of China under the Grant No.61502420, 61802346, and the China Postdoctoral Science Foundation under the Grant No.2015M581957, and Zhejiang Provincial “Ling Yan” Research and Development Project under the Grant No.2022C01088.

REFERENCES

[1] D. Lian, C. Zhao, X. Xie, G. Sun, E. Chen, Y. Rui. (2014). GeoMF: joint geographical modeling and matrix factorization for point-of-interest recommendation, *in: Proceedings of the 20th ACM International Conference on Knowledge Discovery and Data Mining, New York,USA,pp.*831-840.

[2] Shaosheng Cao, Wei Lu, and Qiongkai Xu.(2015). GraRep: Learning graph representations with global structural information”. *in Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. ACM,pp.*891–900.

[3] Mehri Davtalab,Ali Asghar Alesheikh.(2021).A POI recommendation approach integrating social spatio-temporal information into probabilistic matrix factorization. *Knowledge and Information Systems.*vol.63,pp.65–85.

[4] 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,*pp.701–710.

[5] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, Qiaozhu Mei. (2015). LINE: Large-scale information network embedding. *in Proceedings of the 24th International Conference on World Wide Web,*pp.1067–1077.

[6] Hakan Bagci,Pinar Karagoz.(2016).Context-aware location recommendation by using a random walk-based approach. *Knowledge and Information Systems.*vol.47, pp.241–260.

[7] F. Scarselli, M. Gori, A. C. Tsoi, et al. (2009). The graph neural network model. *IEEE Transactions on Neural Networks,* 20(1),pp.61-80.

[8] Will Hamilton, Zhitao Ying, and Jure Leskovec. (2017). Inductive representation learning on large graphs. *In Proceedings of the 31st International Conference on in Neural Information Processing Systems(NIPS'17),Long Beach,USA,*pp.1024-1034.

[9] XiangWang, Xiangnan He, MengWang, Fuli Feng,Tat-Seng Chua.(2019). Neural graph collaborative filtering. *in Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval(SIGIR'19),Paris,France*,pp.165-174.

[10] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, et al. (2018). Graph attention networks. *In Proceedings of ICLR 2018.Vancouver, Canada*, https://arxiv.org/abs/1710.10903.

[11] Qitian Wu, Hengrui Zhang, Xiaofeng Gao, Peng He,Paul Weng,Han Gao,Guihai Chen. (2019). Dual graph attention networks for deep latent representation of multifaceted social effects in recommender systems. *in Proceedings of WWW '19,*pp.2091-2102.

[12] Surong Yan,Haosen Wang,Yixiao Li,Yuan Zheng,Long Han.(2021).Attention-aware metapath-based network embedding for HIN based recommendation *Expert Systems with Applications,*174,114601.

[13] Jean-Benoit Griesner, Talel Abdessalem, and Hubert Naacke. (2015). POI recommendation: towards fused matrix factorization with geographical and temporal influences. *In Proceedings of the 9th ACM Conference on Recommender Systems (RecSys’15). Vienna,Austria*,pp.301-304.

[14] X. Li, G. Cong, X. Li, et al.(2015). Rank-GeoFM: a ranking based geographical factorization method for point of interest recommendation. *in Proceedings of the ACM SIGIR’15, Santiago, Chile,*pp.433-442.

[15] Zhiyuan Zhang, Yun Liu, Zhenjiang Zhang, Bo Shen. (2019). Fused matrix factorization with multi-tag, social and geographical influences for POI recommendation. *World Wide Web,vol.22,pp.*1135-1150.

[16] Dandan Lyu, Ling Chen, Zhenxing Xu, Shanshan Yu.(2020). Weighted multi-information constrained matrix factorization for personalized travel location recommendation based on geo-tagged photos. *Applied Intelligence,vol.50,pp.*924-938.

[17] Xueming Qian,He Feng, Guoshuai Zhao, Tao Mei. (2014). Personalized recommendation combining user interest and social circle. *IEEE Trans. Knowledge and Data Engineering,26(7),pp.*1763-1777.

[18] L. Yao, Q. Z. Sheng, Y. Qin, X. Wang, et al.(2015). Context-aware point-of-interest recommendation using tensor factorization with social regularization.*in Proceedings of the 38th International ACM SIGIR Conference(SIGIR’15), Santiago, Chile,*pp.1007-1010.

[19] D. Yang, D. Zhang, V. W. Zheng, et al. (2015). Modeling user activity preference by leveraging user spatial temporal characteristics in LBSNs. *IEEE Transactions on Systems, Man, and Cybernetics: Systems.45(1),*pp.129-142.

[20] V. W. Zheng, Y. Zheng, X. Xie, et al. (2012). Towards mobile intelligence: learning from GPS history data for collaborative recommendation. *Artificial Intelligence.vol.184,*pp.17-37.

[21] Shenglin Zhao, Tong Zhao, Haiqin Yang, et al. (2016). STELLAR: Spatial-temporal latent ranking for successive point-of-interest recommendation. *In Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI’16).Phoenix,USA*,pp.316-321.

[22] He, J.; Li, X.; Liao, L.; Song, D.; et al. (2016). Inferring a personalized next point-of-interest recommendation model with latent behavior patterns. *In Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI'16),Phoenix,USA,*pp.137-143.

[23] Liu Q, Wu S, Wang L, et al.(2016). Predicting the next location: a recurrent model with spatial and temporal contexts. *In: Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI'16),Phoenix,USA,*pp.194-200.

[24] Dongliang Liao, Weiqing Liu, Yuan Zhong, et al.(2018). Predicting activity and location with multi-task context aware recurrent neural network. *In Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI-18),Stockholm,Sweden,*pp.3435-3441.

[25] Pengpeng Zhao,Haifeng Zhu,Yanchi Liu, et al. (2019). Where to go next: a spatio-temporal gated network for next POI recommendation. *in Proceedings of the 33rd AAAI Conference on Artificial Intelligence(AAAI-19),Hawaii,USA,*pp.5877-5884.

[26] L. Huang, Y. Ma, S. Wang, et al.(2021). An attention-based spatiotemporal LSTM network for next POI recommendation. *IEEE Transactions on Services Computing.14(6),pp.*1585-1597.

[27] Cheng Yang, Maosong Sun, Wayne Xin Zhao, et al. (2017). A neural network approach to joint modeling social networks and mobile trajectories. *ACM Transactions on Information Systems.35(4)pp.*3601-3628.

[28] Khoa D.Doan, Guolei Yang, Chandan K.Reddy. (2019). An attentive spatio-temporal neural model for successive point of interest recommendation. *In Proceedings of the 23rd PAKDD 2019,Macau,China,*pp.346-358.

[29] Dongjing Wang, Dengwei Xu, Dongjin Yu, et al . (2021). Time-aware sequence model for next-item recommendation. *Applied Intelligence, vol.51,pp.*906–920.

[30] Wenqi Fan, Yao Ma, Qing Li, et al. (2020). A graph neural network framework for social recommendation. *IEEE Transactions on Knowledge and Data Engineering(Early Access),DOI:10.1109/ TKDE.2020.3008732*

[31] Yinwei Wei, Xiang Wang, Liqiang Nie, et al. (2019). MMGCN: multi-modal graph convolution network for personalized recommendation of micro-video. *In Proceedings of the 27th ACM International Conference on Multimedia(MM'19),Nice,France,*pp.1437-1445.

[32] Rex Ying, Ruining He, Kaifeng Chen, et al. (2018). Graph convolutional neural networks for web-scale recommender systems. *In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining(KDD '18),London, U K,*pp.974-983.

[33] Yahui Liu, Furao Shen, Jian Zhao. (2019). Pairwise interactive graph attention network for context-aware recommendation.*http://arXiv:1911.07429.*

[34] Yongping Du,Zhi Peng,Jinyu Niu,Jingya Yan.(2022).A unified hierarchical attention framework for sequential recommendation by fusing long and short-term preferences.*Expert Systems with Applications*,201,117102.

[35] Siyuan Zhang,Hong Cheng. (2018). Exploiting context graph attention for POI recommendation in location-based social networks. *In Proceedings of DASFAA 2018,Hsinchu,Taiwan,*pp.83-99.

[36] Kingma, D. and Ba, J. (2015). Adam: a method for stochastic optimization. *in Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015).*San Diego,CA.

[37] He X, Liao L, Zhang H, et al. (2017). Neural collaborative filtering. *in Proceedings of the 26th International Conference on World Wide Web, Perth, Australia*,pp.173-182.

[38] C. Yang, L. Bai, C. Zhang, Q. Yuan, et al. (2017). Bridging collaborative filtering and semi-supervised learning: a neural approach for poi recommendation. *In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(KDD'17),Halifax, Canada*,pp.1245-1254.

Weihua Gong received the Ph.D. degree in Software Engineering from Huazhong University of Science and Technology, Wuhan, China, in 2006. He is currently an associate professor, M.S. supervisor in the School of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, China. His research interests include social network, machine learning.

Kechen Zheng received the B.Eng. degree in Electronic Engineering from Shanghai Jiao Tong University, China, in 2013 and Ph.D. degree in Electronic Engineering in the same university in 2018. He is currently an associate professor in the School of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, China. His research interests include performance evaluation in cognitive networks and social networks, and energy harvesting wireless communication networks.

Shubin Zhang, received the B.S. degree in computer science from Xidian University, Xi'an, China, in 2014, where he received the Ph.D. degree with the School of Cyber Engineering, in 2020. He is currently a Lecturer in the School of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, China. His research interests include device-to-device communications, edge computing and IoT networks.

Ping Hu received the B.Eng. degree in telecommunications engineering from Xidian University, Xi'an, China, in 2006, and the M.Phil. and Ph.D. degrees from the City University of Hong Kong, Hong Kong, China, in 2009 and 2016, respectively. She is currently with the School of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, China. Her research interests include wireless communication, information-theoretic security, and distributed storage systems.