

Geo-ALM: POI Recommendation by Fusing Geographical Information and Adversarial Learning Mechanism

Wei Liu^{1,2}, Zhi-Jie Wang^{1,2}, Bin Yao³ and Jian Yin^{1,2}

¹ Guangdong Key Lab. of Big Data Anal. and Proc., Sun Yat-Sen University, Guangzhou, China

² National Engineering Laboratory for Big Data Analysis and Applications, Beijing, China

³ Department of Comp. Sci. and Eng., Shanghai Jiao Tong University, Shanghai, China
{liuw259, wangzhij5, issjyin}@mail.sysu.edu.cn, yaobin@cs.sjtu.edu.cn

Abstract

Learning user’s preference from check-in data is important for POI recommendation. Yet, a user usually has visited some POIs while most of POIs are unvisited (i.e., negative samples). To leverage these “no-behavior” POIs, a typical approach is pairwise ranking, which constructs ranking pairs for the user and POIs. Although this approach is generally effective, the negative samples in ranking pairs are obtained randomly, which may fail to leverage “critical” negative samples in the model training. On the other hand, previous studies also utilized geographical feature to improve the recommendation quality. Nevertheless, most of previous works did not exploit geographical information comprehensively, which may also affect the performance. To alleviate these issues, we propose a geographical information based adversarial learning model (Geo-ALM), which can be viewed as a fusion of geographic features and generative adversarial networks. Its core idea is to learn the discriminator and generator interactively, by exploiting two granularity of geographic features (i.e., region and POI features). Experimental results show that Geo-ALM can achieve competitive performance, compared to several state-of-the-arts.

1 Introduction

Nowadays, location-based services are widely used in our daily life [Bao *et al.*, 2015]. For instance, Yelp and Meituan can help individuals discover favourite foods, shopping malls, hotels, etc. Foursquare and Wechat can assist people in discovering fun places, friends’ footprint, etc. Almost all these applications incorporate the function of Point-of-Interest (POI) recommendation. Particularly, POI recommendation can help these applications understand individuals’ favourites more deeply, and so it has the potential to provide personalized services for individuals. Besides, it can also help merchants to solicit more potential customers. All in one, POI recommendation is an important task in LBSNs [Cheng *et al.*, 2012; Liu *et al.*, 2016a].

In the real world, it is common that a user may have visited few POIs while most of POIs are unvisited (i.e., negative

samples) [Li *et al.*, 2015; He *et al.*, 2018]. This phenomenon makes it difficult to learn user’s preference, which is important in POI recommendation [Li *et al.*, 2017; Li *et al.*, 2015; Wang *et al.*, 2017b]. To alleviate such an issue, one of representative approaches is the pairwise ranking, which constructs ranking pairs for the user and POIs (including no-behavior POIs) [Li *et al.*, 2015]. A major rule or assumption in constructing ranking pairs is that, users prefer to visited POI than unvisited POIs. Although such a method is generally effective, the negative samples in ranking pairs are obtained randomly, which may fail to leverage “critical” negative samples in the model training, affecting the overall performance [Zhang *et al.*, 2013; Wang *et al.*, 2017a]. On the other hand, previous studies have also attempted to leverage geographical information to strengthen the recommendation performance, since geographical information is closely related to individual behaviors [Ye *et al.*, 2011; Feng *et al.*, 2017; Lian *et al.*, 2014]. Nevertheless, most of previous works exploit geographical information in a single granularity, which may affect the quality of recommendation.

To alleviate the above issues, in this paper we propose a new approach, called *geographical information based adversarial learning model* (Geo-ALM). At a high level, our model can be viewed as a fusion of geographical features and generative adversarial networks (GAN). Our model contains also a discriminator and a generator, which are the core of the adversarial learning mechanism (ALM). The major benefit to use the ALM in our model is that, the *generator* allows us to obtain more critical negative samples to challenge the *discriminator*, which can make the model training more effective. On the other hand, the geographical features are used to learn the extra information to improve embedding from the sample pairs. To well exploit the geographical information, we suggest a two-granularity preference embedding strategy, which considers *POI feature* and *region feature*. To summarize, the main contributions of this paper are as follows.

- We propose a new approach, Geo-ALM, for POI recommendation. To the best of our knowledge, this is the first work to fuse geographical features with generative adversarial networks to achieve POI recommendation.
- We conduct extensive experiments based on two public datasets. The experimental results consistently demonstrate the competitiveness of our proposed method, compared against state-of-the-art methods.

The rest of the paper is organized as follows. Section 2 presents our proposed method. Section 3 discusses and analyzes the experimental results. In Section 4, we review previous works, and finally Section 5 concludes the paper.

2 The Proposed Method

2.1 Preliminaries

We have a set U of users $\{u_1, u_2, \dots, u_{|U|}\}$, and a set V of POIs $\{v_1, v_2, \dots, v_{|V|}\}$, where $|\cdot|$ denotes the cardinality of the corresponding set. For a user u_i (resp., POI v_j), we use \vec{u}_i (resp., \vec{v}_j) to denote its corresponding vector. Given users' check-in behaviors on POIs with geographical information, we aim to recommend each user with n favorite POIs that (s)he has not visited.

In POI recommendation, to leverage the missing data to learn user's preference between visited POIs and unvisited POIs, a representative method is the *pairwise ranking* [Li *et al.*, 2015], in which one usually assumes POI v_j visited by user u_i should be scored larger than unvisited POI $v_{j'}$. In addition, the ranking loss function is based on maximum likelihood estimation, and is defined as

$$L(\theta) = \prod_{u_i \in U, v_j, v_{j'} \in V} p(\hat{x}_{i,j} > \hat{x}_{i,j'} | \theta) \quad (1)$$

where $\hat{x}_{i,j}$ (resp., $\hat{x}_{i,j'}$) means the estimation of user u_i 's preference to POI v_j (resp., $v_{j'}$), θ represents embedding parameters of users and of POIs. As for user u_i 's preference to some POI v_j , it can be learned by sophisticated deep neural networks. Usually, for simplicity it is computed as $\hat{x}_{i,j} = \vec{u}_i^T \vec{v}_j$ [He *et al.*, 2017b].

For a user u_i , its visited POIs are usually called the positive samples. In contrast, the unvisited POIs are usually called the negative samples. To maximize the ranking loss function $L(\theta)$, an important step of the pairwise ranking is to construct sample pairs. For example, for a user u_i , its sample pair could be $\langle (u_i, v_j), (u_i, v_{j'}) \rangle$, where v_j is a visited POI, while $v_{j'}$ is an unvisited POI. For the pairwise ranking method, it constructs sample pairs by randomly choosing the negative samples. We argue that such an approach may fail to leverage "critical" negative samples in the model training. To understand, see an example shown in Fig. 1.

2.2 Solution Overview

At a high lever, our proposed model fuses the merits of the generative adversarial network (GAN) and geographical features seamlessly, in which geographical information is benefit to provide high-quality negative samples (as shown in Fig. 1). Our model contains two modules: discriminator and generator. These two modules, however, are essentially different from that in the conventional GAN. It can be understood from the overall framework shown in Fig. 2.

In brief, in the model pairwise ranking is regarded as *discriminator*, which tries to predict the ranking relationship between generated sample pairs, and is trained to maximize ranking samples' likelihood (recall Eq. 1). On the other hand, the generator continually generates critical negative samples,

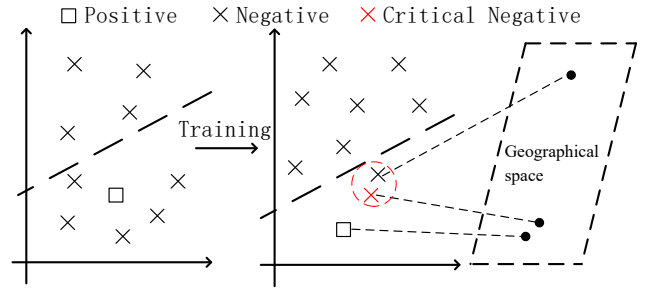


Figure 1: A running example. After several epochs of training, the model can distinguish positive sample from most of negative samples (see the right part). With random sampling, it is hard to choose the critical negative samples (see the ones in the red circle). Moreover, without geographical information GI , similar sample vectors may be hard to differentiate, e.g., the ones in the red circle. Yet, with GI , it is easier to distinguish them (see the black dot).

which are then coupled with positive samples, forming training instances. The adversarial framework interchangeably learns the parameters between two different modules.

Generally, at each training epoch it iterates over the training set to train the generator, while the parameters of the discriminator are fixed. Then, it iterates over the training set again to train the discriminator, while the parameters of the generator is fixed. When the model converges, the embeddings learned by the discriminator are used as user and POI representations for the POI recommendation. Next, we discuss our model in detail. The training process of our proposed model alternately learns the discriminator and the generator.

2.3 The Discriminator Module

The discriminator tries to rank well-matched user-POI tuples (u_i, v_j) before ill-matched ones $(u_i, v_{j'})$. In other words, its goal is to distinguish between visited POIs and unvisited POIs for user u_i as accurately as possible. Essentially, it is a maximum likelihood estimate problem such that we can make the probability, $f_\theta(v_j > v_{j'} | u_i)$, as maximal as possible. Specifically, the probability is computed as

$$f_\theta(v_j > v_{j'} | u_i) = \sigma(\hat{x}_{i,j,j'}) \quad (2)$$

$$\hat{x}_{i,j,j'} = \hat{x}_{i,j} - \hat{x}_{i,j'}$$

where $\sigma(x) = \frac{1}{1+e^{-x}}$, $\hat{x}_{i,j}$ and $\hat{x}_{i,j'}$ represents user u_i 's preference to POI v_j and $v_{j'}$, respectively.

Different from the previous models whose negative samples $v_{j'}$ are obtained by randomly sampling from the whole POI set V , our discriminator uses the negative samples $v_{j'}$ picked by the *generator* (described later). The optimal parameters for the discriminator can be obtained as follows.

$$\theta^* = \arg \max_{\theta} \sum_{u_i \in U} (\mathbb{E}_{v_j \sim p_{true}, v_{j'} \sim f_\phi} [\log f_\theta(v_j > v_{j'} | u_i)]) \quad (3)$$

where p_{true} means the user's true preference distribution on POIs; $v_j \sim p_{true}$ (resp., $v_{j'} \sim f_\phi$) denotes v_j (resp., $v_{j'}$) is sampled from positive samples, i.e., visited POIs, (resp., the generator f_ϕ); $\mathbb{E}_{(\cdot)}[\cdot]$ denotes the expectation of the log likelihood on the sampled pairs $\langle (u_i, v_j), (u_i, v_{j'}) \rangle$.

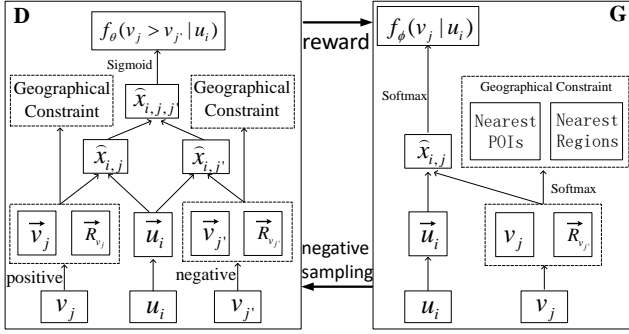


Figure 2: Framework of Geo-ALM, where **D** is Discriminator, **G** is Generator. Discriminator is for learning POI pair’s ranking for each user. Generator provides critical negative samples for discriminator.

As mentioned earlier, geographical information is important when predicting user’s preference to POI. We thus fuse geographical information into the discriminator. Specifically, we employ the following equation to compute $\hat{x}_{i,j}$:

$$\hat{x}_{i,j} = \vec{u}_i^T \vec{R}_{v_j} + \vec{u}_i^T \vec{v}_j^* \quad (4)$$

where R_{v_j} denotes a region containing POI v_j , $\vec{\cdot}$ denotes the feature vector of the corresponding item. Different from many previous works, we here consider two granularities: POI feature and region feature. The POI feature \vec{v}_j^* depicts the user’s preference to POI. The region feature \vec{R}_{v_j} can characterize the sketch of the POIs in the corresponding region, its dot product with user’s embedding can depict user’s geographical and general preference to the region. Meanwhile, the region feature can make the user’s preference to POI more discriminative, which is benefit to the subsequent steps. The intuition of the above idea is that, a user usually picks a POI by two steps: firstly, (s)he may choose a candidate region to visit; and secondly, (s)he may choose a POI in the region to visit. Particularly, in the real world it is common that users usually choose the regions and POIs near to them, or the ones familiar to them, as shown in Fig. 3.

On the other hand, one can easily understand that, when user’s activity data at a POI v_j or a region r_i is sparse, the POI feature \vec{v}_j^* (resp., region feature \vec{r}_i^*) may not be estimated accurately. To address this issue, we exploit the geographical correlation to enhance the prior knowledge, i.e., the model parameters \vec{v}_j^* and \vec{r}_i^* (a.k.a., \vec{R}_{v_j} in Eq. 4). Generally, our idea is inspired by the Skip-gram model [Mikolov *et al.*, 2013] proposed in the natural language processing community. Intuitively, if two POIs v_j and v_k are proximate in the geographical space, then their feature \vec{v}_j^* and \vec{v}_k^* should be similar to each other. This observation allows us to leverage the Skip-gram model to characterize feature similarity. Specifically, given a POI v_j , let C_{v_j} denote a set of POIs nearest to v_j . The objective is to minimize the loss function of predicting the context, based on the embedding of the POI. The loss

function is formulated as

$$\begin{aligned} L(C_{v_j}) &= - \sum_{v_k \in C_{v_j}} \log p(v_k | v_j) \\ &= - \sum_{v_k \in C_{v_j}} \log(\vec{v}_j^T \vec{v}_k^* - \log \sum_{v_{k'} \in V} \exp(\vec{v}_j^T \vec{v}_{k'}^*)) \end{aligned} \quad (5)$$

where v_k is one of the nearest neighbors to v_j .

Similarly, given a region r_i , let C_{r_i} denote a set of regions nearest to r_i . Sometimes, we also call them the surrounding regions of r_i . Let R denote the set of all regions. Then, the loss function for region is formulated as

$$\begin{aligned} L(C_{r_i}) &= - \sum_{r_j \in C_{r_i}} \log p(r_j | r_i) \\ &= - \sum_{r_j \in C_{r_i}} \log(\vec{r}_i^T \vec{r}_j^* - \log \sum_{r_{j'} \in R} \exp(\vec{r}_i^T \vec{r}_{j'}^*)) \end{aligned} \quad (6)$$

where r_j is one of nearest regions (i.e., surrounding regions) to r_i . With the above two loss functions (one is for POIs, the other is for regions), it can allow us to learn discriminator’s parameter θ easier and more accurate. Note that, the parameter ϕ can be also learned using the above functions.

2.4 The Generator Module

The generator tries to generate high-quality unvisited POI for a given user u_i . It hopes to approximate user’s true preference distribution over POIs as much as possible. In other words, the essence of the generative module is to provide the discriminator with high-quality negative samples, which helps the discriminator improve the training performance. Specifically, it can be formulated as follows.

$$f_\phi(v_j | u_i) = \frac{\exp(\hat{x}_{i,j})}{\sum_{v_{j'} \in V \setminus V_{u_i}} \exp(\hat{x}_{i,j})} \quad (7)$$

where V_{u_i} is the POI set visited by user u_i , $\hat{x}_{i,j}$ can be computed as in Eq. 4. Note that, the generator in our model has its own embeddings for users, POIs and regions. The optimal parameters for the generator can be obtained as follows.

$$\phi^* = \arg \min_{\phi} \sum_{u_i \in U} (\mathbb{E}_{v_j \sim p_{true}, v_{j'} \sim f_\phi} [\log f_\theta(v_j > v_{j'} | u_i)]) \quad (8)$$

where the notations such as $v_j \sim p_{true}$ and $\mathbb{E}_{(\cdot)}[\cdot]$ have the similar meanings in Eq. 3.

Since the output of the generator is a discrete index of the POIs, in our model we use the policy gradient based reinforcement learning [Yu *et al.*, 2017] to optimize the generator. Thus, the whole network can be essentially viewed as a policy network in the field of reinforcement learning. For a newly generated negative pair $(u_i, v_{j'})$ together with the positive pair (u_i, v_j) , the reward function calculated by the discriminator is defined as

$$\mathcal{R} = \log(1 + \exp(f_\theta(v_j > v_{j'} | u_i))) \quad (9)$$

Usually, the reward is large, when the generated negative sample and the corresponding positive sample confuse the

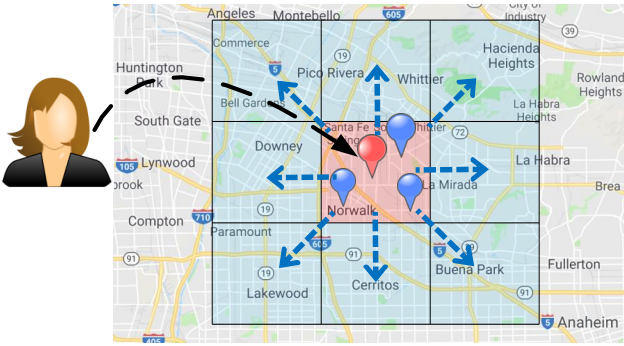


Figure 3: Preference with Geographical Information. User’s preference to POI concludes: user’s geographical interest to the region where the target POI is located and user’s general interest to the target POI; POI’s embedding is related with the nearest POIs’, region’s embedding is related with the adjacent regions’.

discriminator in Eq. 2. This situation essentially encourages to obtain a better negative samples. In our model, the following loss function tries to maximize the expected reward.

$$L(\phi) = \mathbb{E}_{f_\phi(v_j|u_i)}(\mathcal{R}) \quad (10)$$

One can see from the above equation that, in order to achieve a higher reward, the policy used by the generator punishes the negative samples by lowering down their corresponding probability, and encourages the network to distribute more probability to the samples that can perplex discriminator. In addition, when the generator updates, it uses the policy gradient below.

$$\nabla_\phi L(\phi) = \nabla_\phi \log f_\phi(v_j|u_i) \cdot \mathcal{R} \quad (11)$$

3 Empirical Study

3.1 Datasets

In the experiments, we employ two widely-used datasets, one is Foursquare (<https://pan.baidu.com/s/1hrYNwJM>), the other is Gowalla (<https://pan.baidu.com/s/1i4DgFmX>). There are 194,108 check-ins made within Singapore in Foursquare dataset. They are made by 2,321 users on 5,596 POIs from August 2010 to July 2011. The average check-ins of each user is 83.6 and the data sparsity is 1.49%. As for the Gowalla dataset, there are 456,988 check-ins produced within California and Nevada in America. They are produced by 10,162 users on 24,250 POIs from February 2009 to October 2010. The average check-ins per user is 44.9 and the data sparsity is 0.185%. The details about datasets are listed in Table 1.

3.2 Evaluation Metrics

Following [Li *et al.*, 2017; Wang *et al.*, 2018], we use two well-known metrics to evaluate the performance. One is precision@ n , the other is recall@ n , where n is the number

Dataset	#User	POI	#Check-in	#Avg.check-in	Density
Foursquare	2,321	5,596	194,108	83.6	1.49%
Gowalla	10,162	24,250	456,988	44.97	0.18%

Table 1: Details of Two Datasets

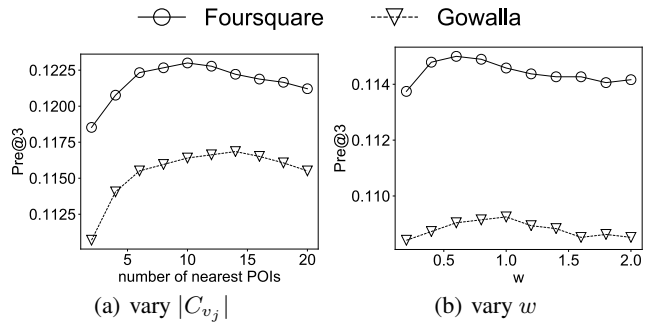


Figure 4: Effect of two newly introduced hyper parameters.

of recommended POIs. For short, we denote them by Pre@ n and Rec@ n , respectively. They are computed as

$$Pre@n = \frac{1}{|U|} \sum_{u=1}^{|U|} \frac{|L_u^n \cap T_u|}{n} \quad (12)$$

$$Rec@n = \frac{1}{|U|} \sum_{u=1}^{|U|} \frac{|L_u^n \cap T_u|}{|T_u|}.$$

where L_u^n denotes the top- n POIs recommended by the model for user u , T_u denotes the POI set that user u really visited. For each experiment, we conduct 5 independent tests and report the average value.

3.3 Hyper Parameters

In our experiments, we mainly focus on investigating the impact of two newly introduced parameters: (i) $|C_{v_j}|$, which is the number of nearest POIs to v_j , and (ii) w , which is the size of the square region. As for these two hyper parameters, in the rest of experiments we choose the values that can achieve the best performance as the default settings. Other hyper parameters follow the settings of previous works. Specifically, for both datasets, we set the dimension number to 16, the learning rate to 0.001, the regularization coefficient of $L(C_{v_j})$ and $L(C_{r_i})$ to 0.01.

Effect of $|C_{v_j}|$: Fig. 4(a) plots the results when we vary $|C_{v_j}|$ from 2 to 20. It can be seen that, for both datasets, when $|C_{v_j}|$ is too small, the performance is poor. This could be mainly because the nearest POIs’ features are too similar in this case, incurring the over-fitting. On the other hand, we observe that, when $|C_{v_j}|$ is too large, the performance turns poor. This could be mainly because, when $|C_{v_j}|$ exceeds a threshold (e.g., $|C_{v_j}| = 10$ in Foursquare, $|C_{v_j}| = 14$ in Gowalla), the set of nearest POIs introduces some unrelated data, which is contrary to the nearest property.

Effect of w : Fig. 4(b) shows the results when we vary w from 0.2 to 2.0 km. One can see that both curves first go up and then go down, with the increase of w . One possible reason is that, when the size of region is less than some threshold, the number of spatial regions will be too much, and so data sparsity may cause the parameters under-fitting. In contrast, when the size of region becomes too large, the region vector \vec{r}_j could introduce too much noise, which is not benefit for depicting spatial features accurately.

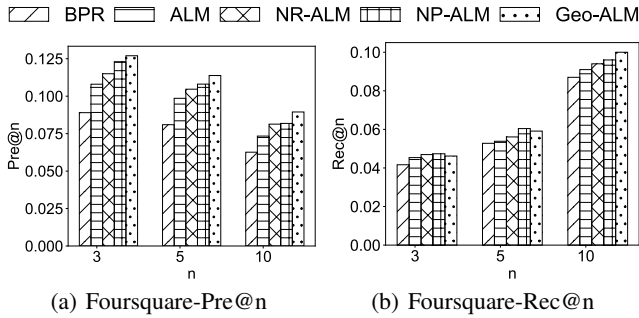


Figure 5: Performance results of various techniques.

3.4 Effectiveness of Techniques

To study the effectiveness of various techniques integrated in our model, we compare our model with (i) the basic **BPR** model [Rendle *et al.*, 2009]; and (ii) **ALM**, which doesn’t consider geographical factor and is the adversarial learning part of Geo-ALM; (iii) **NP-ALM**, which only considers the Nearest POIs of candidate POI as context, without using the region context; and (iv) **NR-ALM**, which only considers the region context when depicting user’s preference to POI, and the Nearest Regions of target region are used as the region context, without using the POI context.

Fig. 5 shows the compared results. It can be seen that our adversarial learning model with dynamic negative sampling can achieve a great improvement, compared against the BPR. It reflects that a high-quality negative sampling is beneficial for general preference learning. Besides, NP-ALM and NR-ALM perform better than ALM, it proves geographical information is significant in POI recommendation. Moreover, NP-ALM is better than NR-ALM, it shows instance granularity is more powerful than region granularity. At last, Geo-ALM is better than all the other models, it demonstrates our model with adversarial learning and two granularity of geographical information is effective.

3.5 Comparing with State-Of-The-Arts

We compare our model with the following state-of-the-art models: (i) **BPR**, a classic model that optimizes the ordering relationship of user’s preferences for the visited POI and the unvisited POI [Rendle *et al.*, 2009]; (ii) **IRenMF**, which incorporates neighboring characteristics into weighted matrix factorization [Liu *et al.*, 2014]; (iii) **Rank-GeoFM**, which is a ranking-based MF model that learns users’ preference rankings for POIs, and includes the geographical influence of neighboring POIs [Li *et al.*, 2015]. It is one of the strongest state-of-the-art top- n recommendation model. (iv) **PACE**, which is a general and principled semi-supervised learning framework. It alleviates data scarcity via smoothing among neighboring POIs, and it treats the geographical context by regularizing POI feature, based on context graphs [Yang *et al.*, 2017]; and (v) **GeoIE**, which exploits geographical information by asymmetry and high variation geographical influence, and constructs a cross-entropy object function to learn model parameters [Wang *et al.*, 2018].

Fig. 6 shows the compared results. It can be seen that, our proposed method, Geo-ALM, outperforms other models

for all these cases, demonstrating the competitiveness of our method. Particularly, when it is compared with the strong competitor, it still obtains about 7% relative improvements on the Foursquare dataset, and more than 5.8% improvements on the Gowalla dataset, respectively. These results essentially demonstrate that dynamic negative sampling with adversarial learning and multiple view geographical features is feasible and effective to improve the recommendation accuracy.

On the other hand, when we look a bit deeper into the results, we find that BPR is poorer than other models. This essentially further validates that geographical information is important in POI recommendation, which is ignored in BPR. In addition, Geo-ALM and GeoIE perform better than PACE, Rank-GeoFM and other baselines, it demonstrates that a fine geographical information representation can improve POI recommendation greatly. Last but not least, Geo-ALM outperforms best in all the models, even better than GeoIE and PACE, it verifies, from another perspective, that the adversarial learning mechanism is helpful to improve POI recommendation performance. Another reason for the improvement could be the high-quality negative samples, which are filtered by the multiple granularity of geographical information.

4 Related Work

This section reviews previous works most related to ours. For clarity, we classify them into three categories.

Negative Sampling: Negative sampling technique has been widely used in natural language processing (NLP) community [Stergiou *et al.*, 2017; Soleimani and Matwin, 2018; Guo *et al.*, 2018; Zhang and Zweigenbaum, 2018]. Recently, it has been also applied to recommendation community, benefiting to the alleviation of data sparsity. The main idea of the negative sampling technique in recommendation community is to leverage not only implicit positive samples but also implicit negative samples when learning user’s preference. For example, [Li *et al.*, 2015] proposed a method, called Rank-GeoFM, which learns the factorization by ranking the POIs correctly. Particularly, both behavior and no-behavior POIs are used to learn the ranking and thus the data sparsity problem can be alleviated. Moreover, they also incorporated geographical information to learn the factorization. Yet, their method chooses negative samples in ranking pairs *randomly*. In contrast, our approach employs the adversarial learning mechanism in which the discriminator and generator are learned interactively, and so our model is iteratively updated, guaranteeing the overall performance. Essentially, the early work [Zhang *et al.*, 2013] have proposed to dynamically choose negative training samples from the ranked list and iteratively update the model. Nevertheless, their work is obviously different from ours in several aspects: (i) they focused on item recommendation, instead of POI recommendation. (ii) we leverage the adversarial learning mechanism and geographical features, which are not covered in their work. Besides, other methods (e.g., SSL [Yang *et al.*, 2017]) that are obviously different from ours, can be also used to leverage negative samples (i.e., unlabelled data).

Adversarial Learning Mechanism: Recently, a widely used adversarial learning model is the generative adversarial net-

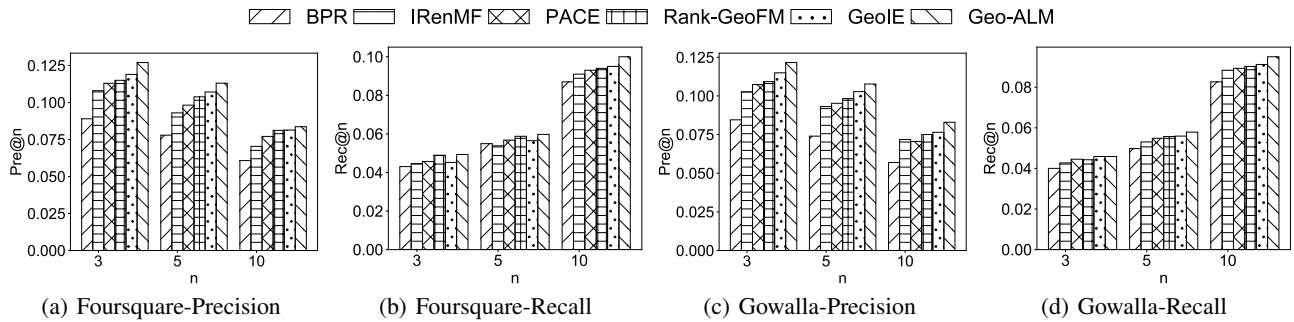


Figure 6: Performance results of comparing with several state-of-the-art models.

work (GAN) [Goodfellow *et al.*, 2014]. Existing studies have demonstrated the effectiveness of GAN in various tasks such as sequence generation [Yu *et al.*, 2017], image captioning [Chen *et al.*, 2017]. In the recommendation community, [Wang *et al.*, 2017a] proposed a method called IRGAN that iteratively optimize a generative retrieval component and a discriminative retrieval component. Moreover, [Zhao *et al.*, 2018] presented a solution that combines GAN with the matrix factorization (MF) and the recurrent neural network (RNN). Recently, [He *et al.*, 2018] proposed an optimization framework, namely Adversarial Personalized Ranking (APR), that enhances the Bayesian personalized ranking (BPR) approach by performing adversarial training. Although we used also the adversarial learning mechanism in our solution, our work is obviously different from the above works in several points: (i) Our work focuses on POI recommendation, instead of sequence generation, image captioning, and/or item recommendation. (ii) Our method fuses the geographical information features into the adversarial learning mechanism, which is not covered in the above works.

Geographical Features: To improve the recommendation accuracy, previous studies have attempted to capture various information [He *et al.*, 2017a], including social and geographical influence [Li *et al.*, 2016], temporal effect [Liu *et al.*, 2016b], and sequential check-ins’ influence [Feng *et al.*, 2017], etc. Among them, the most widely used could be geographical information. For example, [Ye *et al.*, 2011] modelled the geographical influence by power law distribution. [Cheng *et al.*, 2012] suggested to fuse matrix factorization with geographical and social influence. [Lian *et al.*, 2014] proposed to fuse weighted matrix factorization with geographical influence. [Liu *et al.*, 2013] proposed a geographical probabilistic factor analysis framework which strategically takes various factors into consideration. [Li *et al.*, 2015] proposed a ranking based geographical factorization method. [Feng *et al.*, 2017] presented geographical latent representation for predicting future visitors. [Li *et al.*, 2017] proposed user intrinsic and extrinsic interest to POIs, distinguished by user’s geographical activity area. [Wang *et al.*, 2018] Exploited POI-Specific geographical influence. Among these works, the ones highly closest to ours could be [Li *et al.*, 2015], since (i) both of us consider POIs’ geographical feature with neighboring POIs, and (ii) both of us leverage the negative samples. Nevertheless, our work is different from theirs in several aspects at least: (i) Their method chooses

negative samples randomly [Zhao *et al.*, 2016; Chang *et al.*, 2018], while our work introduces the adversarial learning mechanism, which interactively and dynamically chooses the negative training samples. (ii) We model not only POI’s geographical feature but also the region’s feature, which is not adopted in their works.

5 Conclusion

In this paper we have presented a new method, called Geo-ALM, for POI recommendation. Our method absorbs the merits of adversarial learning mechanism and geographical information. The central idea of our method is to learn the discriminator and generator interactively, by exploiting two granularity of geographic features. Extensive experiments based on two public datasets have demonstrated the feasibility and competitiveness of our method. In the future, we would like to incorporate more context information to further improve our method.

Acknowledgments

This work is supported by the National Key R&D Program of China (2018YFB1004400), Key R&D Program of Guangdong Province (2018B010107005, 2019B010120001), National Natural Science Foundation of China (U61811264, U1711262, U1401256, U1501252, U1611264, U1711261).

References

- [Bao *et al.*, 2015] Jie Bao, Yu Zheng, David Wilkie, and Mohamed Mokbel. Recommendations in location-based social networks: a survey. *Geoinformatica*, 19(3):525–565, 2015.
- [Chang *et al.*, 2018] Buru Chang, Yonggyu Park, Donghyeon Park, Seongsoon Kim, and Jaewoo Kang. Content-aware hierarchical point-of-interest embedding model for successive poi recommendation. In *IJCAI*, pages 3301–3307, 2018.
- [Chen *et al.*, 2017] Tseng-Hung Chen, Yuan-Hong Liao, Ching-Yao Chuang, Wan Ting Hsu, Jianlong Fu, and Min Sun. Show, adapt and tell: Adversarial training of cross-domain image captioner. In *ICCV*, pages 521–530, 2017.
- [Cheng *et al.*, 2012] Chen Cheng, Haiqin Yang, Irwin King, and Michael R. Lyu. Fused matrix factorization with geographical and social influence in location-based social network. In *AAAI*, pages 17–23, 12 2012.

- [Feng *et al.*, 2017] Shanshan Feng, Gao Cong, Bo An, and Yeow Meng Chee. Poi2vec: Geographical latent representation for predicting future visitors. In *AAAI*, pages 102–108, 2017.
- [Goodfellow *et al.*, 2014] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Couville, and Yoshua Bengio. Generative adversarial nets. In *NIPS*, pages 2672–2680, 2014.
- [Guo *et al.*, 2018] Guibing Guo, Shichang Ouyang, Fajie Yuan, and Xingwei Wang. Approximating word ranking and negative sampling for word embedding. In *IJCAI*, pages 4092–4098, 2018.
- [He *et al.*, 2017a] Jing He, Xin Li, and Lejian Liao. Category-aware next point-of-interest recommendation via listwise bayesian personalized ranking. In *IJCAI*, pages 1837–1843, 2017.
- [He *et al.*, 2017b] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia hu, and Tat-Seng Chua. Neural collaborative filtering. In *WWW*, pages 173–182, 2017.
- [He *et al.*, 2018] Xiangnan He, Zhankui He, Xiaoyu Du, and Tat-Seng Chua. Adversarial personalized ranking for recommendation. In *SIGIR*, pages 355–364, 2018.
- [Li *et al.*, 2015] Xutao Li, Gao Cong, Xiao-Li Li, Tuan-Anh Nguyen Pham, and Shonali Krishnaswamy. Rankgeofm: A ranking based geographical factorization method for point of interest recommendation. In *SIGIR*, pages 433–442, 2015.
- [Li *et al.*, 2016] Huayu Li, Yong Ge, Richang Hong, and Hengshu Zhu. Point-of-interest recommendations: Learning potential check-ins from friends. In *KDD*, pages 975–984, 2016.
- [Li *et al.*, 2017] Huayu Li, Yong Ge, Defu Lian, and Hao Liu. Learning user’s intrinsic and extrinsic interests for point-of-interest recommendation: A unified approach. In *IJCAI*, pages 19–25, 2017.
- [Lian *et al.*, 2014] 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 *KDD*, pages 831–840, 2014.
- [Liu *et al.*, 2013] Bin Liu, Yanjie Fu, Zijun Yao, and Hui Xiong. Learning geographical preferences for point-of-interest recommendation. In *KDD*, pages 1043–1051, 2013.
- [Liu *et al.*, 2014] Yong Liu, Wei Wei, Aixin Sun, and Chuanyan Miao. Exploiting geographical neighborhood characteristics for location recommendation. In *CIKM*, pages 739–748, 2014.
- [Liu *et al.*, 2016a] Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. Predicting the next location: A recurrent model with spatial and temporal context. In *AAAI*, pages 194–200, 2016.
- [Liu *et al.*, 2016b] Yanchi Liu, Chuanren Liu, Bin Liu, Meng Qu, and Hui Xiong. Unified point-of-interest recommendation with temporal interval assessment. In *KDD*, pages 1015–1024, 2016.
- [Mikolov *et al.*, 2013] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. In *ICLR*, 2013.
- [Rendle *et al.*, 2009] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. Bpr: Bayesian personalized ranking from implicit feedback. In *UAI*, pages 452–461, 2009.
- [Soleimani and Matwin, 2018] Behrouz Haji Soleimani and Stan Matwin. Spectral word embedding with negative sampling. In *AAAI*, pages 5481–5487, 2018.
- [Stergiou *et al.*, 2017] Stergios Stergiou, Zygimantas Straznickas, Rolina Wu, and Kostas Tsioutsoulklis. Distributed negative sampling for word embeddings. In *AAAI*, pages 2569–2575, 2017.
- [Wang *et al.*, 2017a] Jun Wang, Lantao Yu, Weinan Zhang, Yu Gong, Yinghui Xu, Benyou Wang, Peng Zhang, and Dell Zhang. Irgan: A minimax game for unifying generative and discriminative information retrieval models. In *SIGIR*, pages 515–524, 2017.
- [Wang *et al.*, 2017b] Peifeng Wang, Shuangyin Li, and Rong Pan. Incorporating gan for negative sampling in knowledge representation learning. In *AAAI*, 2017.
- [Wang *et al.*, 2018] 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.
- [Yang *et al.*, 2017] 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 *KDD*, pages 1245–1254, 2017.
- [Ye *et al.*, 2011] Mao Ye, Peifeng Yin, Wang-Chien Lee, and Dik-Lun Lee. Exploiting geographical influence for collaborative point-of-interest recommendation. In *SIGIR*, pages 325–334, 2011.
- [Yu *et al.*, 2017] Lantao Yu, Weinan Zhang, Jun Wang, and Yong Yu. Seqgan: Sequence generative adversarial nets with policy gradient. In *AAAI*, pages 2852–2858, 2017.
- [Zhang and Zweigenbaum, 2018] Zheng Zhang and Pierre Zweigenbaum. GNEG: graph-based negative sampling for word2vec. In *ACL*, pages 566–571, 2018.
- [Zhang *et al.*, 2013] Weinan Zhang, Tianqi Chen, Jun Wang, and Yong yu. Optimizing top-n collaborative filtering via dynamic negative item sampling. In *SIGIR*, pages 785–788, 2013.
- [Zhao *et al.*, 2016] Kaiqi Zhao, Gao Cong, and Aixin Sunu. Annotating points of interest with geo-tagged tweets. In *CIKM*, pages 417–426, 2016.
- [Zhao *et al.*, 2018] Wei Zhao, Benyou Wang, Jianbo Ye, Yongqiang Gao, Min Yang, and Xiaojun Chen. PLASTIC: prioritize long and short-term information in top-n recommendation using adversarial training. In *IJCAI*, pages 3676–3682, 2018.