Feature Article

Personalized Geographical Influence Modeling for POI Recommendation

Yanan Zhang

Soochow University

Guanfeng Liu

Macquarie University

An Liu

Soochow University

(Corresponding author: anliu@suda.edu.cn)

Yifan Zhang

Soochow University

Zhixu Li

Soochow University

Xiangliang Zhang

King Abdullah University of Science and Technology

Qing Li

The Hong Kong Polytechnic University

Abstract—Point-of-interest (POI) recommendation has great significance in helping users find favorite places from a large number of candidate venues. One challenging in POI recommendation is to effectively exploit geographical information since users usually care about the physical distance to the recommended POIs. Though spatial relevance has been widely considered in recent recommendation methods, it is modeled only from the POI perspective, failing to capture user personalized preference to spatial distance. Moreover, these methods suffer from a diversity-deficiency problem since they are often based on collaborative filtering which always favors popular POIs. To overcome these problems, we propose in this paper a personalized geographical influence modeling method called PGIM, which jointly learns users' geographical preference and diversity preference for POI recommendation. Specifically, we model geographical preference from three aspects: user global tolerance, user local tolerance, and spatial distance. We also extract user diversity preference from interactions among users for diversity-promoting recommendation. Experimental results on three real-world datasets demonstrate the superiority of PGIM.

Index Terms: H.2.8.d Data mining, I.2.6.g Machine learning

■ **IN THE ERA** of information explosion, fast extracting valuable information from massive data is useful but challenging. POI recommendation is significant in helping users find favorite places from a large number of candidate venues [1], [2]. Early recommendation methods [3] are built on Collaborative Filtering (CF) whose basic idea is users usually like the places visited by others with similar tastes. A recent advance called Neural Collaborative Filtering (NCF) [4] combines CF with deep neural network to better capture complex and non-linear relations among users. Furthermore, geographical information is also widely exploited in POI recommendation [2], [5] since user preference over POIs has typically strong geographical constraints.

One weakness of current POI recommendation methods is that they fail to model personalized distance influence. Although the spatial distance is widely considered in POI recommendation models [6], [7], it is regarded as a user-independent factor. That is, the spatial relevance of two POIs is a constant and only depends on their physical distance. As a result, in the example shown in Fig. 1, KFC and Starbucks are recommended since they are close to the user u . However, if u 's tolerance of distance happens to be high because he/she has a car, u may prefer Pizza Hut since it is a hot restaurant. This toy example implies that a user's tolerance for distance affects the user's geographical preference. In addition, we analyzed the Gowalla dataset and randomly selected 10 users who have checked in a given place. Check-ins of these 10 users near this place are shown in Fig. 2. It is clear from the figure that the geographical influence of the place is user-related, showing the significance of personalized geographical influence modeling in POI recommendation.

Another weakness of CF-based methods is that they bias popular POIs since CF essentially exploits common preferences between users [8]. Consider again the example shown in Fig. 1, assume user u likes KFC's taste, and KFC has only been visited by a few users. On the contrary, u dislikes Starbucks's taste, but Starbucks is welcomed by many users. In general CF-based models, Starbucks is more likely to be recommended to u due to its popularity. As popular

Figure 1. POI Recommendation. Three restaurants with different distances and popularities.

Figure 2. Check-in Distribution. Various colors are used to mark check-in frequencies near a given place.

POIs account for only a small part of all POIs, a favor to popular POIs makes CF-based models suffer from the lack of diversity. To address the problem, the diversity-promoting model [9] raises variousness of recommendation lists by utilizing the heuristic strategy or supervised learning to re-rank POIs based on a certain diversity metric. Generally, diversity and precision are opposites since popular POIs are more likely to appear in a user's check-in list, so the diversity-oriented strategy can achieve high diversity but with a huge sacrifice of precision.

To overcome these two weaknesses, we propose for POI recommendation a **personalized geographical influence modeling** method called PGIM, which captures personalized user distance tolerance for geographical influence and fuses interactions among users in NCF framework for diversity recommendation.

In PGIM, geographical distance influence is modeled from three aspects: user global tolerance, user local tolerance, and spatial distance. *Global tolerance* reflects the user's general tolerance for the distance between any pair of POIs, which is related to external factors such as traffic conve-

Table 1. Notation Descriptions.

Notation	Description
M, N	the number of users and POIs, respectively
U, I	the set of users and POIs, respectively
$\mathbf{R}$	the check-in matrix, $\mathbf{R} \in \mathbb{R}^{M \times N}$
$\mathbf{r}_u$	the check-in data for user u (u -th row of $\mathbf{R}$)
H_u	the set of POIs that user u has visited
H_i	the set of users who have visited POI i
$\mathcal{T}$	the tolerance tensor, $\mathcal{T} \in \mathbb{R}^{M \times N \times N}$
$\mathcal{G}$	the global tolerance tensor, $\mathcal{G} \in \mathbb{R}^{M \times N \times N}$
$\mathcal{L}$	the local tolerance tensor, $\mathcal{L} \in \mathbb{R}^{M \times N \times N}$
$\mathbf{D}_u$	the paired distance matrix for POIs in H_u
lon_i, lat_i	the longitude and latitude of POI i
d_{ij}	the distance between POI i and j
y_{ui}	the overall preference for u to i
y_{ui}^g	the geographical preference for u to i
y_{ui}^d	the diversity preference for u to i

nience. For example, if a user is in the center of the city with convenient transportation, it is likely to take the same time to walk to a restaurant 1km away as to take the bus to a restaurant 10km away. *Local tolerance* reveals a user's specific tolerance to the distance to a given POI, which is positively related to the user's preference over the POI. For example, assume user u has visited place i many times and place j only once. It is more likely for u to visit a POI near i , though it is farther than another POI around j . Global tolerance is affected by the external factors varying by users, while local tolerance is determined by user preference to POIs. It is therefore beneficial to merge these two kinds of tolerances to better capture user tolerance to distance. Furthermore, combining subjective user tolerance with objective spatial distance can get accurate modeling of spatial relevance.

On the other hand, we model *user diversity preference* to improve recommendation diversity. By taking users as documents and POIs as words, we combine the TF-IDF idea with NCF to learn user diversity preference from the check-ins of unpopular POIs. Different from traditional methods that search unpopular POIs heuristically from the whole data set, our method can raise the recommendation probability of long-tail terms that users are interested in, thus achieving diversity with high precision.

Preliminary

We use fancy uppercase letters to represent three-dimension tensors (i.e., $\mathcal{T}$), bold uppercase letters to represent matrices (i.e., $\mathbf{A}$) and bold lowercase letters to represent vectors (i.e., $\mathbf{a}_i$).

In particular, the lowercase letter a_{ij} corresponds to (i, j) -th element in the matrix $\mathbf{A}$ or the j -th element in the vector $\mathbf{a}_i$. Besides, uppercase letters represent sets (i.e., A). Tab.1 summarizes basic notations used in the paper. Based on it, our problem is defined as follow:

POI Recommendation. Given a set U of users with their check-ins $\mathbf{R}$ and a set I of POIs with their locations (lon, lat) , our goal is to obtain a set of POIs $\{i \mid i \in I\}$ for a user u , which contains k unvisited POIs that u is most interested in.

Proposed Method

As discussed earlier, our PGIM method can capture user diversity preference and user geographical preference simultaneously. We first introduce diversity preference which fairly reflects a user's tendency towards POIs regardless of their popularity. Then, we model user geographical preference by considering geographical information of POI. A three-dimension tensor $\mathcal{T}$ is created in this step where the element t_{uij} denotes user u 's tolerance to the distance between POIs i and j . Finally, we describe how to evaluate a user's *overall preference* to POIs based on diversity preference and geographical preference.

Diversity Preference

Fig.3 illustrates the architecture of *diversity preference modeling*, which aims at raising recommendation diversity through estimating user u 's diversity preference for POI i (i.e., y_{ui}^d).

Input and Embedding Layer. The Input Layer has three inputs: the one-hot vector $\mathbf{x}_u$ of the target user u , the one-hot vector $\mathbf{y}_i$ of the target POI i , and the one-hot vector $\mathbf{x}_v$ of each user v in historical users H_i . Then the Embedding Layer transforms these one-hot vectors into low-dimensional dense embeddings:

$$\mathbf{p}_u = \mathbf{x}_u \mathbf{P}, \quad \mathbf{q}_i = \mathbf{y}_i \mathbf{Q}, \quad \mathbf{p}_v = \mathbf{x}_v \mathbf{P}, \quad v \in H_i \setminus u \quad (1)$$

where $\mathbf{P} \in \mathbb{R}_+^{M \times K}$ and $\mathbf{Q} \in \mathbb{R}_+^{N \times K}$ represent embedding matrices of users and POIs, respectively. K denotes the embedding size. We utilize ReLU function to handle the constraints $\mathbf{P} \geq 0$ and $\mathbf{Q} \geq 0$ in Eq.1, that is, if $\mathbf{P} < 0$ or $\mathbf{Q} < 0$ during training iterations, we reset $\mathbf{P} = 0$ or $\mathbf{Q} = 0$.

Geographical Influence Modeling

Figure 3. Diversity Preference Modeling

Interaction Layer. Interactions between the target user and historical users are modelled explicitly on this layer. We use the ratio of corresponding elements in the two embedding vectors to capture the deep interaction relationship between the target user u and a historical user v :

$$\mathbf{rel}_{uv} = \left[\frac{p_{u1}}{p_{v1}}, \frac{p_{u2}}{p_{v2}}, \dots, \frac{p_{uK}}{p_{vK}} \right], v \in H_i \setminus u \quad (2)$$

where p_{uk} and p_{vk} represent k -th ($k \in [1, K]$) element of user embedding $\mathbf{p}_u$ and $\mathbf{p}_v$, respectively. Correspondingly, $\frac{p_{uk}}{p_{vk}}$ is the k -th element of the vector $\mathbf{rel}_{uv}$.

Our motivation here is inspired by TF-IDF, which reflects how important a word w is to a document d by exploring the relative frequency of w in d . Generally, if a user u has visited some unpopular POIs, then these places are more likely to reflect u 's personal preference. This fact illustrates that u 's preference to POI i decreases with the increase of other users' preference for i . Considering that positive embedding matrices ensure a positive correlation between user preference and user embedding, we conclude that when calculating u 's preference for i , u 's embedding should decrease with the increase of other users' embeddings. Therefore, we propose to learn user relative embedding for user's relative preference, which is defined as the ratio of user embedding to historical user embedding.

Pooling Layer. This layer integrates multiple user relative embeddings with the simple average pooling and receives user diversity embedding

$\mathbf{div}_{ui}$. The function is formulated as follow:

$$\mathbf{div}_{ui} = \frac{1}{|H_i| - 1} \left(\sum_{v \in H_i \setminus u} rel_{uv} \right) \quad (3)$$

Concat and Deep Interaction Layer. Inspired by NCF [4], we concatenate $\mathbf{div}_{ui}^T$ and $\mathbf{q}_i^T$ directly, followed by MLP to extract valuable information. The detailed process is defined as follows:

$$\mathbf{z}_L = \delta_L(\mathbf{W}_L \delta_{L-1}(\dots \delta_1(\mathbf{W}_1 \begin{bmatrix} \mathbf{div}_{ui}^T \\ \mathbf{q}_i^T \end{bmatrix} + \mathbf{b}_1) \dots) + \mathbf{b}_L) \quad (4)$$

where $\mathbf{W}_l$, $\mathbf{b}_l$ and $\mathbf{z}_l$ denote the weight matrix, the bias vector and the output vector in the l -th layer, respectively. δ_l is Rectified Liner Unit (ReLU) activation function.

Prediction Layer. The Prediction Layer infers u 's diversity preference score for i through a Fully Connected Layer, which takes $\mathbf{z}_L$ as the input and outputs a prediction value y_{ui}^d . We project the conversion function as follow:

$$y_{ui}^d = \mathbf{w}^T \mathbf{z}_L + b \quad (5)$$

where $\mathbf{w}$ and b denote the weight vector and the bias, respectively. In summary, y_{ui}^d reflects u 's diversity preference which eliminates the influence of common preference.

Geographical Preference

Given user u 's historical check-ins H_u , our purpose is to estimate u 's geographical preference for the unvisited POI j . We give concrete implementation steps below.

Global Tolerance. The spatial distance between two POIs in H_u implicitly reflects u 's acceptable distance range. Since global tolerance defines user general acceptance of spatial distance, we can obtain u 's tolerance based on the distance information contained in H_u .

Specifically, we take the paired distance matrix $\mathbf{D}_u$ as u 's input, where the (i, j) -th element represents the spatial distance between the i -th POI and the j -th POI in H_u . Then we employ CNN to extract high-order features among distances. Here, MLP is inappropriate since the interconnection rule of all the neurons in adjacent layers may result in many parameters when inputting a matrix. We favor CNN as it stacks layers in a parameter sharing manner.

Before convolution, the model introduces the zero padding which pads $\mathbf{D}_u$ to an $m \times m$ matrix $\mathbf{D}'_u$ where m is the number of check-ins of the user who has the most check-ins among all users. Then, the output feature map of the Convolution Layer at $\mathbf{D}'_u$ can be formulated as follow:

$$c_u^{f,i,j} = \phi\left(\sum_{a=0}^{E-1} \sum_{b=0}^{E-1} \mathbf{D}'_{Si+a, Sj+b} \cdot \mathbf{E}_{a,b}^f + b^f\right) \quad (6)$$

where the number of filters is denoted as F and the stride is S . Kernel matrix of filter f is $\mathbf{E}^f \in \mathbb{R}^{E \times E}$. $c_u^{f,i,j}$ denotes u 's (i, j) -th element of f -th feature map and $\phi()$ is the ReLU function. Then the max-pooling function is applied to the matrix $\mathbf{C}_u^f$ for dimension reduction.

The Convolution Layer can extract the local features of a matrix. To identify global feature, i.e. global tolerance, we introduce a Fully Connected Layer to aggregate these local parts. u 's global tolerance g_{uij} ((u, i, j) -th element of $\mathcal{G}$) for the distance between i in H_u and an unvisited POI j is:

$$g_{uij} = \text{ReLU}(\mathbf{W}_f \eta(t_u) + \mathbf{b}_f), i \in H_u, j \in V - H_u \quad (7)$$

where $\eta()$ denotes the flatten operation which flattens a tensor with size (h, w, c) to a vector with size $(1, h \times w \times c)$. $\mathbf{W}_f$ and $\mathbf{b}_f$ are weights and bias of the Fully Connected Layer, respectively. Note that the value of g_{uij} is not affected by i and j because u 's global tolerance is only related to u 's own characteristics.

Local Tolerance. Unlike global tolerance, local tolerance is directly related to i in H_u . Actually, local tolerance reflects the fact that if u prefers i , u may visit i 's neighbours. Fig.2 depicts user local tolerance by counting the check-in distribution of 10 users near POI i (i corresponds to '0km' in the x -axis). From the figure, the more times users check in i , the more frequencies they visit near i with a wider range. That is, u 's local tolerance for the distance between i in H_u and an unvisited POI j is positively relevant to u 's diversity preference to i . We denote u 's local tolerance l_{uij} ((u, i, j) -th element of $\mathcal{L}$) as follow:

$$l_{uij} \propto y_{ui}^d, i \in H_u, j \in V - H_u \quad (8)$$

where $\propto$ is the proportional symbol.

Geographical Preference. User geographical preference consists of three parts: global tolerance, local tolerance and spatial distance. Mathematically, u 's geographical preference for POI j is defined as follow:

$$\begin{aligned} y_{uj}^g &= \frac{1}{|H_u|} \sum_{i \in H_u} t_{uij} \psi(d_{ij}) \\ &= \frac{1}{|H_u|} \sum_{i \in H_u} (\alpha g_{uij} + \beta l_{uij}) \psi(d_{ij}) \end{aligned} \quad (9)$$

where t_{uij} ((u, i, j) -th element of $\mathcal{T}$) denotes u 's tolerance for the distance between i in H_u and an unvisited POI j . α and β are smoothing factors to trade-off global and local tolerance. Besides, $\psi(x) = a \cdot x^b$ is a distance function that denotes user growing preference with decreased distance, where a and b are hyper-parameters.

The Fusion of Overall Preference

To model the *overall preference* of a user for a POI, we combine user geographical preference and user diversity preference. Specifically, user u 's overall preference for POI i is denoted as follow:

$$y_{ui} = \rho y_{ui}^g + \varrho y_{ui}^d \quad (10)$$

where ρ and ϱ are used to distribute the importance of both. However, the rating in $\mathbf{R}$ is a kind of implicit feedback and only equals to 0 or 1. To fit $\mathbf{R}$ effectively, we deal with the preference score y_{ui} as follows:

$$s_{ui} = \sigma(y_{ui}) = \frac{1}{1 + e^{-y_{ui}}} \quad (11)$$

where $\sigma()$ denotes the sigmoid function which controls s_{ui} between 0 and 1. Finally, we make top- k recommendation to users by sorting all unvisited POIs and extracting the first k terms.

To make the prediction ranking fit $\mathbf{R}$, we adopt *binary cross-entropy* loss function to optimize the model. This method has superiority in accelerating optimization by randomly choosing quantitative negative samples for each positive check-in of user u (negative sampling) rather than all negative items. The loss function can be formalized as follow:

$$\begin{aligned} L &= \sum_{u \in U} \sum_{i \in I} \sum_{j \in \{i\} \cup \text{neg}(i)} \\ &\quad \left\{ \varsigma_{ij} \log [\sigma(y_{uj})] + (1 - \varsigma_{ij}) \log [1 - \sigma(y_{uj})] \right\} \end{aligned} \quad (12)$$

Geographical Influence Modeling

Table 2. Statistics of the evaluation datasets.

Dataset	#user	#POI	#rating	%rating
Yelp	26534	37161	197205	0.02%
Foursquare	9529	21862	312484	0.15%
Gowalla	2987	6576	130513	0.66%

where $neg(i)$ denotes u 's negative samples, ς_{ij} is an indicator function which returns 1 if i equals to j and 0 otherwise. Furthermore, Adam algorithm is used to optimize parameters.

Experiments

Datasets

We use three real-world datasets from different location-based social networks for evaluation: Yelp, Foursquare and Gowalla. The statistics of these datasets are shown in Tab.2. Each record in these datasets contains user ID, POI ID and POI location information. These datasets are pre-processed by eliminating users with less than 10 interactions and POIs with less than 20 interactions. For each user, we sort his/her check-in records chronologically and take the first 90% of check-ins as training data and the remaining 10% as testing data. POIs in the trainset are considered to be visited, while those in the testset are regarded as unvisited. We aim at maximizing the intersection between recommended POIs and unvisited POIs.

Evaluation Metrics

To evaluate the effectiveness of the proposed models, we adopt two wide-used evaluation metrics, namely, $pre@k$ and $rec@k$. Their definitions are as follows:

$$pre@k = \frac{1}{M} \sum_{u \in U} \frac{|R_u \cap V_u|}{|V_u|} \quad (13)$$

$$rec@k = \frac{1}{M} \sum_{u \in U} \frac{|R_u \cap V_u|}{|R_u|} \quad (14)$$

where k is the number of POIs recommended for each user. R_u represents user u 's real check-ins in the testset. V_u represents top- k recommended POIs for user u . Actually, $pre@k$ implies how many recommendation terms are visited truly and $rec@k$ implies how many ground check-ins are recommended.

Baseline Methods

To illustrate the effectiveness of PGIM, we use the following representative methods for comparison.

GeoMF [3]: It is a weighted matrix factorization model which considers the geographical aggregation phenomenon by augmenting user latent factors with user activity area vectors and POI latent factors with POI influence area vectors.

RankGeoFM [7]: It is a ranking based model which proposes an extensible method to incorporate the geographical influence and other context information.

NeuMF [4]: It is a deep neural network variant of MF, which combines *GMF* and *MLP* to learn user and item embeddings.

Geo-Teaser [5]: It proposes a geo-temporal sequential model which combines the temporal POI embedding model and the geographically hierarchical pairwise preference ranking model.

PACE [1]: It is a general and principled combination of CF and SSL based on neural networks.

GeoIE [2]: It integrates specific-POI geographical influence into a standard MF model which models the geographical influence between two POIs using three factors: the geo-influence of POI, the geo-susceptibility of POI, and their physical distance.

APOIR [10]: It unifies reinforcement learning and MF into an adversarial learning framework.

PGIM-D: It recommends POIs for users according to user diversity preferences. This is a special case of PGIM, i.e., when $\rho = 0$ in Eq.10.

PGIM-G: It recommends POIs for users according to users geographical preferences. This is another special case of PGIM, i.e., when the framework in Fig.3 no longer has Interaction Layer and Pooling Layer.

Among these methods, GeoMF, RankGeoFM, Geo-Teaser, GeoIE and PGIM-G capture geographical influence. Specially, Geo-Teaser employs BPR and the others adopt MF. The rest including NeuMF, PACE, APOIR and PGIM-D utilize neural network to improve the performance of recommendation.

Parameter Settings

The default values of some parameters are set as follows. For the learning rate γ , we set a small

Figure 4. Performance comparison on three datasets w.r.t. $pre@k$.

Figure 5. Performance comparison on three datasets w.r.t. $rec@k$.

Figure 6. Performance comparison w.r.t. $NDCG@k$ and diversity.

value 0.001 to ensure fine learning. For the neural network, we set the number of MLP layers to 3, the size of the layers to $32 \rightarrow 16 \rightarrow 8$, and the embedding size is 16. For the geographical influence, we set a to 0.01 and b to -0.02 in the distance function. For the model optimization, we set the batch size as a constant value 258 and model 8 negative samples for each positive entry. Besides, smoothing factors $\alpha, \beta, \rho, \varrho$ are initialized to 1 and constantly updated during model training. The recommendation list length

k is tested in $\{1, 5, 10\}$. For all baselines, we set their parameter values as suggested in the original papers.

Performance Comparison

Fig.4 and Fig.5 show $pre@k$ and $rec@k$ of all models on three datasets with different k values. By analysing the results, we have the following findings:

(1). PGIM consistently achieves the best performance on all datasets and significantly outperforms other state-of-the-art methods. Taking the Gowalla data for example (Fig.4(c) and Fig.5(c)), PGIM improves precision by 12.4% and recall by 22.8% on average compared with APOIR which performs the second best.

(2). PGIM-G defeats the corresponding comparative models. In particular, it performs better than GeoIE which stands out among the methods for comparison. The success comes from PGIM-G's capability of capturing veritable requirements with personalized geographical tolerance. Besides, PGIM-D performs better than PACE and

Geographical Influence Modeling

Figure 7. Parameter influences on experimental results.

NeuMF, but is slightly worse than APOIR, as it pursues a trade-off between recommendation precision and diversity.

(3). With the increase of recommendation list length k , all models keep consistent improvement on recall and deterioration on precision. It makes sense as surplus recommendation often means a small increase in interested POIs but conversely a large increase in uninterested POIs.

(4). By observing the effectiveness of all models on different datasets, data sparsity negatively reduces precision and recall. We conclude that larger and sparser datasets usually have more POIs that are visited not very often and this fact may result in worse preference extraction. Note that Geo-Teaser has significant performance change probably because a small dataset cannot meet the needs of sequence learning.

NDCG Analysis

Normalized discounted cumulative gain (NDCG) is a rank-based metric, which measures ranking performance of recommendation list. Fig. 6(a) shows models' NDCG with different k values on the Gowalla dataset. From the figure, PGIM achieves the best performance in general, indicating the validity of personalized geographical influence and diversity preference modeling. In addition, NDCG increases as k increases. This is because valid POIs in recommendation lists increases with k 's rising.

Diversity Analysis

We follow diversity evaluation measure in [9], where diversity is defined as a proportion of exposed POIs across all recommendation lists among total POIs. From Fig.6(b), PGIM succeeds

in mining more long-tail terms by learning user diversity preference. On average, PGIM achieves 18.7% on diversity over the second best model, showing the validity of paying great attention to a small proportion of long-tail check-ins.

Parameter Tuning

Epoch. Fig.7(a) shows the training loss and recommendation performance in different epochs on Gowalla data. The results in other datasets show similar trends, so they are not reported here. From the overall trend, the increase of iteration epochs is accompanied by the decrease of training loss and the improvement of recommendation performance. The most valuable update occurs in the first 75 iterations. After that, the training loss and recommendation performance change slowly. About 200-th iteration is an important turning point, after which the recommendation performance degrades, although the training loss keeps decreasing. This results from too many iterations leading to over-fitting of the training data. The above findings provide empirical evidence for the effectiveness of the binary cross-entropy loss in optimizing our model.

Layers. Fig.7(b) summarizes the performance of PGIM with different number of layers in Deep Interaction Layer. MLP- i indicates that MLP with i layers, while MLP-0 denotes Concat Layer is directly connected to Prediction Layer. Factor- i represents the embedding with length i . From the results, the proposed model is improved by stacking more non-linear layers. Notably, when the embedding size is 16, MLP-3 is more suitable than MLP-4. We think the phenomenon comes from the over-fitting of the training data. Besides,

the growth of the embedding size has a positive effect on the model performance, which illustrates long embedding has the ability to express comprehensive information.

Negative Samples. We study the effect of negative sampling ratio of $\{1, 2, 4, 8, 12, 16\}$ on the model performance with a constant value 16 for the embedding size and 3 for the layer number. From Fig.7(c), the performance increases with the rise of negative sample ratio up to a certain value. Specifically, the optimal ratios are 4 for Gowalla and 8 for Yelp and Foursquare. We argue that Yelp and Foursquare are larger than Gowalla, thus they need more negative examples.

Conclusion

We propose a novel approach PGIM for POI recommendation which learns user personalized geographical preference over POIs in an NCF framework. PGIM explores user tolerance for distances among POIs and extracts user diversity preference to POIs for diversity recommendation. Comprehensive experiments have demonstrated PGIM enhances recommendation performance and is better than state-of-the-art methods. In the future, we plan to enhance PGIM in several ways. First, we note that user geographical preference typically changes over time. For instance, users may accept places farther away on weekends than on weekdays. Unfortunately, PGIM cannot capture this change in the current stage, so it is worthy to explore time-dependent geographical preference. Secondly, PGIM cannot effectively model the geographical relations among recently visited POIs when modeling user short-term preference. This is another issue that should be studied in the future.

REFERENCES

1. Carl Yang, Lanxiao Bai, Chao Zhang, Quan Yuan, and Jiawei Han. 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*, pages 1245–1254. ACM, 2017.
2. Hao Wang, Huawei Shen, Wentao Ouyang, and Xueqi Cheng. Exploiting poi-specific geographical influence for point-of-interest recommendation. In *IJCAI*, pages 3877–3883, 2018.
3. Defu Lian, Cong Zhao, Xing Xie, Guangzhong Sun, Enhong Chen, and Yong Rui. Geomf: joint geographical modeling and matrix factorization for point-of-interest recommendation. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 831–840. ACM, 2014.
4. Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web*, pages 173–182. International World Wide Web Conferences Steering Committee, 2017.
5. Shenglin Zhao, Tong Zhao, Irwin King, and Michael R Lyu. Geo-teaser: Geo-temporal sequential embedding rank for point-of-interest recommendation. In *Proceedings of the 26th international conference on world wide web companion*, pages 153–162. International World Wide Web Conferences Steering Committee, 2017.
6. Yong Liu, Wei Wei, Aixin Sun, and Chunyan Miao. Exploiting geographical neighborhood characteristics for location recommendation. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 739–748. ACM, 2014.
7. Xutao Li, Gao Cong, Xiao-Li Li, Tuan-Anh Nguyen Pham, and Shonali Krishnaswamy. Rank-geomf: A ranking based geographical factorization method for point of interest recommendation. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 433–442. ACM, 2015.
8. Azin Ashkan, Branislav Kveton, Shlomo Berkovsky, and Zheng Wen. Optimal greedy diversity for recommendation. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.
9. Peizhe Cheng, Shuaiqiang Wang, Jun Ma, Jiankai Sun, and Hui Xiong. Learning to recommend accurate and diverse items. In *Proceedings of the 26th international conference on World Wide Web*, pages 183–192. International World Wide Web Conferences Steering Committee, 2017.
10. Fan Zhou, Ruiyang Yin, Kunpeng Zhang, Goce Trajcevski, Ting Zhong, and Jin Wu. Adversarial point-of-interest recommendation. In *The World Wide Web Conference*, pages 3462–34618. ACM, 2019.

Yanan Zhang is a graduate student in the Department of Computer Science and Technology, Soochow University, Suzhou, China. Her research interests include POI recommendation system, trajectory prediction and information security. She re-

Geographical Influence Modeling

ceived the B.S. degree in software engineering from Soochow University in 2017. Contact her at 20175227082@stu.suda.edu.cn.

Guanfeng Liu is a lecturer in the Department of Computing, Macquarie University, Sydney, Australia. His research interests include graph database, social computing and trust computing. He received the Ph.D. degree in Computer Science from Macquarie University in 2013. Contact him at guanfeng.liu@mq.edu.au.

An Liu is a Professor in the Department of Computer Science and Technology, Soochow University, Suzhou, China. His research interests include spatial database, crowd-sourcing, recommendation system, data security and cloud/service computing. He received the Ph.D. degree in computer science from both City University of Hong Kong (CityU) and University of Science and Technology of China (USTC) in 2009. Contact him at anliu@suda.edu.cn.

Yifan Zhang is a graduate student in the Department of Computer Science and Technology, Soochow University, Suzhou, China. His research interests include spatial database, recommendation system and deep learning. He received the B.S. degree in software engineering from Soochow University in 2017. Contact him at 20175227050@stu.suda.edu.cn.

Zhixu Li is an associate professor in the Department of Computer Science and Technology, Soochow University, Suzhou, China. His research interests include data cleaning, information extraction and retrieval, machine learning, deep learning, knowledge graph and crowd-sourcing. He received the B.S. and M.S. degree in computer science from Renmin University, Beijing China in 2006 and 2009 respectively, with the Ph.D. degree in computer science from the University of Queensland in 2013. Contact him at zhixuli@suda.edu.cn.

Xiangliang Zhang is an Associate Professor and directs the Machine Intelligence and Knowledge Engineering (MINE) Laboratory in King Abdullahs University of Science and Technology (KAUST), Saudi Arabia. Her research interests include diverse areas of machine learning and data mining. She received the Ph.D. degree in computer science with great honors from INRIA-University Paris-Sud 11, France in 2010. Contact her at Xiangliang.Zhang@kaust.edu.sa.

Qing Li is a Chair Professor at the Department of Computing, the Hong Kong Polytechnic University.

His research interests include multi-modal data management, conceptual data modeling, social media, web services and e-learning systems. He received the B.S. from Hunan University (Changsha), with M.S. and Ph.D. degrees from the University of Southern California (Los Angeles), all in computer science. Contact him at qing-prof.li@polyu.edu.hk.