Disentangling Geographical Effect for Point-of-Interest Recommendation

Yingrong Qin, Chen Gao, Yue Wang, Shuangqing Wei, *Senior Member, IEEE*, Depeng Jin, *Member, IEEE*, Jian Yuan, and Lin Zhang

Abstract—Point-of-Interest (POI) recommendation has drawn a lot of attention in both academia and industry. It utilizes user check-in data, aiming at recommending unvisited POIs to users. To address the data-sparsity problem, geographical information of POIs is often incorporated into recommender systems. However, most of the existing approaches model geographical impact in an implicit way, in which geographical information is encoded as auxiliary vectors for learning unified representations of users and POIs. Following this paradigm, the embedding of POIs can not reflect geographical similarity directly; thus, an explicit modeling approach is needed as geography is of great importance in POI recommendation. To address challenges in disentangling geographical effect, we proposed a disentangled representation learning method named *DIG* (short for **D**isentangled embedding of user **I**nterest and POIs' **G**eographical information). Aiming at decoupling the geographical factor and the user interest factor thoroughly, we first proposed a geo-constrained negative sampling strategy, which helps to find reliable negative samples for the two factors. Second, a geo-enhanced soft-weighted loss function was proposed to quantify the trade-off between the two factors in loss computation. Extensive experiments have been conducted on two real-world datasets, and results have demonstrated the significant improvement of DIG at $3.92\% \sim 20.32\%$ on recall, and $2.53\% \sim 11.48\%$ on hit ratio, compared with other state-of-the-art approaches.

Index Terms—Disentangled embedding, geographical effect, graph neural networks, point-of-Interest recommendation

1 Introduction

In recent years, location-based social networks (LBSN) have appeared and entered into our daily life, benefiting from the prevalence of mobile networks and the convenience of mobile terminals. Various Point-of-Interest (POI) recommendation approaches have been proposed to enhance user experience and produce more commercial profits, such as [1], [2], [3], [4], [5], [6], [7], [8].

However, the geographical information of POIs distinguishes POI recommendation from traditional recommendation scenarios. Specifically, geographical attributes of POIs add constraints to interactions between users and POIs [9], [10], [11], [12], [13], [14]. Unlike interactions (click, add to cart, buy, etc.) between users and commodities on E-Commerce platforms, the check-in behavior of users requires extra cost, and the consumed time varies due to the geographical distance between the user's current location and the location of a target POI. Besides, neighborhood influence works as an uncontrollable variable in POI recommendation. Taking the restaurant recommendation as an example, a driving user is more likely to prefer restaurants closer to a spot with parking space. Therefore, POIs such as parking lots and restaurants in the same region can influence each other. However, even the same POI influences its neighbors differently, and the neighborhood of different POIs varies greatly. Moreover, the impact on the neighbors also decreases as distance increases [3], [15]. Thus, the influence from the neighborhood of POIs is non-negligible in POI recommendation, and should be captured well in the representation learning process for performance improvement.

In this paper, we aim to provide an approach that can model geographical attributes of POIs explicitly, while previous studies [1], [2], [3], [4], [15], [16], [17], [18] only model the geographical factor in an implicit way. We classify existing methods into three categories to explain the major drawback. The first category utilizes geographical information in a regularization-based way, in which geographical information is not encoded by latent vectors but is used to form extra regularization terms [1]. The second category learns latent vectors for geographical information, but either these latent vectors work as hidden vectors for a unified representation [3], [15], or is optimized by the fusion of user interest and geography, etc., [4], [16], [17]. The last category, which uses separate embedding space to encode geographical

- *Yingrong Qin, Yue Wang, Depeng Jin, and Jian Yuan are with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China. E-mail: qyr16@mails.tsinghua.edu.cn, wangyue@mail.tsinghua.edu.cn, {jindp, jyuan}@tsinghua.edu.cn.*
- *Chen Gao is with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China, and also with the Huawei Noah's Ark Lab, Beijing 100000, China. E-mail: chgao96@gmail.com.*
- *Shuangqing Wei is with the Department of Electrical and Computer Engineering, Louisiana State University, Baton Rouge, LA 70803 USA. E-mail: swei@lsu.edu.*
- *Lin Zhang is with the Tsinghua Shenzhen International Graduate School, Shenzhen 518055, China. E-mail: linzhang@tsinghua.edu.cn.*

Manuscript received 24 October 2021; revised 15 September 2022; accepted 30 October 2022. Date of publication 14 November 2022; date of current version 21 June 2023.
This work is supported in part by the National Natural Science Foundation of China under Grants 62272262, U20B2060, and U21B2036, in part by the National Key Research and Development Program of China under Grant 2022YFB3104702, in part by the Guoqiang Institute, Tsinghua University under Grant 2021GQG1005, and in part by the fellowship of China Postdoctoral Science Foundation under Grants 2021TQ0027 and 2022M710006.
(Corresponding author: Chen Gao.)
Recommended for acceptance by E. Chen.
Digital Object Identifier no. 10.1109/TKDE.2022.3221873

information, [2], [18], still belongs to the implicit paradigm for two reasons. First, the representations in geographical space are pre-computed by using coordinates of POIs and check-in frequencies of users and are fixed during optimization, which can lead to sub-optimal performance. Second, the use of check-in frequency, which is closely associated with user interest, mingles user interest with geography again. Therefore, the drawback of modeling geographical information in POI recommendation is obvious, and approaches of explicit modeling are required.

To avoid rough fusion of the geographical factor with other non-geographical factors, we adopt a disentangled representation learning approach in which separate subspaces are explicitly learned for different factors (e.g., interest and geographical factors). There are two significant challenges in disentangling geographical effects, and they are summarized as follows.

- *Distinguishing geographical factor from non-geographical factors.* Geographical factor is often deeply fused with non-geographical factors like user interest in POI recommendation. For example, a driving user may visit a shopping mall with an underground parking lot multiple times, while never visiting another mall without parking places. In this situation, we can not judge whether the latter one is less attractive or not, as the parking lot is an important and uncontrollable variable. Therefore, it is challenging to achieve disentanglement of the geographical factor and other factors.
- *Trade-off between geographical factor and non-geographical factors.* As mentioned, different embedding spaces are used for different factors in the disentangled embedding learning process. Therefore, how to balance representations of different spaces in loss computation is a challenging problem, as contributions of different factors vary in different interactions.

To address the challenges mentioned above, we propose a disentangled approach named DIG, namely, **D**isengled user **I**nterest factor and POI **G**eographical factor for POI recommendation. To be specific, representations of users and POIs are obtained by concatenating representations from two separate spaces, and a geo-constrained negative sampling strategy is proposed to generate representative negative samples that can balance different spaces. Moreover, a geo-enhanced soft-weighted loss function is designed to estimate the weights of loss.

To summarize, our major contributions are listed below:

- We propose a novel disentangled representation learning approach that clearly models the user interest factor and POI geography factor, the two deeply fused factors in POI recommendation.
- We also propose a geo-constrained negative sampling strategy and a geo-enhanced soft-weighted loss function, which distill POI geographical information to assist disentanglement.
- Extensive experiments are conducted on two real-world LBSN datasets, and our proposed model outperforms the state-of-the-art algorithms. In addition, experimental results demonstrate the effectiveness of disentanglement in improving representation learning.

TABLE 1
Notations Frequently Used in This Paper

Notations	Descriptions
$\lvert\cdot\rvert$	The cardinality of a set
$\langle\cdot\rangle$	Inner product
$\mathcal{U}, \mathcal{I}$	The set of users, POIs
I_u^+, I_u^-	POIs visited by user u, POIs not visited by user u
U_i^+, U_i^-	Users who visited POI i, users who didn't visit POI i
$\mathcal{C}_u$	Candidate set for user u
$\mathcal{O}$	Triplet set of $(u,i,j), u \in \mathcal{U}, i \in I_u^+, j \in I_u^-$
$\mathbf{U}, \mathbf{I}$	Embedding of all users, all POIs
$\vec{u}, \vec{i}$	Embedding of user u, POI i
d_i, d_g	Dimension of interest space, geography space
n_i, n_g	The number of negative samples
$\vec{u}^{\text{int}}, \vec{i}^{\text{int}}$	Sub-vectors of user u, POI i in interest space
d_m	Distance margin used in negative sampling
m_f	Distance expanding margin used to enlarge d_m
$\vec{u}^{\text{geo}}, \vec{i}^{\text{geo}}$	Sub-vectors of user u, POI i in geography space
M, N	The number of users, POIs
x_i, y_i	Latitude and longitude of POI i
$d(i,j)$	Geographical distance between POI i and POI j
$\mathcal{L}$	Loss function

Moreover, since real-world recommender systems can always be split into matching, ranking, and re-ranking stages [19], [20], [21], our proposed method, DIG, focuses on the matching stage specifically. During this stage, recommender systems will select dozens of candidates from the item set accurately. In addition, the interest factor mentioned in this paper is a broad concept, which is similar to [11]. Specifically, user preferences are grouped into two parts: the geographical part and the non-geographical part. We name the former part the geographical factor, which is tightly associated with users' tolerance of travel time. Similarly, we name the latter part interest factor, including properties of users, the popularity of POIs, etc. Extensive experimental results have demonstrated that our DIG decouples these two factors effectively. The rest of this paper is organized as follows. We first formulate the studied problem in this paper in Section 2. We then describe our proposed DIG method in Section 3, following which we present the experimental results in Section 4. We then review the related work in Section 5 and discuss three important questions about the concepts and the scope of this paper in Section 6. Finally, we conclude this paper in Section 7.

2 Problem Formulation

In this section, we formulate the POI recommendation problem to solve in this paper. Notations frequently used in this paper are listed in Table 1.

Given M users and N POIs, let $\mathcal{U}$ denote the user set and $\mathcal{I}$ the POI set. For a specific user u, I_u^+ indicates POIs visited by user u, and I_u^- indicates those unvisited ones. (x_i, y_i) denotes the latitude and longitude of POI $i, i \in \mathcal{I}$. The goal of recommender system is to generate candidate sets $\mathcal{C}_u =$

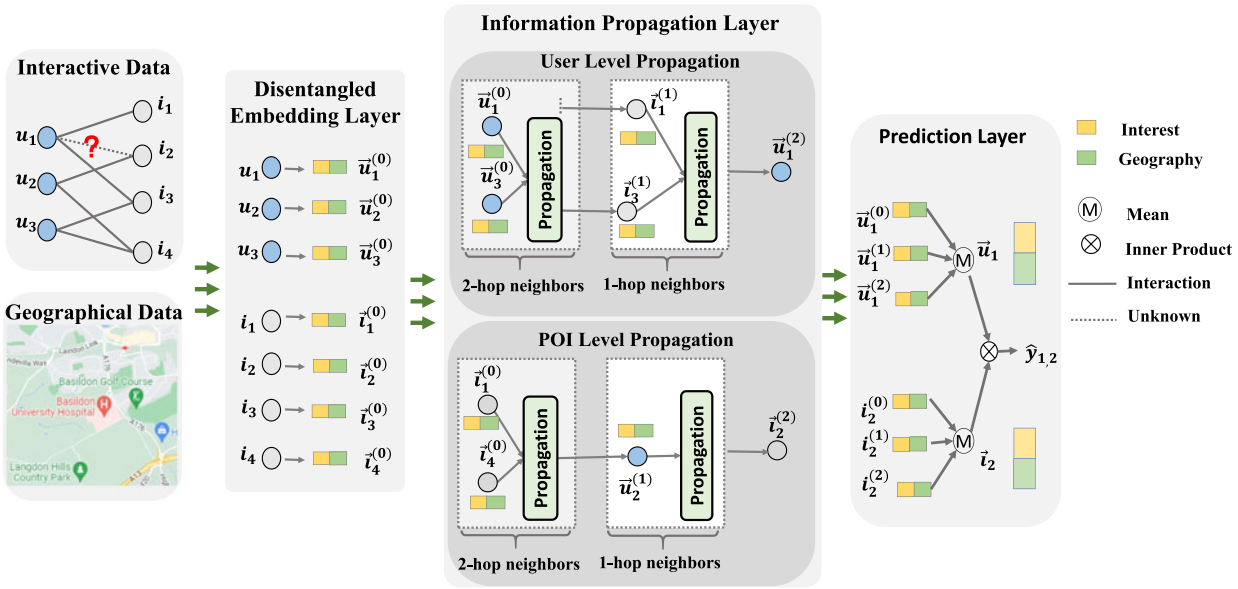


Fig. 1. An illustration of our DIG model (best view in color). The model input is made up of two parts: user-POI interaction data and geographical data. Taking these data as input, DIG's disentangled embedding layer can generate disentangled components for the interest factor (yellow color) and geographical factor (green color), respectively. These disentangled embeddings will be optimized in the Information Propagation Layer, leveraging both user-POI bipartite graph structure and geographical constraints. After convergence, the candidate set of each user is generated by ranking the matching scores between user embeddings and all un-visited POIs' embeddings.

$\{j|j \in I_u^-\}, \forall u \in \mathcal{U}$ for each user by utilizing user embedding $\mathbf{U} = [\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_M]$ and POI embedding $\mathbf{I} = [\vec{i}_1, \vec{i}_2, \ldots, \vec{i}_N]$, which are optimized by using interaction data between users and POIs. To be specific, $\mathcal{C}_u, \forall u \in \mathcal{U}$ is generated by calculating inner product between user $\vec{u}$ and POIs $\vec{i}$, where $i \in I_u^-$. $\mathcal{C}_u$ consists of POIs with top-K largest inner product values, where K is the predefined size of the candidate set.

3 Methodology

In this section, we present the proposed model DIG in detail in five parts:

- the disentangled embedding layer, which describes disentanglement within each layer;
- the information propagation layer, which depicts how information propagates on a graph;
- the prediction layer that calculates scores between user-POI pairs using disentangled embedding;
- the geo-constrained negative sampling strategy, which helps disentanglement by curriculum learning under geographical constraints;
- the geo-enhanced soft-weighted loss function, which distills geographical information to balance user interest factor and geographical factor in final prediction.

The first four components are illustrated in Fig. 1, in which the yellow color indicates representations of interest factor, and the green color indicates representations of the geographical factor. When inputting geographical data and user-POI interactions into DIG, disentangled embedding of users and POIs will first be initialized in the disentangled layer. Then, embedding will be optimized by a GNN model, and information is propagated on a user-POI bipartite graph in the information propagation layer. Ultimately, recommended candidates will be generated in the prediction layer by ranking the inner product of a specific user and all un-visited POIs.

3.1 Disentangled Embedding Layer

As we want to learn disentangled representation for user interest factor and POI geographical factor, two separate spaces are used to model these two factors, respectively. Following disentangled models in [22], [23], [24], embedding of user i is obtained by concatenation as follows:

$$\vec{u}_i = [\vec{u}_i^{int} || \vec{u}_i^{geo}], \quad i = 1, \ldots, M, \tag{1}$$

where $\vec{u}_i^{int} \in \mathbb{R}^{d_i}$, indicates the sub-vector of interest space for user i, and $\vec{u}_i^{geo} \in \mathbb{R}^{d_g}$ indicates its sub-vector of geographical space. $||$ denotes vector concatenation operation.

Similarly, the embedding of POI k is defined as:

$$\vec{i}_k = [\vec{i}_k^{int} || \vec{i}_k^{geo}], \quad k = 1, \ldots, N, \tag{2}$$

where $\vec{i}_k^{int} \in \mathbb{R}^{d_i}$, and $\vec{i}_k^{geo} \in \mathbb{R}^{d_g}$. Therefore, the embedding of all users can be represented as follows,

$$\mathbf{U} = [\mathbf{U}^{int} || \mathbf{U}^{dis}], \tag{3}$$

and the embedding of all POIs as follows:

$$\mathbf{I} = [\mathbf{I}^{int} || \mathbf{I}^{dis}], \tag{4}$$

where $\mathbf{U} \in \mathbb{R}^{(d_i+d_g)\times M}$, $\mathbf{I} \in \mathbb{R}^{(d_i+d_g)\times N}$.

Note that $\mathbf{U}$ and $\mathbf{I}$ are optimized in an end-to-end fashion in DIG. Traditional recommender approaches like MF [25] and NCF [26] feed interactions of user-item pairs in the representation learning process, in which only first-order connectivity is utilized. In contrast, the embedding of users and POIs are optimized by propagating them on the user-POI interaction graph. In this way, more representative embedding can be obtained as high-order connectivity is also used during representation optimization [27].

3.2 Information Propagation Layer

As users who visit a POI can be regarded as the feature of that POI, and POIs that are visited by a user can be treated as the user's feature [27], embedding propagation is conducted by aggregating feature vectors from the directly connected users or POIs over the user-POI interaction graph. Following [27], information aggregation of interest representation in layer $l+1$ can be defined as:

$$\vec{u}^{\text{int}(l+1)} = \sum_{i \in I_u^+} \frac{1}{\sqrt{|I_u^+|}\sqrt{|U_i^+|}} \vec{i}^{\text{int}(l)}, \tag{5}$$

$$\vec{i}^{\text{int}(l+1)} = \sum_{u \in U_i^+} \frac{1}{\sqrt{|U_i^+|}\sqrt{|I_u^+|}} \vec{u}^{\text{int}(l)}, \tag{6}$$

where U_i^+ indicates users that have visited POI i, and symmetric normalization term $\frac{1}{\sqrt{|I_u^+|}\sqrt{|U_i^+|}}$ follows the design of [28]. Moreover, self-connection is excluded as [29] does.

Similarly, information aggregation of geographical embedding in layer $l+1$ is computed as:

$$\vec{u}^{\text{geo}(l+1)} = \sum_{i \in I_u^+} \frac{1}{\sqrt{|I_u^+|}\sqrt{|U_i^+|}} \vec{i}^{\text{geo}(l)}, \tag{7}$$

$$\vec{i}^{\text{geo}(l+1)} = \sum_{u \in U_i^+} \frac{1}{\sqrt{|U_i^+|}\sqrt{|I_u^+|}} \vec{u}^{\text{geo}(l)}, \tag{8}$$

In addition, user embedding, POI embedding of layer l is obtained as:

$$\mathbf{U}^{(l)} = [\mathbf{U}^{int\ (l)} || \mathbf{U}^{geo\ (l)}], \quad \mathbf{I}^{(l)} = [\mathbf{I}^{int\ (l)} || \mathbf{I}^{geo\ (l)}]. \tag{9}$$

As illustrated in Fig. 1, first-order proximity is preserved by aggregating information from the directly connected neighbors, and high-order proximity can be captured via traversing through multiple layers of the model.

3.3 Prediction Layer

After propagating with L layers, we obtain user embedding $\{\vec{u}^{(1)}, \ldots, \vec{u}^{(L)}\}$ and POI embedding $\{\vec{i}^{(1)}, \ldots, \vec{i}^{(L)}\}$. Following [29], linear operation instead of concatenation is leveraged. The final user embedding and POI embedding are computed as:

$$\vec{u} = \frac{1}{L+1}\left(\vec{u}^{(0)} + \vec{u}^{(1)} + \cdots + \vec{u}^{(L)}\right), \tag{10}$$

$$\vec{i} = \frac{1}{L+1}\left(\vec{i}^{(0)} + \vec{i}^{(1)} + \cdots + \vec{i}^{(L)}\right), \tag{11}$$

where $\vec{u}^{(0)}$ and $\vec{i}^{(0)}$ are generated by model initialization.

Finally, inner product, a widely used user-item interaction function [23], [27], [29], is conducted to compute the score between user-POI pair (u, i) as follows:

$$\hat{y}_{ui} = \langle \vec{u}, \vec{i} \rangle. \tag{12}$$

Note that the higher the score $\hat{y}_{ui}$ is, the more likely that user u will visit POI i.

3.4 Negative Sampling Strategy

After describing the whole architecture of model DIG, the key design, geo-constrained negative sampling strategy, which can help to distinguish interest space and geographical space, is presented in this part. Negative sampling is widely used in the optimization process, since it can save time overhead and storage overhead while guaranteeing representative representation learning. However, utilizing a random sampling strategy only is not enough for disentanglement representation learning. That is, all negative samples are treated equally without differentiating interest factor and geographical factor. Since interactions between different user-POI pairs can be dominated by different factors, it is reasonable to choose representative negative samples for these two factors, respectively. However, it is impractical to divide a negative sample into the interest side or geographical side totally in the POI recommendation scenario. To address this challenge, a geo-constrained negative sampling strategy is proposed, which is composed of two parts, the geo-constrained sampling part and the random sampling part. Moreover, curriculum learning is used for learning accurate representations of interest space.

Geo-constrained Sampling Part

This design is inspired by an intuitive idea that un-visited POIs, which are near a specific user's visited areas, are not visited because of the influence of the interest factor rather than the geographical factor. Following this intuitive but reasonable idea, inference is made via leveraging geographical prior information, and triplet set $\mathcal{O}_1$ is specially sampled to benefit user interest space. $\mathcal{O}_1$ is defined as follows:

$$\mathcal{O}_1 = \{(u, i, j) | u \in \mathcal{U}, i \in I_u^+, j \in I_u^-, d(i, j) < d_m\}, \tag{13}$$

where d_m indicates distance margin, ensuring that negative samples in $\mathcal{O}_1$ locate in the nearby regions with positive items. Triplet set $\mathcal{O}_1$ is necessary for distinguishing interest factor and geographical factor, and experiment results will be shown in Section 4.4.2 to demonstrate its necessity.

Random Sampling Part. Although random sampling only is not enough for the disentangled representation learning process, without it can also lead to sub-optimal embedding learning. That is due to the fact that random sampling can cover the distribution of all the facts. Similar to Equation (13), triplet set $\mathcal{O}_2$ from the random sampling part is defined as:

$$\mathcal{O}_2 = \{(u, i, j) | u \in \mathcal{U}, i \in I_u^+, j \in I_u^-\}, \tag{14}$$

Curriculum Learning. Inspired by curriculum learning [30], which first feeds easy training data to the model and then gradually increases the difficulty, the area of regions valid for $\mathcal{O}_1$ is enlarged gradually in the training process. As mentioned that nearby POIs should contribute more to interest space, the curriculum learning is implemented via the following steps: (1) d_m is initialized with a small value, which returns negative examples that benefit interest space with higher confidence; (2) keep d_m for several epochs, get triplets from both $\mathcal{O}_1$ and $\mathcal{O}_2$, and feed them into DIG to optimize both interest representations and geography representations; (3) after several epochs, d_m will be expanded by a preset

expanding factor m_f; (4) repeat (2) and (3) until an early stop is triggered.

3.5 Optimization

After $\mathcal{O}_1$ and $\mathcal{O}_2$ are defined, parameter optimization with our proposed negative sampling strategy will be described in the following part. In this section, we will explain how disentangled representations of users and POIs are obtained via loss computation and gradient backward.

3.5.1 Choice of Weights Estimation Function f

Before diving into the details of loss computation, we will first explain how we select weights estimation function f, which balances loss optimization in interest space and geographical space ultimately. As weights of user interest factor and geographical factor vary greatly in encouraging check-in behaviors of users, it is critical to estimate the weights to ensure reasonable loss computation in different spaces. Recall that the interest factor should dominate the decision of user u in a small region where $d(i,j)$ is small, $i \in I_u^+$, $j \in I_u^-$. Therefore, we hope $f(d(i,j)) \to 0$ when $d(i,j) \to 0$. But for un-visited POI j with very large $d(i,j)$, $f(d(i,j))$ should satisfy $f(d(i,j)) \propto d(i,j)$. Moreover, the range of $f(d(i,j))$ is $[0,1)$ when $d(i,j)$ ranges from 0 to $+\infty$. Last but not least, $f(d(i,j))$ must be differentiable for all $d(i,j) \in [0,+\infty)$. Combining the above aspects, $f(d(i,j))$ is computed as follows:

$$f(d(i,j)) = \tanh(\lambda d(i,j)), \tag{15}$$

where λ is a hyper-parameter, and can be tuned easily during the experiment process.

3.5.2 Soft-Weighted Loss Computation in Sub-Spaces

As weights of the two mingled factors can be estimated by using f in the negative sampling process, we then propose the soft-weighted loss function to compute loss during optimization. For the purpose of learning model parameters, Bayesian personalized ranking (BPR) [31] loss is widely used in POI recommendation scenario [32]. It is computed as follows:

$$\mathrm{BPR}(\vec{u},\vec{i},\vec{j}) = -\sum_{(u,i,j)\in\mathcal{O}} \ln\sigma(\langle\vec{u},\vec{i}\rangle - \langle\vec{u},\vec{j}\rangle), \tag{16}$$

where $\mathcal{O}$ denotes triplet set generated by negative sampling, σ denotes sigmoid function, and $\langle\cdot,\cdot\rangle$ indicates the inner product of two vectors. Our proposed soft-weighted loss function weights BPR loss of interest space and BPR loss of geographical space according to the intuition: interest representations can dominate the user decision process in a small region, while geographical representations contribute more to remote POIs. Consider an extreme situation in which there are two next-door POIs, j_1, j_2, which are unvisited by user u. As they are next-door POIs, $d(j_1,j_2)$ can be regarded as 0. In addition, they share the same neighborhood. Thus, $\vec{j}_1^{\,\mathrm{geo}}$ should be similar to $\vec{j}_2^{\,\mathrm{geo}}$, and whether user u will check in j_1 or/and j_2 is depending on how much $\vec{j}_1^{\,\mathrm{int}}$ or/and $\vec{j}_2^{\,\mathrm{int}}$ match with $\vec{u}^{\,\mathrm{int}}$. That is why we think interest representations should dominate the decision process within small regions. For unvisited POI j far from all positive POIs in I_u^+, the geographical factor should play a more important role than the interest factor, as previous research has found that remote places will be checked in with low probability for the unbearable time cost on the way [9], and users tend to visit POIs close to their checked-in POIs in history [33]. That is why we assume the geography factor should impact more on remote decision making.

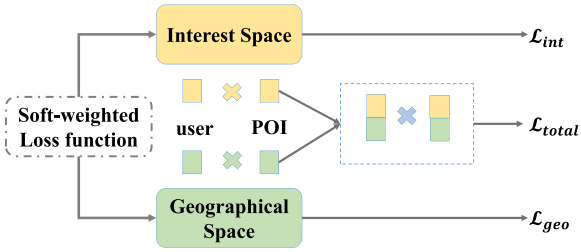


Fig. 2. An illustration of loss computation in DIG. The total losses consist of loss from the interest space, loss from the geographical space, and loss from the total space.

Therefore, the geo-enhanced soft-weighted loss function is computed as follows:

$$\mathcal{L}_{\mathrm{int}} = (1 - f(d(i,j)) \cdot \mathrm{BPR}(\vec{u}^{\,int}, \vec{i}^{\,int}, \vec{j}^{\,int}), \tag{17}$$

$$\mathcal{L}_{\mathrm{geo}} = f(d(i,j)) \cdot \mathrm{BPR}(\vec{u}^{\,geo}, \vec{i}^{\,geo}, \vec{j}^{\,geo}), \tag{18}$$

where f is the soft-weighted function, balancing user interest factor and POI geography factor under the geographical constraint $d(i,j)$.

Note that the triplet set $\mathcal{O}$ used in approach DIG is the union of $\mathcal{O}_1$ and $\mathcal{O}_2$. Although interest space benefits more from $\mathcal{O}_1$, $\mathcal{O}_2$ is not absent from calculating $\mathcal{L}_{\mathrm{int}}$. It is the same with $\mathcal{O}_2$ and $\mathcal{L}_{\mathrm{geo}}$. During the experiments, the size of the two triplet sets subjects to $|\mathcal{O}_1| : |\mathcal{O}_2| = d_i : d_g$. Therefore, both interest space and geographical space can receive balanced information after deploying our negative sampling strategy, and effective optimization process is guaranteed.

3.5.3 Loss Computation in the Total Space

Since the ranking layer of DIG uses the embedding of total space, the main task should be predicting user-POI interaction based on the embedding after concatenation. Therefore, BPR optimization should also be computed using the total-size embedding $\vec{u},\vec{i},\vec{j}$ for each triplet (u,i,j) in addition to the soft-weighted loss function used to compute $\mathcal{L}_{\mathrm{int}}$ and $\mathcal{L}_{\mathrm{geo}}$. Let $\mathcal{L}_{\mathrm{total}}$ denotes BPR loss of $(\vec{u},\vec{i},\vec{j})$, the loss is computed as:

$$\mathcal{L}_{\mathrm{total}} = \mathrm{BPR}(\vec{u}^{\,\mathrm{int}}||\vec{u}^{\,\mathrm{geo}}, \vec{i}^{\,\mathrm{int}}||\vec{i}^{\,\mathrm{geo}}, \vec{j}^{\,\mathrm{int}}||\vec{j}^{\,\mathrm{geo}}), \tag{19}$$

and illustrated in Fig. 2, the final loss function used in DIG is computed as:

$$\mathcal{L} = \mathcal{L}_{\mathrm{total}} + \alpha \cdot \mathcal{L}_{\mathrm{int}} + \beta \cdot \mathcal{L}_{\mathrm{geo}}, \tag{20}$$

where α and β are hyper-parameters. Since estimating user check-in behavior using the embedding of the total space is the main task, we set $0 < \alpha < 1, 0 < \beta < 1$ during the training process. By tuning α and β, embedding of interest space and geographical space will also be optimized via geographical prior knowledge.

TABLE 2
Basic Statistics of Two Datasets

Dataset	# Users	# POIs	# Records	Sparsity
Yelp	30,887	18,995	860,888	99.85%
Gowalla	22,224	18,442	701,500	99.82%

DIG decouples interest factor and geographical factor from user-POI interaction data and geographical data (longitude and latitude of POIs, distances between POIs) effectively via deploying modules mentioned above. Once auxiliary data, such as customized user options, is available, DIG can model them by adapting model input and disentangled embedding layer.

4 Experiments

Extensive experiments have been conducted on two real-world datasets to evaluate our proposed approach, DIG. We aim to answer four research questions as follows.

- *RQ1*: As DIG disentangles user interest factor and geographical factor for POI recommendation, how does DIG perform when compared with the state-of-the-art recommendation algorithms?
- *RQ2*: How to demonstrate the disentanglement of user interest factor and geographical factor?
- *RQ3*: How does our proposed geo-constrained negative sampling strategy benefit the disentanglement representation learning?
- *RQ4*: Since two explicable factors are decoupled in DIG, which factor is more important, the interest factor or the geographical factor?

4.1 Experimental Settings

4.1.1 Datasets

Two real-world public datasets, Yelp[1] and Gowalla,[2] are used in our experiments, which provide user-POI check-in data, check-in time, and coordinates of POIs. We filter out users with less than 10 POIs in the two datasets, and reserve POIs with more than 10 users in Yelp and users with more than 15 users in Gowalla. The whole dataset is split to training set, validation set, and testing set with the ratio of $7:1:2$ and is kept the same for all baselines. Statistics of these two datasets are shown in Table 2.

4.1.2 Evaluation Metrics

To compare the recommendation performance of different approaches, we choose to rank all items instead of generating a sampled subset [23], [27]. In other words, all the unvisited POIs I_u^- will be taken into account when a model is generating recommendation candidate set $\mathcal{C}_u$ for all users. Although this operation costs more time, the results are more reliable [36]. Two widely used top-K evaluation metrics [24], [37], Recall and Hit Ratio (HR) are adopted in the evaluation phase. Let $\mathcal{T}_u$ indicate the test positive set of user u, and Recall@K and HR@K are computed as follows:

$$\text{Recall@}K = \frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}}\frac{|\mathcal{T}_u \cap \mathcal{C}_u|}{|\mathcal{T}_u|}, \tag{21}$$

$$\text{HR@}K = \frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}}\frac{\mathbb{1}(|\mathcal{T}_u \cap \mathcal{C}_u| > 0)}{|\mathcal{T}_u|}, \tag{22}$$

where $\mathbb{1}$ is the indicating function, returning 1 when $\mathcal{T}_u \cap \mathcal{C}_u \neq \phi$ and 0 otherwise. Since only the first few POIs in candidate sets can be exposed to users with high probability, K is set to small numbers $3, 5, 10, 15$ in the evaluation phase following existing works [15], [34], [35].

4.1.3 Baselines

We compare the performance of DIG with multiple state-of-the-art approaches. These approaches include geography-related methods and a newly published disentangled representation learning method.

- *MF-BPR [31]* This model uses BPR loss to compute a pairwise loss function, and tries to complete empty entries of the user-item interaction matrix by computing the inner product of the corresponding user vector and item vector, which are learned using non-empty entries. This model can be very competitive when feeding in enough negative samples during our experiments.
- *APOIR-G [4]* APOIR is composed of a generator that generates candidates of POIs for each user, and a discriminator which evaluates candidates from the generator. The reward of this generating adversarial network is computed from social information and geographical information. By training the generator and discriminator together, it can achieve good performance in POI recommendation. To keep the comparison fair, we simplify APOIR to APOIR-G by removing the social-related part in reward computation and replacing GRU units with an MF-based representation learning module.
- *STGCN-G [38]* STGCN is a GNN-based POI recommendation approach, which encodes geographical information and timeline into GNN architecture. POIs are first mapped into different regions leveraging the geohash function while 24 time-bins are used to encode time information. Different spatial-temporal types are assigned with different edge types on the graph. To keep the comparison fair, we simplify STGCN to STGCN-G by removing the time module.
- *AutoInt* [39] AutoInt first maps both numerical features and categorical features into the same low-dimension space. Afterward, a multi-head neural network with a self-attentive design is deployed for the purpose of learning the high-order feature interactions of input features. Following [38], [40], we transform geographical information into regions via the geohash function.
- *xDeepFM* [41] This model combines a compressed interaction network with a classical deep neural network. During experiments, geographical information is encoded in the same way as AutoInt.
- *GCN [28]* This is a spectral GCN approach, introducing a propagation rule for neural networks which

1. https://www.yelp.com/dataset/
2. http://snap.stanford.edu/data/loc-gowalla.html

TABLE 3
Performance Comparison on the Yelp and Gowalla Datasets Where topK = 3, 5, 10, 15

		Yelp Dataset							
Group	**Method**	**Recall@3**	**HR@3**	**Recall@5**	**HR@5**	**Recall@10**	**HR@10**	**Recall@15**	**HR@15**
Baselines	APOIR-G	0.00934	0.04076	0.01460	0.06187	0.02692	0.10733	0.03782	0.14576
	STGCN-G	0.01020	0.04562	0.01628	0.07126	0.02820	0.11769	0.03989	0.15890
	Auto-Int	0.00105	0.00550	0.00174	0.00903	0.00351	0.01748	0.00530	0.02561
	xDeepFM	0.00250	0.01246	0.00383	0.01930	0.00682	0.03367	0.00949	0.04510
	MF-BPR	0.03144	0.14134	0.04566	0.18878	0.07294	0.26684	0.09486	0.32066
	GCN	0.01832	0.07926	0.02797	0.11611	0.04884	0.18858	0.06590	0.24178
	GCMC	0.01908	0.08337	0.02878	0.12109	0.04974	0.19266	0.06882	0.25240
	NGCF	0.02517	0.10516	0.03787	0.15123	0.06334	0.23422	0.08717	0.30298
	DGCF	0.02954	0.12010	0.04369	0.16949	0.07259	0.25998	0.09486	0.32066
	LightGCN	0.03266	0.13566	0.04809	0.18867	0.07846	0.27955	0.10368	0.34436
Our Model	DIG	**0.03557**	**0.15123**	**0.05168**	**0.20475**	**0.08335**	**0.29375**	**0.10913**	**0.35688**
% Improvement	-	8.91	11.48	7.47	8.52	6.23	5.08	5.26	3.64
		Gowalla Dataset							
Group	**Method**	**Recall@3**	**HR@3**	**Recall@5**	**HR@5**	**Recall@10**	**HR@10**	**Recall@15**	**HR@15**
Baselines	APOIR-G	0.01605	0.07379	0.02497	0.10759	0.04400	0.17472	0.06072	0.22620
	STGCN-G	0.01367	0.06520	0.02044	0.09526	0.03511	0.15398	0.04876	0.20329
	Auto-Int	0.00446	0.02249	0.00683	0.03406	0.01238	0.06052	0.01763	0.08383
	xDeepFM	0.01470	0.07132	0.02150	0.10088	0.03437	0.15348	0.04482	0.19402
	MF-BPR	0.04479	0.21384	0.06345	0.28033	0.09817	0.38427	0.12521	0.45178
	GCN	0.02560	0.12424	0.03727	0.17355	0.06056	0.26117	0.07883	0.32019
	GCMC	0.02606	0.12873	0.03798	0.17900	0.06153	0.26731	0.08117	0.32928
	NGCF	0.04070	0.18774	0.05729	0.25010	0.08937	0.35468	0.11685	0.42886
	DGCF	0.04282	0.19749	0.05988	0.26389	0.09194	0.36669	0.11750	0.43829
	LightGCN	0.04498	0.20929	0.06351	0.27841	0.09830	0.38706	0.12679	0.45970
Our Model	DIG	**0.04840**	**0.22707**	**0.06783**	**0.29612**	**0.10435**	**0.40395**	**0.13176**	**0.47132**
% Improvement	-	7.60	8.49	6.19	6.36	5.63	4.36	3.92	2.53

(The best performance is reported in bold fonts.)

can be operated on graphs. Fast approximate convolutions are achieved via Chebyshev polynomials. Therefore, it is applied for fast and scalable node classification problems by learning node embedding on the graph.

- *GCMC [42]* This model uses GCN encoder [28] to learn user embedding and item embedding, which transforms the interaction matrix completion problem to edge prediction on a user-item bipartite graph. However, only the similarity between connected nodes can be captured in GCMC.
- *NGCF [27]* This model encodes high-order CF signals by propagating messages between users and items over the bipartite graph. Nonlinear activation functions and embedding transformation matrices in GCN are reserved during propagation across layers.
- *LightGCN [29]* This model reserves neighborhood aggregation in GCN, but discards all nonlinear activation functions and weighting matrices. After this simplification, one-hop similarity and multi-hop similarity are captured well, and notable performance promotion is achieved.
- *DGCF [23]* This disentangled embedding approach decouples four inexplicable factors for POI recommendation, where each factor corresponds to a chunk of embedding. Model parameters are optimized using BPR loss optimization and L_2 norm regularization. Despite the good performance it achieves, the interpretability of disentanglement over POI datasets is weak.

4.1.4 Hyper-Parameter Settings

We implement our DIG model in Pytorch. The total embedding size of DIG remains the same with all baselines (i.e., $d_i + d_g = 128$) as in [23]. In this way, the model size of DIG is identical to MF, LightGCN, and DGCF, but is lighter than other baselines. Thus fair performance comparison is guaranteed [24]. All models are optimized with Adam optimizer [43]. As for hyper-parameters, learning rate is searched in $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$, and the number of layers is tuned amongst $\{1, 2, 3, 4\}$ for GCN based model. The number of negative samples is set equivalent, and positive check-ins are blocked during the negative sampling process for all methods in Table 3. In addition, negative sampling in DIG also observes $|\mathcal{O}_1| : |\mathcal{O}_2| = d_i : d_g = 32 : 96$.

4.2 Performance Comparison (RQ1)

Extensive experiments have been conducted to explore how disentangled embedding in DIG can improve the effectiveness of representation learning, and the best average performance is reported in Tables 3 and 4. In Table 3, we compare DIG with state-of-the-art approaches (mostly GNN-based approaches) on both Yelp and Gowalla. In Table 4, we compare DIG with state-of-the-art methods specific for POI

TABLE 4
Performance Comparison With Geo-Based Models on Yelp

Meodels	Recall@5	Recall@10	Recall@20	Recall@50
IRenMF [17]	0.02564	0.04923	0.06872	0.13590
GeoMF [18]	0.02718	0.05077	0.06974	0.14513
Rank-GeoMF [16]	0.02821	0.05179	0.07179	0.15538
Geo-Teaser [34]	0.02615	0.02779	0.04974	0.08205
GeoIE [3]	0.03282	0.05538	0.07590	0.15897
APOIR [4]	0.03538	0.06103	0.08359	0.16564
GPR [35]	0.03897	0.06410	0.08871	0.17231
DIG	**0.04042**	**0.06633**	**0.10674**	**0.19054**
Improvement %	3.73	3.48	20.32	10.58

(The best performance is reported in bold font.)

recommendation that consider geographical information. The observations are listed as follows.

- *DIG outperforms all baselines in all cases*. As shown in Table 3, DIG improves Recall@3 and HR@3 over LightGCN by $8.91\%, 11.48\%$ on Yelp and $7.60\%, 8.49\%$ on Gowalla. Therefore, the impact of disentanglement in improving representation learning is remarkable. The notable improvements on small K values can improve the user experience greatly because users tend to focus on the first few recommended items in most cases. Thus incorporating disentanglement into POI recommender systems is challenging but promising. Moreover, Table 4 shows the performance comparison between DIG and state-of-the-art geo-based POI recommendation approaches. As the Yelp dataset is exactly the same as the GPR [35] paper used, the results in [35] are directly used. Embedding size of DIG is set as 32 (the same size as used in Table 4), where $d_i = 8, d_g = 24$. As reported in Table 4, DIG outperforms all the geo-based approaches. In addition, although GPR distills information from geography together with time information, DIG outperforms GPR at Recall@20 by 20.32% and Recall@50 by 10.58%.
- *Leveraging geographical information is hard*. Most geographical based approaches, such as IRenMF [17], GeoMF [18], GeoIE [3] etc., don't achieve good performance on Yelp. APOIR-G does not achieve good performance, and this may result from removing social information in its reward computation process, which is critical in GANs. We try our best to implement them by ourselves, but we cannot obtain promising results. In summary, modeling geographical information in implicit ways is hard to guarantee good performance.
- *GNN-based models are powerful*. NGCF, DGCF, and LightGCN outperform most geo-based information in most cases as reported. Although these models leverage user-POI interaction only, the rule of information propagation between users and POIs help them perform well.
- *Disentanglement of explicable factors is more powerful*. Since DGCF shows weaker performance in all circumstances when it is compared with DIG, the advantage of decoupling interpretive factors is obvious. The number of intents is set to 4 in DGCF, which is consistent with the published best parameter for POI recommendation in [23]. This comparison also demonstrates the effectiveness of explicable disentanglement in representation learning.

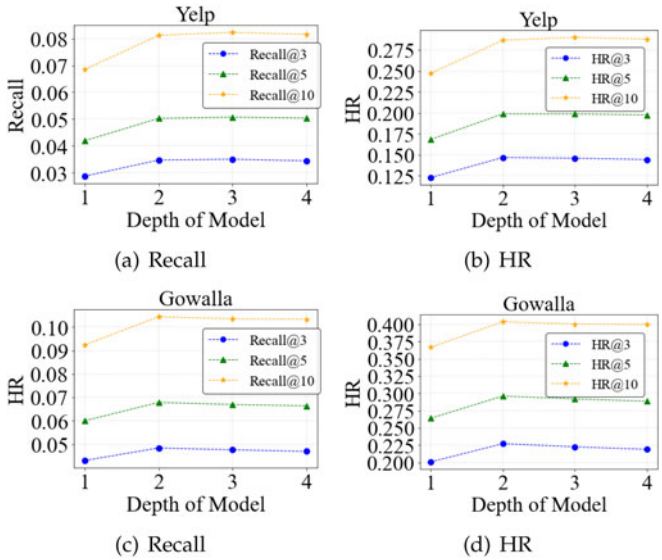


Fig. 3. Impact of model depth of DIG on Yelp and Gowalla.

4.3 Hyper-Parameters Study

In this part, the effects of hyper-parameters, especially those hyper-parameters specially designed for disentanglement, are investigated. We start by exploring the impact of layer number of DIG, and then we investigate the influence of dimension combination of disentanglement. Moreover, the effects of margin and region expanding factors are also analyzed. Finally, how α and β influence disentanglement balance and how λ affects the soft-weighted loss function is studied.

Effect of Depth. To investigate how DIG is affected by the number of layers, the depth of DIG is searched in the range of $\{1, 2, 3, 4\}$ to verify whether a deeper network can ensure better performance.As it is shown in Fig. 3, increasing layer number L from 1 to 2 for the 128-dim DIG model brings significant gain, but only negligible improvement when changing L from 2 to 3. Slightly worse performance is observed when comparing cases of $L = 4$ and $L = 2$. Moreover, training time and the number of parameters of DIG will increase when stacking more layers. Therefore, we choose $L = 2$ for DIG in this paper.

Effectofα, β, λ. To study how α and β influence loss balance in Equation (20) and how λ affects soft-weighted loss function computation, we search α and β together in the range of $\{0, 1e^{-4}, 1e^{-3}, 1e^{-2}, 1e^{-1}\}$ in a grid-search way, and tune λ among $\{\frac{1}{20}, \frac{1}{15}, \frac{1}{10}, \frac{1}{5}\}$. Note that the bigger λ is, the smaller the area in which the user interest factor dominates should be. It is found that $\lambda = \frac{1}{10}$ works well on two datasets. Table 5 shows the details of model performance on dataset Yelp at Recall@15 when tuning α and β. In summary, 1) keeping α constant, larger β ($\beta = 0.1$ or 0.01) will result in performance drops; 2) keeping β constant, larger α will lead to significant performance improvement; 3) the best performance is found with experiment setting $\alpha = 0.1, \beta = 0.0001$. Here, the influence of

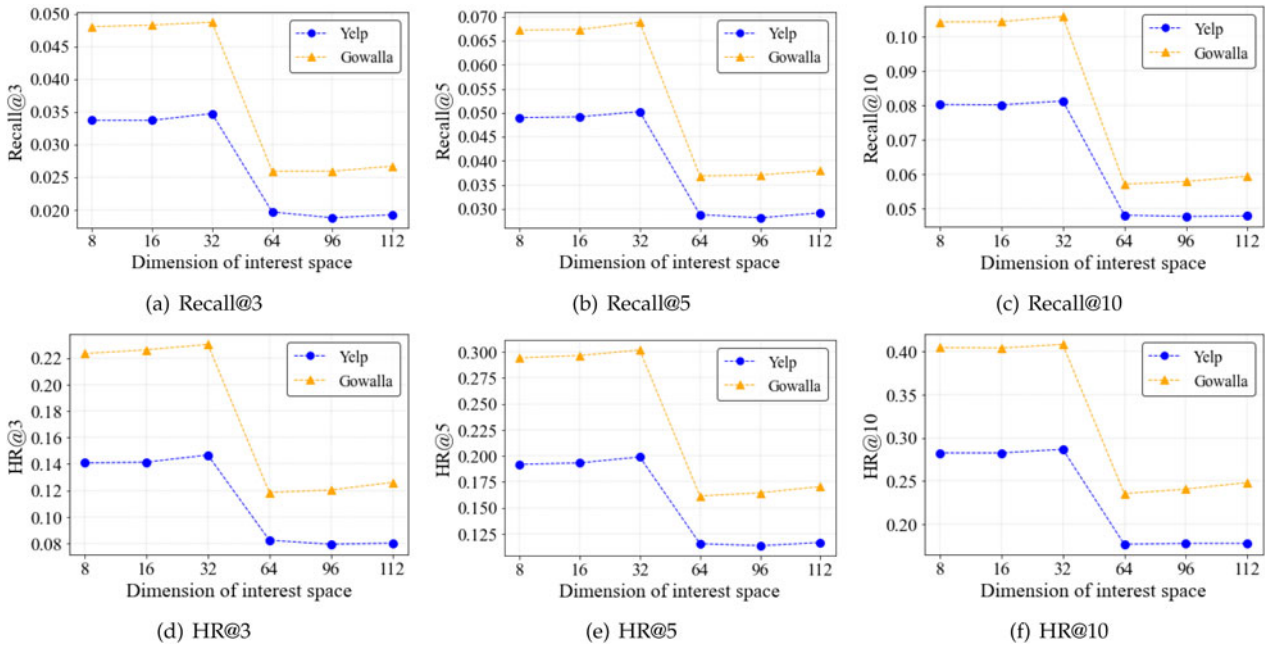


Fig. 4. Performance over Yelp and Gowalla when different dimension combinations are used. The dimension of total embedding is fixed as 128, and the best performance is achieved when the dimension of interest space is 32 on both datasets.

β is not only associated with its absolute value, but is also associated with the absolute values of gradients. Thus, small β is also effective and necessary. In addition, when implementing DIG in other datasets, these three hyper-parameters can be tuned in the same way mentioned above.

Dimension Combination. Existing disentangled representation learning approaches keep embedding size equivalent for each factor decoupled in their work. As factors in previous works are inexplicable in most scenarios, varying embedding sizes for different factors may only increase the workload on the parameter-tuning process. But it is unfair to keep embedding size equivalent in our proposed method, because the two interpretive factors, the user interest factor and POI geographical factor, might need different dimensions to achieve effective encoding. This is straightforward due to the different amounts of information carried by these two factors, and the factor with more information certainly deserves encoding with a higher dimension. Motivated by this, we search the space of different dimension combinations in this work. In addition, to keep the comparison in Table 3 fair, we hold $d_i + d_g = 128$. Different dimension combinations of $[d_i, d_g]$ are tried while other parameters remain the same. Fig. 4 showed the best dimension combination is $[32, 96]$.

Margin and Expanding Factor. In order to investigate how margin d_m and region expanding factor m_f affect the performance, experiments are conducted and repeated to report average performance, as shown in Fig. 5 and Tables 6 and 7. Although it seems hard to determine the initial value of d_m and the expanding factor m_f during the adaptive negative sampling process, experiments have demonstrated these two hyper-parameters are easily tuned. During the experiments, we found $d_m = 4$ km is a good initialization for Yelp dataset, while $dm = 2$ km is a good initialization for Gowalla dataset (shown in Fig. 5). Besides, $m_f = 1.1$ is also a good choice when tuned in the range of $\{1.1, 1.2, 1.3, 1.4, 1.5\}$ (shown in Tables 6 and 7). During training, d_m is expanded after every 8 epochs, where $d_m = d_m \times m_f$. Best performances are found with $m < 10$ km in most cases. Therefore, concerns about $d_{m_{\max}}$ is not necessary.

TABLE 5
Recall@15 of DIG With Grid-Search of α, β on Yelp

Recall@15	$\alpha = 0$	$\alpha = 0.0001$	$\alpha = 0.001$	$\alpha = 0.01$	$\alpha = 0.1$
$\beta = 0$	0.09703	0.10334	0.10871	0.10773	0.10660
$\beta = 0.0001$	0.10805	0.10809	0.10897	0.10887	0.10913
$\beta = 0.001$	0.10728	0.10593	0.10761	0.10680	0.10742
$\beta = 0.01$	0.09067	0.09071	0.08976	0.09198	0.08851
$\beta = 0.1$	0.09418	0.09546	0.09460	0.09393	0.09349

(The best values: $\alpha = 0.1, \beta = 0.0001$.)

4.4 Ablation Study

In this section, two groups of experiments are executed and reported. The first group of experiments is designed to dem-

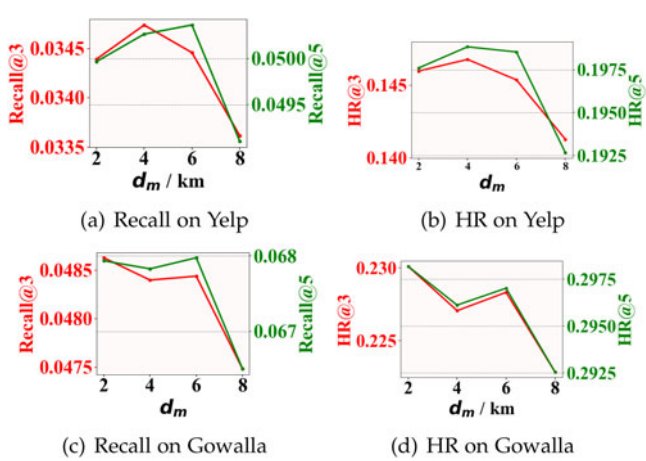


Fig. 5. Effects of d_m (initialized margin in negative sampling process) on Yelp and Gowalla.

TABLE 6
Effects of m_f (Margin Expanding Factor Which Controls Negative Sampling Process) on Yelp Dataset

m_f	Recall@3	HR@3	Recall@5	HR@5	Recall@10	HR@10
1.1	0.03474	0.14677	0.05027	0.19884	0.08120	0.28671
1.2	0.03358	0.14042	0.04898	0.19220	0.07957	0.28224
1.3	0.03314	0.13877	0.04867	0.19173	0.07960	0.28298
1.4	0.03345	0.13973	0.04910	0.19307	0.07958	0.28206
1.5	0.03295	0.13770	0.04849	0.19130	0.07833	0.28018

TABLE 7
Effects of m_f (Margin Expanding Factor Which Controls Negative Sampling Process) on Gowalla Dataset

m_f	Recall@3	HR@3	Recall@5	HR@5	Recall@10	HR@10
1.1	0.04840	0.22707	0.06783	0.29612	0.10435	0.40395
1.2	0.04727	0.22255	0.06624	0.29277	0.10280	0.40045
1.3	0.04652	0.21730	0.06545	0.28706	0.10189	0.39735
1.4	0.04593	0.21541	0.06508	0.28574	0.10086	0.39330
1.5	0.04638	0.21750	0.06541	0.28753	0.10120	0.39636

onstrate the effectiveness of disentanglement, while the second group of experiments is designed to evaluate the effectiveness of the proposed negative sampling strategy.

4.4.1 Effectiveness of the Disentanglement (RQ2)

Although DIG proves its superiority via the notable performance improvement on the two real-world public datasets, compared with the state-of-the-art approaches, there is still one important question that remains to be answered. How can we demonstrate the effectiveness of disentanglement? To answer this question, we designed nearby-region ranking evaluations for further demonstration. To be specific, areas are circled with the radius of $r = 1, 2, 3, 4, 5$ km, and centers are the locations of positive items in $\mathcal{T}_u, \forall u \in \mathcal{U}$. $\mathcal{T}_u$ in Equations (21) and (22) are no longer chosen from I_u^-, but is from the subset of $\{j | j \in \bigcup_{i \in \mathcal{T}_u} d(i,j) < r, j \in I_u^-\}$. During the evaluation phase, the first 32-dim vectors taken from LightGCN, named LGN-32, are compared with the 32-dim interest representations and the 96-dim geographical representations of DIG, which are named DIG-Int, DIG-Geo, respectively. This nearby-region evaluation is designed to check whether interest representations of DIG can dominate users' decision-making in nearby areas.

The observations from Fig. 6 are as follows:

- DIG-Int outperforms DIG-Geo and LGN-32 in all cases, showing its leading role in small regions when geographical influence can be regarded the same or only change a little. This result is consistent with our expectation of well-decoupled interest representations.
- DIG-Geo performs worst in all cases on Gowalla and in most cases on Yelp. The incapability of achieving accurate prediction in small regions relying on geography representations, is desirable indeed. The 96-dim representations taken out from DIG performs worse than the 32-dim representations from LightGCN in most case, implying the fact that the interest factor has hardly been captured in it.
- The consistent results on Yelp and Gowalla prove the effectiveness of disentanglement between user interest factor and POI geography factor in our proposed model.

4.4.2 The Effectiveness of Negative Sampling Strategy (RQ3)

To demonstrate the effectiveness of negative sampling strategy proposed in this paper, ablation study is conducted by removing $\mathcal{O}_1$ in Equation (13) and $\mathcal{O}_2$ in Equation (14) alternately.

According to the results shown in Tables 8 and 9, discarding triplet set $\mathcal{O}_2$, which is generated by random sampling, can greatly damage model performance. This is not surprising as random sampling helps nodes in the user-POI bipartite graph to reserve global similarity. Although the impact of $\mathcal{O}_1$ is not significant under all evaluation metrics on Yelp and Gowalla, it is indispensable in improving user experience due to the 4.5% gain on Yelp and 7.29% gain on Gowalla at Recall@3.

4.5 Analysis of Geographical Factor and Interest Factor (RQ4)

As two interpretable factors are disentangled in this work, one question arises naturally, which factor is more important in POI recommendation? Previous studies have found the significant impact of geographical factor [1], [3], [15], [18]. The best dimension combination is $d_i = 32, d_g = 96$, and it is consistent with the results of previous works, for higher dimensions implying a more important role played by the geography factor in POI recommendation. Combined with the results of nearby area evaluation, we can draw the conclusion that geographical factor contributes more in the concatenation embedding space, while user interest factor dominates check-in behaviors of users within small areas, where geographical constraint can be negligible.

5 Related Work

5.1 Point-of-Interest Recommendation

Geographical information of POIs plays an important role in POI recommendation. It has been found that users will visit the neighborhoods of their visited POIs with a higher probability [18], [44], [45]. Therefore, many approaches have been proposed, and they can be classified into two groups according to how they model geographical influence. The first group models geographical impact via self-defined quantitative calculation rules. Specifically, [1] proposed a way of calculating geographical-related score (added to the final prediction score) using the physical distance between visited POI and unvisited POI pairs, and integrated geographical scores of all such pairs to assist recommending candidates generation. Similarly, SAE-NAD [46] introduced a Gaussian radial basis function kernel together with inner product operations for distance modeling, inspired by the skip-gram model [47]. Moreover, [4] designed a generative adversarial network (GAN) and incorporated geographical information into the reward

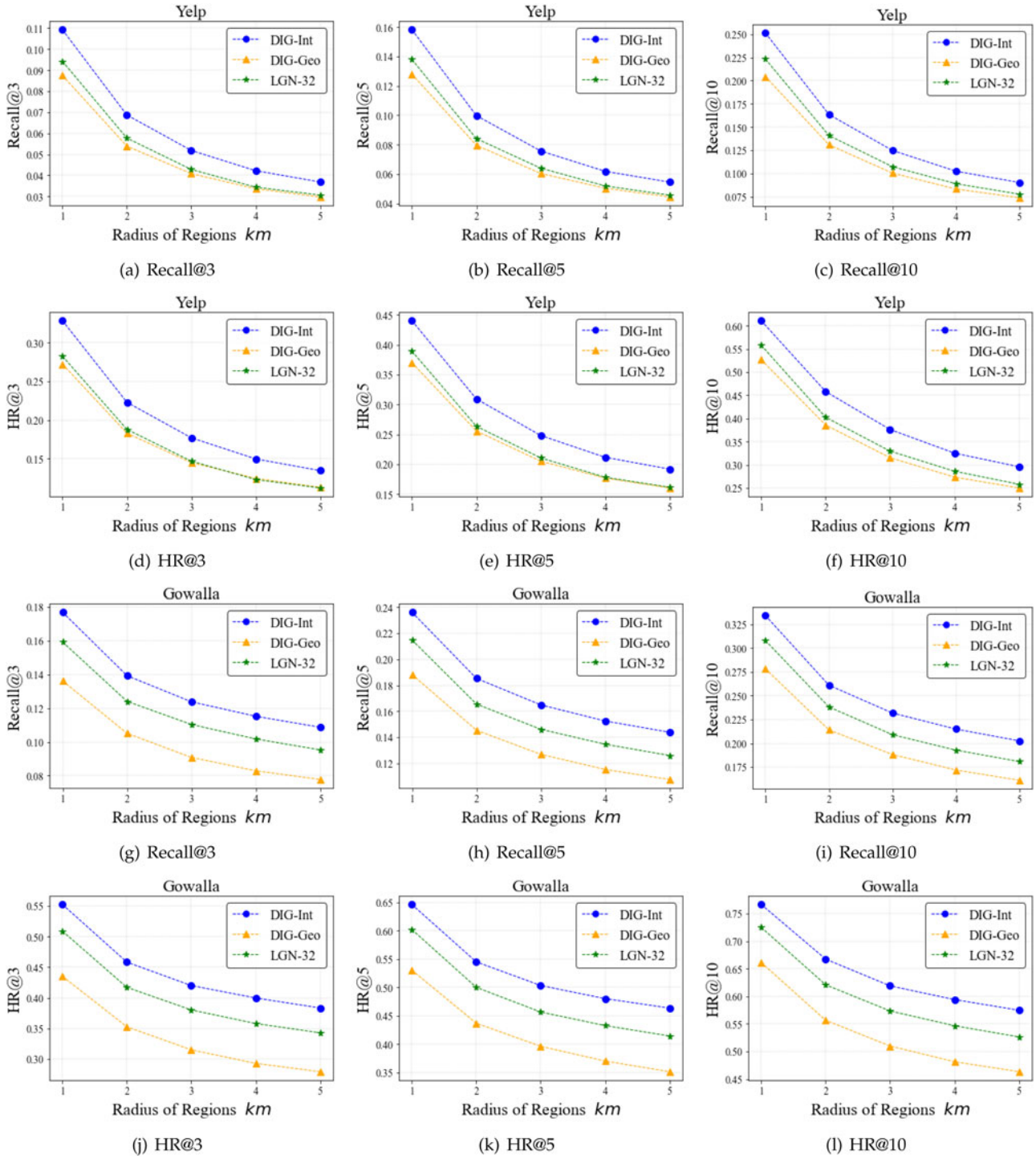


Fig. 6. Recommendation performance comparison of nearby-region evaluation over different region radius. Candidate set of each user is generated among those un-visited POIs which are near to his/her interacted POIs, constrained by the given radius of regions. Performance of DIG-Int is computed by using the 32-dim interest embedding of our DIG, DIG-Geo by the 96-dim geographical embedding of our DIG, while LGN-32 is computed using the first 32-dim embedding of [29].

computation process. Besides, GeoMF [18] and GeoMF++ [2] captured geographical influence via their self-defined rules, which leveraged both user check-in frequency and spatial distances between POIs. Despite the effectiveness of these models, the fusion of geographical factor with other factors (such as user preference reflected by visiting frequency, or friendships between users), can lead to sub-optimal representation learning. In contrast, the second group utilized latent vector/space for geographical influence modeling. For instance, GeoIE [3] utilized a latent vector space to model the POI in-going influence and out-going influence based on geographical distance, while Geo-ALM [5], another GAN-based POI recommendation approach, was proposed with a latent space to learn region similarity. Although all the above-mentioned works have exploited how to leverage geographical information in POI recommendation, the final

TABLE 8
Effect of Negative Sampling Strategy on Yelp

$\mathcal{O}$	Recall@3	HR@3	Recall@5	HR@5	Recall@10	HR@10
$\mathcal{O}_1$	0.01995	0.08231	0.02940	0.11640	0.04815	0.17807
$\mathcal{O}_2$	0.03323	0.13853	0.04866	0.19088	0.07851	0.27965
$\mathcal{O}_1 \cup \mathcal{O}_2$	0.03474	0.14677	0.05027	0.19884	0.08120	0.28671

TABLE 9
Effect of Negative Sampling Strategy on Gowalla

$\mathcal{O}$	Recall@3	HR@3	Recall@5	HR@5	Recall@10	HR@10
$\mathcal{O}_1$	0.04334	0.20674	0.06106	0.27190	0.09336	0.37308
$\mathcal{O}_2$	0.04511	0.21150	0.06397	0.28112	0.09844	0.38808
$\mathcal{O}_1 \cup \mathcal{O}_2$	0.04840	0.22707	0.06783	0.29612	0.10435	0.40395

representation learning was optimized by considering multiple factors together. Thus, they only modeled geography in an implicit way.

5.2 Graph Neural Networks for Recommendation

Another research line is graph neural networks (GNN), which have shown their superiority over different recommendation tasks. With the strong power of learning representations, GNN has been widely applied in recommender systems [26], [28], [48], [49], [50]. First, GCN [28] demonstrated its efficiency on the semi-supervised node classification task by using the information propagation rule between neighbors on the graph. Then, GraphSAGE [48] used aggregators to learn node features from its neighborhood, and its model layers could also help to capture graph structure. Subsequently, NGCF [27] was proposed, incorporating high-order collaborative signals by multi-hop message propagation between users and items on the graph. For the purpose of implementing GNN models over large-scale datasets and online recommendation systems, PinSAGE [49] proposed a new sampling strategy, which sampled a fixed-size neighborhood and computed different importance scores by random walk. Meanwhile, SGCN [50] simplified GCN [28] by choosing linear activation functions and collapsing weighting matrices of different layers to one weighting matrix. Moreover, LightGCN [29] only reserved neighborhood aggregation to enable information propagated over bipartite graph structure. Although non-linearity and extra weighting matrices were all discarded, LightGCN obtained remarkable performance improvement over two real-world public datasets. In our work, we leveraged the information propagation rule used in LightGCN [29], and extended its architecture to the disentangled representation learning process.

5.3 Disentangled Embedding for Recommendation

Disentangled embedding has first shown its superiority in computer vision area, such as [51], [52], [53]. For instance, [54] and [55] learned disentangled components according to different categories, [56] emphasized different viewpoints, and [57] captured different components in videos via disentanglement. One earlier work, DisenGCN [22], disentangled node embedding in the graph by deploying their proposed neighborhood routing mechanism to the GCN model, and the convergence of neighborhood routing was proved theoretically. Subsequently, this representation learning paradigm was explored in recommendation scenarios [23], [24], [58] as robust embedding of users and/or items could improve performance in recommendation. Moreover, DGCF [23] decoupled different intent factors and added an independence regularization module to guarantee the disentanglement. Despite the novelty and effectiveness of existing approaches, most of them suffered from lacking convinced interpretability of decoupled factors. In other words, although the core of disentangled embedding was to decouple different factors, most of the existing methods treated the number of factors K as a hyper-parameter, which made those factors physically inexplicable. Specifically, MacridVAE [58] manually interpreted different factors on a shopping dataset, but no similar analysis was reported over movie-related datasets. Differently, DICE [24] successfully disentangled two explicable factors, the user interest factor, and the conformity factor. However, its framework could not handle the unique geographical features of POI recommendation. Contrasting with existing methods, we explored different embedding sizes to develop the disentangling approach in POI recommendation. Moreover, the disentanglement of user interest factor and geographical factor was physically explicable, which also distinguished our method from the existing work.

6 Discussions

More Auxiliary Data. Real-world recommender systems may describe the properties of POIs, including brands, categories, etc., which can serve as auxiliary data for improving recommendation. These customized options are helpful in recommender systems, and they can provide a better user experience. However, such customized-options data is not available in public datasets, including the Yelp dataset and the Gowalla dataset. If such filter data is available, our DIG can also use it. Specifically, such auxiliary data can be used in the embedding layer of DIG by the projection of corresponding embedding matrices. Moreover, designing DIG as a two-factor disentangled model (decouples user interest and geography in POI recommendation) is motivated by two reasons as follows. (1) Geographical information of POIs (longitude and latitude) affect user behavior greatly, as shown in existing works of POI recommendation [3], [18], [32], [35]; (2) Geographical information is available in most datasets, compared with properties of POIs and customized filters.

The Relation Between Geography-Aware Preference and Interest-Aware Preference.Users can make decisions based on many aspects, such as timeline, the actual properties of the POI, the user's property, etc. In this paper, we divide user preferences towards these aspects into two parts, the geographical part and the non-geographical part (we name it as interest). In other words, the interest factor mentioned in our paper is presumed to have a broader scope. It is worth mentioning that there is a similar definition in Section 2.3 of [11].

The Geographical Factor in POI Recommendation. There are many situations in which people might choose longer-

distance travels. However, from the perspective of visiting probability, short-distance POIs are preferred. Previous works [3], [18], [32], [35] have reached a consensus that users tend to visit nearby areas with a higher probability. As the first law of geography stated [12], "everything is related to everything, but near things are more related than distant things." In other words, nearby unvisited POIs will be visited with a higher probability, and remote POIs with a lower probability.

7 Conclusion and Future Work

In this work, we proposed DIG, a GNN-based disentangled representation learning approach for POI recommendation task, which has successfully disentangled the user interest factor and POI geographical factor. The disentanglement is ensured by deploying our proposed geo-constrained negative sampling strategy and the geo-enhanced soft-weighted loss function to DIG during training. We also conducted extensive experiments on two real-world LBSN datasets, Yelp and Gowalla, sets, and the results have demonstrated the superiority of DIG. Besides, despite the effectiveness of our proposed DIG working as a two-factor disentangled model, it should be extended to a multi-factor disentangled model to incorporate temporal influence, etc.

In the future, we will exploit the multi-factor disentangled representation learning model in POI recommendation scenario, which will take side information such as timeline, user hobbies, etc. into consideration for the purpose of understanding user preference better as well as generating more accurate recommendation.

References

[1] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee, "Exploiting geographical influence for collaborative point-of-interest recommendation," in *Proc. 34th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2011, pp. 325–334.

[2] D. Lian, K. Zheng, Y. Ge, L. Cao, E. Chen, and X. Xie, "GeoMF++ scalable location recommendation via joint geographical modeling and matrix factorization," *ACM Trans. Inf. Syst.*, vol. 36, no. 3, pp. 1–29, 2018.

[3] H. Wang, H. Shen, W. Ouyang, and X. Cheng, "Exploiting POI-specific geographical influence for point-of-interest recommendation," in *Proc. Int. Joint Conf. Artif. Intell.*, 2018, pp. 3877–3883.

[4] F. Zhou, R. Yin, K. Zhang, G. Trajcevski, T. Zhong, and J. Wu, "Adversarial point-of-interest recommendation," in *Proc. World Wide Web Conf.*, 2019, pp. 3462–34618.

[5] W. Liu, Z.-J. Wang, B. Yao, and J. Yin, "Geo-ALM: POI recommendation by fusing geographical information and adversarial learning mechanism," in *Proc. Int. Joint Conf. Artif. Intell.*, 2019, pp. 1807–1813.

[6] H. Liu et al., "EDMF: Efficient deep matrix factorization with review feature learning for industrial recommender system," *IEEE Trans. Ind. Informat.*, vol. 18, no. 7, pp. 4361–4371, Jul. 2022.

[7] D. Li et al., "CARM: Confidence-aware recommender model via review representation learning and historical rating behavior in the online platforms," *Neurocomputing*, vol. 455, pp. 283–296, 2021.

[8] H. Liu et al., "Multi-perspective social recommendation method with graph representation learning," *Neurocomputing*, vol. 468, pp. 469–481, 2022.

[9] W. R. Tobler, "A computer movie simulating urban growth in the detroit region," *Econ. Geogr.*, vol. 46, pp. 234–240, 1970.

[10] W.-Y. Zhu, W.-C. Peng, L.-J. Chen, K. Zheng, and X. Zhou, "Modeling user mobility for location promotion in location-based social networks," in *Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2015, pp. 1573–1582.

[11] B. Liu, Y. Fu, Z. Yao, and H. Xiong, "Learning geographical preferences for point-of-interest recommendation," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2013, pp. 1043–1051.

[12] H. J. Miller, "Tobler's first law and spatial analysis," *Ann. Assoc. Amer. Geographers*, vol. 94, no. 2, pp. 284–289, 2004.

[13] N. Lim, B. Hooi, S.-K. Ng, Y. L. Goh, R. Weng, and R. Tan, "Hierarchical multi-task graph recurrent network for next POI recommendation," in *Proc. 45th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2022, pp. 1133–1143.

[14] M. Fan, Y. Sun, J. Huang, H. Wang, and Y. Li, "Meta-learned spatial-temporal POI auto-completion for the search engine at Baidu maps," in *Proc. 27th ACM SIGKDD Conf. Knowl. Discov. Data Mining*, 2021, pp. 2822–2830.

[15] B. Chang, Y. Park, D. Park, S. Kim, and J. Kang, "Content-aware hierarchical point-of-interest embedding model for successive POI recommendation," in *Proc. Int. Joint Conf. Artif. Intell.*, 2018, pp. 3301–3307.

[16] X. Li, G. Cong, X.-L. Li, T.-A. N. Pham, and S. Krishnaswamy, "Rank-GeoFM: A ranking based geographical factorization method for point of interest recommendation," in *Proc. 38th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2015, pp. 433–442.

[17] Y. Liu, W. Wei, A. Sun, and C. Miao, "Exploiting geographical neighborhood characteristics for location recommendation," in *Proc. 23rd ACM Int. Conf. Conf. Inf. Knowl. Manage.*, 2014, pp. 739–748.

[18] D. Lian, C. Zhao, X. Xie, G. Sun, E. Chen, and Y. Rui, "GeoMF: Joint geographical modeling and matrix factorization for point-of-interest recommendation," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2014, pp. 831–840.

[19] C. Li et al., "Multi-interest network with dynamic routing for recommendation at Tmall," in *Proc. 28th ACM Int. Conf. Inf. Knowl. Manage.*, 2019, pp. 2615–2623.

[20] C. Gao, X. Wang, X. He, and Y. Li, "Graph neural networks for recommender system," in *Proc. 15th ACM Int. Conf. Web Search Data Mining*, 2022, pp. 1623–1625.

[21] J. Wang, P. Huang, H. Zhao, Z. Zhang, B. Zhao, and D. L. Lee, "Billion-scale commodity embedding for e-commerce recommendation in Alibaba," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2018, pp. 839–848.

[22] J. Ma, P. Cui, K. Kuang, X. Wang, and W. Zhu, "Disentangled graph convolutional networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 4212–4221.

[23] X. Wang, H. Jin, A. Zhang, X. He, T. Xu, and T.-S. Chua, "Disentangled graph collaborative filtering," in *Proc. 43rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2020, pp. 1001–1010.

[24] Y. Zheng, C. Gao, X. Li, X. He, Y. Li, and D. Jin, "Disentangling user interest and conformity for recommendation with causal embedding," in *Proc. World Wide Web Conf.*, 2021, pp. 2980–2991.

[25] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30–37, 2009.

[26] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua, "Neural collaborative filtering," in *Proc. World Wide Web Conf.*, 2017, pp. 173–182.

[27] X. Wang, X. He, M. Wang, F. Feng, and T.-S. Chua, "Neural graph collaborative filtering," in *Proc. 42nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2019, pp. 165–174.

[28] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. 5th Int. Conf. Learn. Representations*, 2017, pp. 1–14.

[29] X. He, K. Deng, X. Wang, Y. Li, Y. Zhang, and M. Wang, "LightGCN: Simplifying and powering graph convolution network for recommendation," in *Proc. 43rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2020, pp. 639–648.

[30] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, 2009, pp. 41–48.

[31] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, "BPR: Bayesian personalized ranking from implicit feedback," in *Proc. 25th Conf. Uncertainty Artif. Intell.*, 2009, pp. 456–461.

[32] S. Xu, X. Fu, J. Cao, B. Liu, and Z. Wang, "Survey on user location prediction based on geo-social networking data," in *Proc. World Wide Web Conf.*, 2020, pp. 1621–1664.

[33] C. Cheng, H. Yang, I. King, and M. Lyu, "Fused matrix factorization with geographical and social influence in location-based social networks," in *Proc. AAAI Conf. Artif. Intell.*, 2012, pp. 17–23.

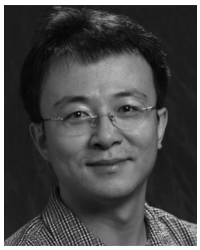
[34] S. Zhao, T. Zhao, I. King, and M. R. Lyu, "Geo-teaser: Geo-temporal sequential embedding rank for point-of-interest recommendation," in *Proc. 26th Int. Conf. World Wide Web Companion*, 2017, pp. 153–162.

[35] B. Chang, G. Jang, S. Kim, and J. Kang, "Learning graph-based geographical latent representation for point-of-interest recommendation," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage.*, 2020, pp. 135–144.
[36] W. Krichene and S. Rendle, "On sampled metrics for item recommendation," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2020, pp. 1748–1757.
[37] P. Sun, L. Wu, and M. Wang, "Attentive recurrent social recommendation," in *Proc. 41st Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2018, pp. 185–194.
[38] H. Han et al., "STGCN: A spatial-temporal aware graph learning method for poi recommendation," in *Proc. IEEE Int. Conf. Data Mining*, 2020, pp. 1052–1057.
[39] W. Song et al., "AutoInt: Automatic feature interaction learning via self-attentive neural networks," in *Proc. 28th ACM Int. Conf. Inf. Knowl. Manage.*, 2019, pp. 1161–1170.
[40] J. Jin, Z. Xiao, Q. Qiu, and J. Fang, "A geohash based Place2vec model," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 3344–3347.
[41] J. Lian, X. Zhou, F. Zhang, Z. Chen, X. Xie, and G. Sun, "xDeepFM: Combining explicit and implicit feature interactions for recommender systems," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2018, pp. 1754–1763.
[42] R. V. D. Berg, T. N. Kipf, and M. Welling, "Graph convolutional matrix completion," in *Proc. 23th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2017, pp. 51–56.
[43] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Representations*, 2015, pp. 1–15.
[44] H. Wang, M. Terrovitis, and N. Mamoulis, "Location recommendation in location-based social networks using user check-in data," in *Proc. 21st ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, 2013, pp. 374–383.
[45] J.-D. Zhang and C.-Y. Chow, "GeoSoCa: Exploiting geographical, social and categorical correlations for point-of-interest recommendations," in *Proc. 38th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2015, pp. 443–452.
[46] C. Ma, Y. Zhang, Q. Wang, and X. Liu, "Point-of-interest recommendation: Exploiting self-attentive autoencoders with neighbor-aware influence," in *Proc. 27th ACM Int. Conf. Inf. Knowl. Manage.*, 2018, pp. 697–706.
[47] M. Tomás, C. Kai, C. Greg, and D. Jeffrey, "Efficient estimation of word representations in vector space," in *Proc. 1st Int. Conf. Learn. Representations*, 2013, pp. 1–12.
[48] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 1025–1035.
[49] R. Ying, R. He, K. Chen, P. Eksombatchai, W. L. Hamilton, and J. Leskovec, "Graph convolutional neural networks for web-scale recommender systems," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2018, pp. 974–983.
[50] F. Wu, A. Souza, T. Zhang, C. Fifty, T. Yu, and K. Weinberger, "Simplifying graph convolutional networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6861–6871.
[51] T. Liu, J. Wang, B. Yang, and X. Wang, "NGDNet: Nonuniform Gaussian-label distribution learning for infrared head pose estimation and on-task behavior understanding in the classroom," *Neurocomputing*, vol. 436, pp. 210–220, 2021.
[52] H. Liu, T. Liu, Z. Zhang, A. K. Sangaiah, B. Yang, and Y. Li, "ARHPE: Asymmetric relation-aware representation learning for head pose estimation in industrial human–computer interaction," *IEEE Trans. Ind. Informat.*, vol. 18, no. 10, pp. 7107–7117, Oct. 2022.
[53] H. Liu, T. Liu, Y. Chen, Z. Zhang, and Y.-F. Li, "EHPE: Skeleton cues-based Gaussian coordinate encoding for efficient human pose estimation," *IEEE Trans. Multimedia*, early access, Aug. 08, 2022, doi: 10.1109/TMM.2022.3197364.
[54] S. Gidaris, P. Singh, and N. Komodakis, "Unsupervised representation learning by predicting image rotations," in *Proc. 6th Int. Conf. Learn. Representations*, 2018, pp. 1–16.
[55] E. Dupont, "Learning disentangled joint continuous and discrete representations," in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst.*, 2018, pp. 708–718.
[56] S. A. Eslami et al., "Neural scene representation and rendering," *Science*, vol. 360, no. 6394, pp. 1204–1210, 2018.
[57] J.-T. Hsieh, B. Liu, D.-A. Huang, L. Fei-Fei, and J. C. Niebles, "Learning to decompose and disentangle representations for video prediction," in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst.*, 2018, pp. 515–524.
[58] J. Ma, C. Zhou, P. Cui, H. Yang, and W. Zhu, "Learning disentangled representations for recommendation," in *Proc. 14th ACM Conf. Recommender Syst.*, 2020, pp. 43–52.

Yingrong Qin received the BS degree from the School of Information Technology and Engineering, Shandong University, in 2016. She is currently working toward the PhD degree with the Department of Electronic Engineering, Tsinghua University. Her interests include representation learning and recommender systems.

Chen Gao received the bachelor's and PhD degrees from the Department of Electronic Engineering, Tsinghua University and Huawei Noah's Ark Lab, in 2016 and 2021, respectively. He is now a postdoc researcher with the Department of Electronic Engineering, Tsinghua University and Huawei Noah's Ark Lab. His research mainly focuses on data mining and information retrieval, especially on recommender system. He has more than 30 publications in journals and conferences such as the *IEEE Transactions on Knowledge and Data Engineering*, *ACM Transactions on Knowledge Discovery from Data*, SIGIR, WWW, ICLR, KDD, IJCAI, etc.

Yue Wang received the BSc and PhD degrees from the Electronic Engineering Department, Tsinghua University, in 1999 and 2005, respectively. He is currently an Associate Professor with Tsinghua University. His research interests include computer networks, data fusion, and complex networks.



Shuangqing Wei (Senior Member, IEEE) received the BE and ME degrees in electrical engineering from Tsinghua University, Beijing, China, in 1995 and 1998, respectively, and the PhD degree from the University of Massachusetts, Amherst, MA, USA, in 2003. He is currently a tenured professor with the School of Electrical Engineering and Computer Science, Louisiana State University (LSU), Baton Rouge, LA, USA, where he also holds the Michel B. Voorhies distinguished professorship of electrical engineering. His current research interests include information theory, communication theory, and high-dim statistical inference in complex systems and networks, as well as the development of theoretical understanding about deep neural networks.

Depeng Jin (Member, IEEE) received the BS and PhD degrees from Tsinghua University, Beijing, China, in 1995 and 1999, respectively, both in electronics engineering. Now, he is a professor with Tsinghua University and vice chair with the Department of Electronic Engineering. He was awarded National Scientific and Technological Innovation Prize (Second Class) in 2002.

Jian Yuan received the MS degree in signals and systems from Southeast University, Nanjing, China, in 1989, and the PhD degree in electrical engineering from the University of Electronic Science and Technology of China, in 1998. He holds the position of professor of electronic engineering with Tsinghua University, Beijing, China. His main interests include complexity science and Urban computing.

Lin Zhang received the BSc, MSc, and PhD degrees from Tsinghua University, Beijing, in 1998, 2001, and 2006, respectively. He was a visiting professor with the University of California at Berkeley, from 2011 to 2013. His research interests include efficient protocols for sensor networks, statistical learning and data mining algorithms for sensory data processing, and information theory. Since 2006, he has been implementing wireless sensor networks in a wide range of application scenarios, including underground mine security, precision agriculture, industrial monitoring, and also the 2008 Beijing Olympic Stadium (the Bird's Nest) structural security surveillance Project and a metropolitan area sensing and operating network (MASON) in Shenzhen. He received the IEEE/ACM SenSys 2010 Best Demo Awards, the IEEE/ACM IPSN 2014 Best Demo Awards, and the IEEE CASE 2013 Best Paper Awards, and the Excellent Teacher Awards from Tsinghua University, in 2004 and 2010.