IJCNN 2019. International Joint Conference on Neural Networks. Budapest, Hungary. 14-19 July 2019

ADPR: An Attention-based Deep Learning Point-of-Interest Recommendation Framework

Junjie Yin[1], Yun Li[1], Zheng Liu[1], Jian Xu[2], Bin Xia[1*], Qianmu Li[2]
[1]Jiangsu Key Laboratory of Big Data Security and Intelligent Processing,
Nanjing University of Posts and Telecommunications, Nanjing, China
[2]School of Computer Science and Engineering,
Nanjing University of Science and Technology, Nanjing, China
{1017041117,liyun,zliu,bxia}@njupt.edu.cn, dolphin.xu@njust.edu.cn, liqianmu@126.com

***Abstract*—With the development of location-based social networks (LBSNs), Point-of-Interest (POI) recommendation has attracted lots of attention. Most of the existing studies focus on recommending POIs to users based on their recent check-ins. However, the recent check-ins may contain some daily check-ins that users are not really interested in. If a model treats the recent check-ins equally, it is non-trivial to capture the actual preference of users. To address the issue of mining the actual preferences of users in the POI recommendation, we propose an attention-based deep learning POI recommendation framework (ADPR), which consists of a latent representation method and an attention-based deep convolutional neural network. To learn the embedding of users and POIs, we propose a latent representation method, which incorporates the geographical influence and the categories of POIs to capture the relationships between POIs better. Further, we propose an attention-based deep convolutional neural network, which employs the attention mechanism to filter the important information in the recent check-ins, to recommend POIs to users based on the latent representations of users and the recent check-ins. We conduct experiments on a real-world LBSN dataset to evaluate our framework, and the experimental results show the effectiveness of our framework.**

***Index Terms*—POI recommendation, attention-based, latent representation, CNN**

I. INTRODUCTION

With the rapid development of mobile devices, GPS, and Web 2.0 technologies, LBSNs have attracted a lot of attention. The significant characteristic of LBSNs is that users can share their visiting experience via check-in records in corresponding POIs. As social interaction platforms, LBSNs can make full use of the rich information (e.g., social relationships and check-in history) to capture user preferences and recommend interesting POIs for these users. The task of recommending new POIs that users are interested in, which is known as POI recommendation, has been extensively researched [1]–[6].

The movements of users are influenced by both their recent visits and personal preferences [3]. Many works about POI recommendation focus on modeling users' personal preferences and recommend POIs based on users' recent check-ins [7]–[9]. However, the recent check-ins of a user usually contain some daily check-ins, such as the check-in at a bus station, which are not locations that the user is really interested in. The task of POI recommendations is to discover POIs which users are interested in rather than to predict users' daily movement. For example, a user, who likes amusement parks, prefers a recommendation list including new amusement parks even if she has not visited a amusement park for a long time, to a recommendation list only including some locations she daily visited. If a model treats all check-ins in users' recent check-ins equally, it is non-trivial for the model to capture the actual preference of users. On the one hand, a POI recommendation model should capture users' actual preference; On the other hand, a user's recent check-ins, which captures the context, also has great influence on the user's future check-ins [9]. Therefore, how to distinguish check-ins that users are really interested in from users' daily check-ins has became a problem in the POI recommendation. In this paper, we solve this problem in two steps: mining users' personal preferences from their historical check-ins and using their personal preferences to filter the POIs which are appropriate in the recent check-ins.

The historical check-in data can reflect the personal preference of the user. However, due to the high sparsity of check-in data, it is challenging to model users' personal preferences from their historical check-in sequences. The technique of word2vec [10], [11], which is an effective way of word representation in natural language processing (NLP), can capture the sequential semantic relationships between words. Recently, the word2vec algorithm has been utilized to model users' sequential check-ins [12] and has also been improved for POI recommendation [13]. In [13], the hierarchical softmax technique [14] is exploited to incorporate geographical influence in generating latent representations for POIs. However, except for geographical influence, many other factors have great impact on the check-in behavior of users, such as the categories of POIs. Therefore, in this paper, we aim to incorporates the geographical influence and the categories of POIs to generate embedding vectors of POIs for improving the performance of POI recommendations in the following stage.

Besides the historical check-in data, the recent check-in records which reflect the behavioral preference is also important for capturing the user preferences. Therefore, to model a user's recent check-ins better, we need to pay attention to these check-in records. Recent researches show that the CNN, which is the state-of-the-art deep learning methodology, has advantages in extracting high level features. So we exploit the CNN to extract features from users' recent check-ins. Then, we

978-1-7281-2009-6/$31.00 ©2019 IEEE paper N-20072.pdf

exploit attention mechanisms, which is widely used in many fields [15]–[17], to capture a user's actual preference from features extracted from the recent check-ins. In our model, we use the attention mechanism to determine which features are more important to precisely describe the user preferences.

In this paper, to address the issue of mining users' actual preferences from users' recent check-ins in the POI recommendation, we propose an attention-based deep learning framework (ADPR), which consists of a latent representation method and a deep convolutional neural network employing the attention mechanism. To generate representations for users' personal preferences and POIs, we propose a latent representation method incorporating the geographical influence and the categories of POIs. Then, we use the generated latent representation of users and POIs as the input to the attention-based deep convolutional neural network for the task of POIs recommendations. The major contributions of our work are summarized as follows:

- We propose a latent representation model for users and POIs, which incorporates geographical influence and the categories of POIs.
- We propose a deep convolutional neural network employing the attention mechanism for POI recommendation to address the issue of mining users' actual preferences from users' recent check-ins.
- We conduct the extensive experiments to compare our framework with the state-of-the-art approaches, and the experimental results show the effectiveness of our framework.

II. Related Work

The POI recommendation has attracted a lot of attention recently. Most of the previous methods are based on traditional methods such as Collaborative Filtering [2], [18] and Factorization models [19], [20], and capture users' preference by exploiting the geographical information [13], which is usually considered as the basic characteristic of a POI. However, some other crucial factors, such as the categories of POIs, should also be considered in the POI recommendation.

The technique of word2vec [10] are developed for the NLP task. There are two traditional solutions [11]: the hierarchical softmax and the negative sampling. Although word2vec techniques are proposed for natural language processing, they can also be adapted to other areas as latent representation methods. In particular, check-in data have sequential semantic relationships like the text data and follows the same distribution as does the word frequency distribution. So word2vec techniques can be adapted to process check-in data. Besides, the latent representation methods improved by POI-specific factors have achieved good results in the POI recommendation [9], [13]. [12] utilizes the word2vec technique to model the check-in sequence and [13] incorporate the geographical influence by proposing a novel latent representation model POI2Vec based on the hierarchical softmax technique. However, the movements of users are influenced by many factors. To capture the relationships between POIs better, we exploit negative sampling and improve it for POI recommendation by incorporating the geographical influence and the categories of POIs.

TABLE I
Key Notations in the paper

Notation	Description
$U = \{u_1, u_2, \cdots, u_m\}$	Set of users
$L = \{l_i, l_2, \cdots, l_n\}$	Set of POIs
$H(u_k)$	Check-in history of user u_k
X	Set of check-in records
$r_{(u_k, l_i, t)}$	The record of u_k visiting l_i at time t
lon_i,lat_i	Longitude and latitude of POI l_i
$Cat(l_i)$	Category of POI l_i
$d(l_i, l_j)$	Linear distance between POI l_i and POI l_j
D	Dimension of latent representation
Γ_N^-	Set of negative samples

Recently, deep learning models have rapid development and are widely used in various fields, including POI recommendation [21], [22]. Most of existing deep learning POI recommendation models are based on the recurrent neural network (RNN). However, RNN has some drawbacks in modeling the sequence [23]. Compared to RNNs, the convolutional neural network, which is originally developed for computer vision [24] and is widely used in many other fields [25]–[28] do not depend on the computations of the previous time step and therefore allow parallelization over every element in a sequence. So we use CNNs to extract features from users' recent check-ins. Besides, attention mechanisms, which are widely used in many fields [15]–[17], can be utilized in POI recommendation to filter the important information [29]. So we exploit the attention mechanism in our deep deep convolutional neural network to filter the important features from the recent check-ins.

III. Proposed Attention-based Deep Learning Framework

In the task of POI recommendations, given a set of users U and a set of POIs L, we aim to capture the preferences of users and provide lists of exciting POIs for each user. In this paper, we propose an attention-based deep learning framework which consists of a latent representation method and a deep convolutional neural network employing the attention mechanism. In the following sections, first, we introduce a latent representation method incorporating the geographical influence and the categories of POIs. Second, we describe the proposed deep convolutional neural network employing the attention mechanism in details. The key notations mentioned in this paper are summarized in Table I.

A. Latent Representation for users and POIs

1) Latent Representation Method: For a given check-in $r_{(u_k, l_i, t)}$, we define the context of $r_{(u_k, l_i, t)}$ as $C(r_{(u_k, l_i, t)}) =$

$\{r_{(u_k,l_j,t')}, 0 \leq |t-t'| \leq \tau\}$, where τ is the time window. We assume that the user preferences and the influence of contexts are independent. Therefore, given the user u_k and the context $C(r_{(u_k,l_i,t)})$, the probability that u_k visits l_i at time t is

$$\begin{aligned} & p(r_{(u_k,l_i,t)}|u_k, C(r_{(u_k,l_i,t)})) \\ & = p(r_{(u_k,l_i,t)}|u_k) \times p(r_{(u_k,l_i,t)}|C(r_{(u_k,l_i,t)})), \end{aligned} \tag{1}$$

where $p(r_{(u_k,l_i,t)}|u_k)$ and $p(r_{(u_k,l_i,t)}|C(r_{(u_k,l_i,t)}))$ describes the influence of the user preferences and the contexts, respectively. The influence of the user preferences is estimated by

$$p(r_{(u_k,l_i,t)}|u_k) = \sigma(\vec{u_k} \cdot \vec{l_i}) = \frac{1}{1+e^{-\vec{u_k}\cdot\vec{l_i}}}, \tag{2}$$

where $\vec{u_k} \in \mathbb{R}^D$ is the latent representation of u_k and $\vec{l_i} \in \mathbb{R}^D$ is the latent representation of l_i. The influence of the contexts is estimated by

$$\begin{aligned} p(r_{(u_k,l_i,t)}|C(r_{(u_k,l_i,t)})) &= \sigma(\vec{l_i} \cdot \vec{Z}(C(r_{(u_k,l_i,t)}))) \\ &= \frac{1}{1+e^{-\vec{l_i}\cdot\vec{Z}(C(r_{(u_k,l_i,t)}))}}, \end{aligned} \tag{3}$$

where $\vec{Z}(C(l_i^{u_k}))$ is the contextual factor:

$$\vec{Z}(C(r_{(u_k,l_i,t)})) = \frac{1}{|C(r_{(u_k,l_i,t)})|} \sum_{r_{(u_k,l_j,t')} \in C(r_{(u_k,l_i,t)})} \vec{l_j}, \tag{4}$$

where $|C(r_{(u_k,l_i,t)})|$ is the number of check-ins in $C(r_{(u_k,l_i,t)})$.

Our target is to maximize $p(r_{(u_k,l_i,t)}|u_k, C(r_{(u_k,l_i,t)}))$ for all observed check-ins in the training set, meanwhile, minimize $p(r_{(u_k,l_j,t)}|u_k, C(r_{(u_k,l_i,t)}))$ for all $l_j \neq l_i$. However, the complexity of computing all probability is $O(|L| \times |X|)$, which is quiet expensive when $|L|$ is large. So we utilize the negative sampling technology [11] to reduce the computational complexity to $O(|X|)$. The loss function for $\forall r_{(u_k,l_i,t)} \in X$ is transformed into:

$$\begin{aligned} & loss(r_{(u_k,l_i,t)}) = -log\, p(r_{(u_k,l_i,t)}|u_k, C(r_{(u_k,l_i,t)})) \\ & - \sum_{l_{neg} \in \Gamma_N^-} log(1 - p(r_{(u_k,l_{neg},t)}|u_k, C(r_{(u_k,l_i,t)}))), \end{aligned} \tag{5}$$

where Γ_N^- is a set of N negative samples for the context $C(r_{(u_k,l_i,t)})$. Therefore, the overall objective function is:

$$E = \sum_{r \in X} loss(r) \tag{6}$$

2) Negative Sampling Technology: In a traditional task of word2vec [11], a word is often considered as a positive sample due to the frequent occurrence with the target context. To efficiently train the word2vec model, the words, which are frequent in the corpus but never appear in the target context, are selected as the negative samples. However, it is not suitable to produce negative samples based on the popularity of POIs for the recommendation task. The geographical distance and the category of POI are more important to affect the check-in behavior of users. Therefore, we select the negative samples (i.e., POIs) simultaneously considering the influence of geographical distance and the categories. For example, the check-in record $r_{(u_k,l_i,t)}$ is the positive sample under the context $C(r_{(u_k,l_i,t)})$. The unobserved check-in record $r_{(u_k,l_j,t)}$ is more likely to be the negative sample for $C(r_{(u_k,l_i,t)})$ if $Cat(l_j) = Cat(l_i)$ and l_j is geographically close to l_i. The probability that a POI l_{neg} is selected as the negative sample of $C(r_{(u_k,l_i,t)})$ is estimated by

$$p_N(l_{neg}|C(r_{(u_k,l_i,t)})) = \frac{f(d(l_i, l_{neg}))}{\sum_{\{l_j|Cat(l_j)=Cat(l_i)\}} f(d(l_i, l_j))}, \tag{7}$$

where $f(d(l_i, l_j))$ is defined as follow:

$$f(d(l_i, l_j)) = \frac{1}{d(l_i, l_j)} \tag{8}$$

3) Parameter Learning: We adopt the stochastic gradient descent (SGD) to optimize the objective function Eqn. 6. In each iteration, the set of parameters $\Theta = \{\{\vec{u_k}|1 \leq k \leq m\}, \{\vec{l_i}|1 \leq i \leq n\}\}$ is updated using:

$$\Theta = \Theta + \gamma \frac{\partial}{\partial \Theta} loss(r_{(u_k,l_i,t)}), \tag{9}$$

where γ is the learning rate. The optimization algorithm for the proposed latent representation method is summarized in Algorithm 1.

Algorithm 1 Latent Representation Method for users and POIs

Input: U, L, X, τ, D, γ, N
Output: $\Theta = \{\{\vec{u_k}|1 \leq k \leq m\}, \{\vec{l_i}|1 \leq i \leq n\}\}$

Initialize Θ with normal distribution $\mathcal{N}(0,1)$
repeat
  for all check-in $r_{(u_k,l_i,t)} \in X$ **do**
    Clear Γ_N^-
    for all $l_{neg} \in \{l_j|Cat(l_j) = Cat(l_i)\}$ **do**
      compute $p_N(l_{neg}|C(r_{(u_k,l_i,t)}))$ with Eqn. 7
    end for
    $t = 0$
    while $t < N$ **do**
      Sample a $l_{neg} \in \{l_j|Cat(l_j) = Cat(l_i)\}$
      if $l_{neg} \notin \Gamma_N^-$ **then**
        Add l_{neg} to Γ_N^-
        $t \leftarrow t + 1$
      else
        Continue
      end if
    end while
    Update Θ with Eqn. 9
  end for
until convergence

B. Attention-based Deep Convolutional Neural Network

After generating the latent representation for users and POIs, we propose a deep convolutional neural network employing the attention mechanism to recommend POIs to users. Figure 1 illustrates the architecture of the network.

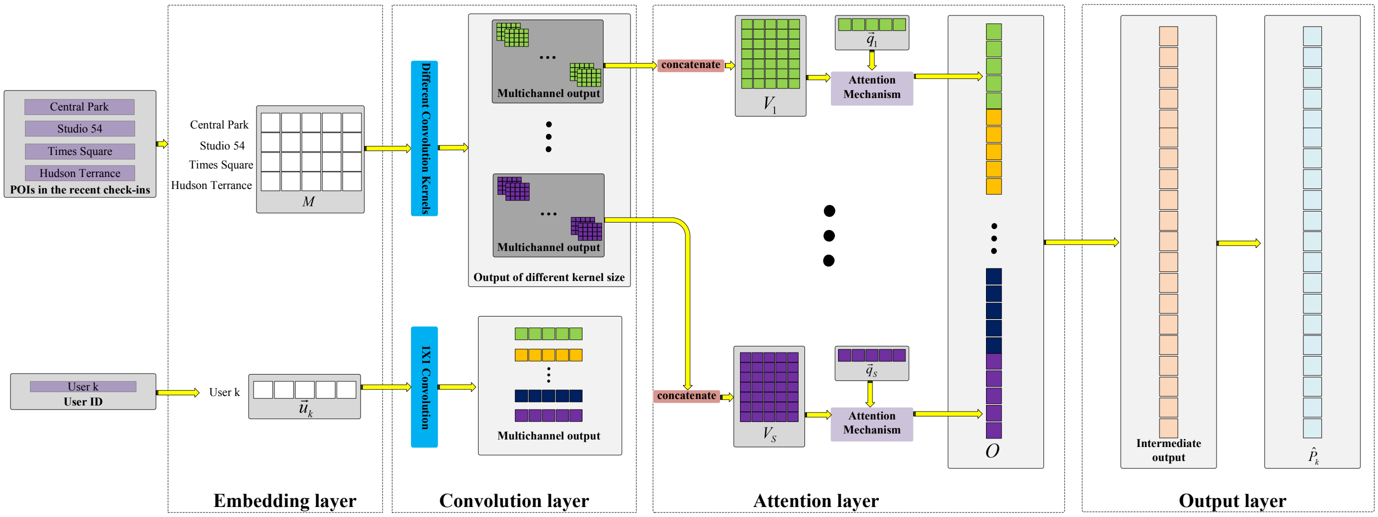


Fig. 1. The architecture of the attention-based deep convolutional neural network

1) Embedding Layer: The embedding layer consists of the embedding vectors of users $\{\vec{u_k}|1 \leq k \leq m\}$ and POIs $\{\vec{l_i}|1 \leq i \leq n\}$ that are initialized using the pre-trained latent representation. The input of the embedding layer is the user id of u_k and an ordered POI id sequence $[l_1, l_2, \cdots, l_h]$ based on the POIs in the recent check-ins of u_k. The output of the embedding layer consists of a matrix $M \in \mathbb{R}^{h \times D}$ and the user embedding $\vec{u_k}$. The matrix M is the concatenation of the ordered POI id sequence $[l_1, l_2, \cdots, l_h]$.

$$M = \begin{bmatrix} \vec{l_1} \\ \vec{l_2} \\ \vdots \\ \vec{l_h} \end{bmatrix}. \tag{10}$$

2) Convolution Layer: The convolution layer is used to extract features from M and $\vec{u_k}$. To extract features from M, we use S convolution kernels that have different sizes. Each convolution kernel has Ω output channels. For example, if the size of the convolution kernel is $s_i \times 1$, we transform M into a matrix $\hat{M} \in \mathbb{R}^{(h+2s_i-2) \times D}$ by padding $s_i - 1$ zero vectors on both the top and bottom side of M. Then, the output of each channel of the convolution kernel is $v_i^\omega \in \mathbb{R}^{(h+s_i-1) \times D}$:

$$v_i^\omega = ReLU(W_i^\omega * \hat{M}_{(j:(j+s_i-1),:)} + b_i^\omega), \tag{11}$$

where $*$ is a convolution operator, b_i^ω is a bias for W_i^ω and $ReLU$ is the activation function. Then, we concatenate all v_i^ω to generate the final output of this convolution kernel:

$$V_i = \begin{bmatrix} v_i^1 \\ v_i^2 \\ \vdots \\ v_i^\Omega \end{bmatrix}, \tag{12}$$

where $V_i \in \mathbb{R}^{\Omega(h+s_i-1) \times D}$. To extract features from $\vec{u_k}$, we use a 1×1 convolution kernel that has S output channels. The output of this convolution kernel consists of S vectors $\vec{q_1}, \vec{q_2}, \cdots, \vec{q_S}$. So the final output of the convolution layer consists of $\{V_i|1 \leq i \leq S\}$ and $\{\vec{q_i}|1 \leq i \leq S\}$.

3) Attention Layer: The attention layer is used to filter important information from features extracted by the convolution. In the attention layer, the output of the convolution layer $\{V_i|1 \leq i \leq S\}$ and $\{\vec{q_i}|1 \leq i \leq S\}$ is transformed into S feature map pairs (V_i, q_i), and the attention mechanism is employed to each feature map pair $(V_i, \vec{q_i})$. V_i is considered as the feature of the recent check-ins of u_k. The attention mechanism allocates different attentions (i.e., weights) to V_i, which can enhance the impact of the check-ins that u_k actually interested in. $\vec{q_i}$ that extracted from $\vec{u_k}$ describes the user preference. The rows of V_i that are more similar to $\vec{q_i}$ are more important to mine the user preference. Therefore, the weight of the row in V_i is positive correlated with the similarity between this row and $\vec{q_i}$, which can be measured by the dot product of this row and $\vec{q_i}$. The weight of the row $V_i(n,:)$ is computed by:

$$a_n = \frac{exp(\frac{V_i(n,:) \cdot \vec{q_i}}{\sqrt{D}})}{\sum_{j=1}^{\Omega(h+s_i-1)} exp(\frac{V_i(j,:) \cdot \vec{q_i}}{\sqrt{D}}))}, \tag{13}$$

where $\frac{1}{\sqrt{D}}$ is the scaling factor [17]. Therefore, the final output of the attention layer $O \in \mathbb{R}^{SD}$ is generated by:

$$O = Concat(\vec{o}_1, \vec{o}_2, \cdots, \vec{o}_S), \tag{14}$$

where $\vec{o}_i$ is the weighted sum of the rows of V_i :

$$\vec{o}_i = \sum_{n=1}^{\Omega(h+s_i-1)} a_n \cdot V_i(j,:). \tag{15}$$

4) Output Layer: The output layer, which consists of two fully-connected layers, uses the output of the attention layer O as the input and outputs a vector $\hat{P}_{u_k} \in \mathbb{R}^{|L|}$, which represents the probability of a user's check-in of a POI. We use ReLU on the first fully-connected layer layer as the activate fuction.

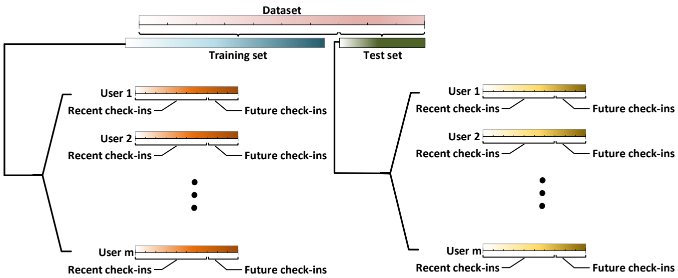


Fig. 2. The structure of the training set and the test set

To prevent overfitting, we employ dropout on the first fully-connected layer. On the second fully-connected layer, we use sigmoid as the activate function.

5) Parameter Learning: The loss of each iteration is the cross entropy between P_{u_k} and $\hat{P}_{u_k}$:

$$loss(P_{u_k}, \hat{P}_{u_k}) = \sum_{i=1}^{|L|}[\hat{P}(i)logP(i)+(1-\hat{P}(i))log(1-P(i))], \tag{16}$$

where $\hat{P}_{u_k}$ is the output of the attention-based deep convolutional neural network for u_k and P_{u_k} is the real posterior probability of the future check-ins of u_k that is estimated by:

$$P_{u_k}(i) = \begin{cases} 1, \text{ if } u_k \text{ visits } l_i \text{ in the future} \\ 0, \text{ else} \end{cases} \tag{17}$$

The objective of the deep convolutional neural network employing the attention mechanism is as follow:

$$\theta = \min_{\theta} \sum_{u_k \in U} loss(P_{u_k}, \hat{P}_{u_k}), \tag{18}$$

where θ are the parameters of the network and updated using SGD.

IV. EXPERIMENTS

A. Setup

In this paper, we perform our evaluation on a real-world LBSN dataset [29]. The dataset contains 419,509 check-in records published by 49,823 users among 18,899 POIs from August 2012 to July 2013 in Manhattan via Foursquare API. The dataset includes 16 categories (e.g., Shop, Service, and Food) of POIs. To reduce the effects of inactive users and POIs, we remove the users who have fewer than 5 check-ins, and the POIs which have been visited by fewer than 5 users. After the data preprocessing, there remain 17,564 users, 7,734 POIs and 339,634 check-ins.

We construct the training set using the first 70% chronological check-ins and the test set using the last 30% check-ins. In the training set, we use the early 70% of a user's check-ins as her recent check-ins and the last 30% as the future check-ins. In the test set, we also use the first 70% check-ins of each user as her recent check-ins and the rest as her future check-ins. Figure 2 illustrates the framework of the training set and the test set. First, to learn the latent representation of users and POIs, we set $\tau = 12\,hours$, the dimension of latent representation $D = 200$, the numbers of negative samples for each actual check-in record $N = 5$, and the learning rate as $\gamma = 0.001$. Second, to train the deep convolutional neural network employing the attention mechanism, we set the learning rate of embedding layer as 0.001 to tune the embedding, and the learning rate of other parameters as 0.025. In the convolution layer, we use four kernel sizes are 1×1, 2×1, 4×1, and 8×1, and set the number of channels $\Omega = 16$. In the output layer, we set the dropout rate of 0.5.

B. Evaluation Metrics

We adopt three widely-used metrics for evaluation [30], $Precision@K$, $Recall@K$, and $F1@K$, where K is the number of POIs in the recommendation list of each user.

$$Precision@K = \frac{1}{|U|}\sum_{u \in U}\frac{|R_u^K \bigcap T_u|}{|R_u^K|}, \tag{19}$$

$$Recall@K = \frac{1}{|U|}\sum_{u \in U}\frac{|R_u^K \bigcap T_u|}{|T_u|}, \tag{20}$$

$$F1@K = \frac{2 \cdot Precision@K \cdot Recall@K}{Precision@K + Recall@K}, \tag{21}$$

where R_u^K is the set of top K POIs in the recommendation list of user u, and T_u is user u's ground truth set of POIs.

C. Baselines

We compare the proposed method with the following POI recommendation methods:

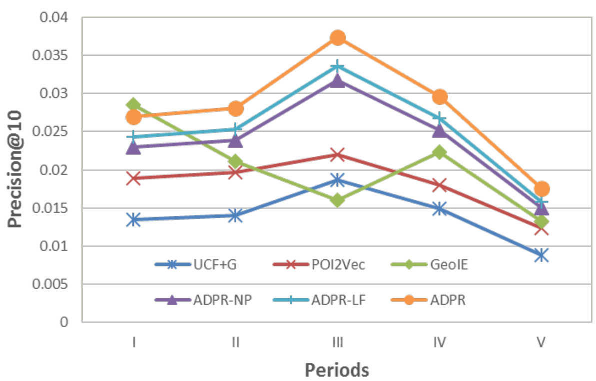
- UCF+G [2]: It is an improved user-based collaborative filtering method capturing check-in probability with distance by using a power-law function.
- POI2Vec [13]: It is a latent representation model that joint captures POI sequential and user preference to recommend POIs. It incorporates the geographical influence of POIs in learning the latent representations of POIs.
- GeoIE [9]:It is a POI recommendation model exploiting POI-specific geographical influence. It incorporates the geographical influence between two POIs using three factors: the geo-influence of POI, the geo-susceptibility of POI, and their physical distance.
- ADPR-NP: ADPR-NP is a variant of the proposed framework ADPR, which generate negative samples based on the popularity of POIs.
- ADPR-LF: ADPR-LF is a variant of the proposed framework ADPR, which fix the parameters of the embedding layer in the deep convolutional neural network employing the attention mechanism.

D. Results and Discussion

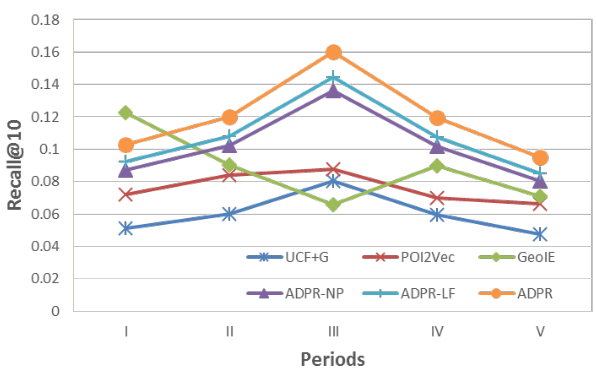
1) The Performance of TopK Recommendation: To make a fair comparison, we set the parameters of baselines same as our model. Table II shows the performance ($Precision@K, Recall@K, F1@K$) of ADPR and baselines on the LBSN dataset, where $K = 5, 10, 15, 20$. ADPR, ADPR-NP and ADPR-LF all outperform UCF-G, POI2Vec

TABLE II
THE PERFORMANCE OF TOPK RECOMMENDATION

Model	Precision				Recall				F1			
	Top-5	Top-10	Top-15	Top-20	Top-5	Top-10	Top-15	Top-20	Top-5	Top-10	Top-15	Top-20
UCF+G	0.1000	0.0715	0.0577	0.0509	0.0529	0.0669	0.0756	0.0845	0.0692	0.0692	0.0654	0.0635
POI2Vec	0.1400	0.1000	0.0808	0.0713	0.0741	0.0936	0.1059	0.1183	0.0969	0.0967	0.0917	0.0890
GeoIE	0.1500	0.1072	0.0866	0.0764	0.0793	0.1003	0.1134	0.1268	0.1038	0.1036	0.0982	0.0954
ADPR-NP	0.1700	0.1215	0.0981	0.0866	0.0899	0.1137	0.1286	0.1437	0.1176	0.1175	0.1113	0.1081
ADPR-LF	0.1800	0.1286	0.1039	0.0917	0.0950	0.1204	0.1361	0.1521	0.1244	0.1244	0.1178	0.1144
ADPR	0.2000	0.1429	0.1155	0.1019	0.1058	0.1338	0.1513	0.1690	0.1384	0.1382	0.1310	0.1271



(a) Precision in different periods



(b) Recall in different periods

Fig. 3. Performance in different periods

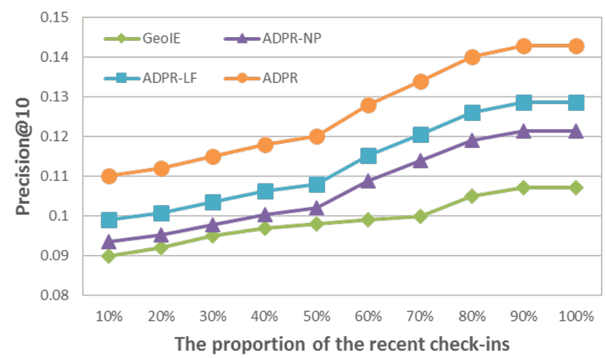


Fig. 4. Performance with different amounts of recent check-ins

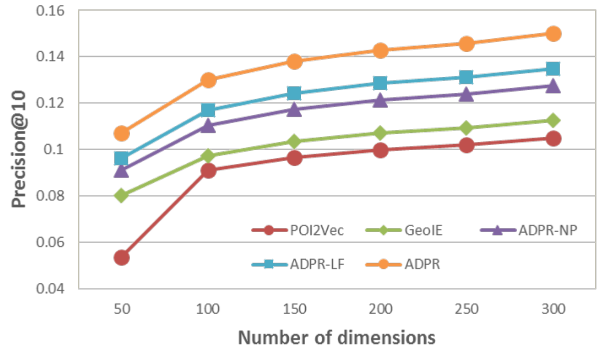
and GeoIE, which shows the effectiveness of our framework. Note that ADPR achieves better results than ADPR-NP, which shows that it is reasonable to incorporate the influence of geographical distance and the categories. In addition, ADPR performances better than ADPR-LF, which shows that tuning the latent representations improves the model.

2) The Performance in Different Periods: In this experiment, we divide a user's future check-ins (the last 30% check-ins of each user in the test set) into 5 parts in chronological order. Then, we evaluate the performance ($Precision@10, Recall@10$) of ADPR and baselines in these 5 parts. The result is shown in Figure 3. ADPR outperforms baselines in the most periods. Only the performance of GeoIE is better than ADPR in the first interval and has a distinctly different trend with other models. The reason of this phenomenon is that GeoIE is more affected by geographical location than other models. The scope of activities of users in the short term is limited. So the performance of GeoIE is very good at the first period and then decreases rapidly. The preformance of ADPR is high in all first three periods, which shows that ADPR better balance the impact of multiple factors. The performance of all models is bad at the last period because the influence of recent check-ins is timeliness.

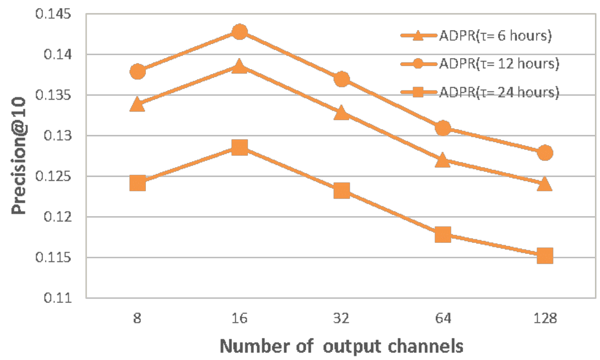
3) The Performance with Different Amounts of Recent Check-ins: In this experiment, we evaluate the performance of ADPR and baselines with different amounts of recent check-ins in the test set (the last 30% check-ins of each user in the test set). We only evaluate GeoIE, ADPR-NP, ADPR-LF and ADPR because UCF-G and POI2Vec are not influenced by the recent check-ins. The result is shown in Figure 4. The abscissa indicates the proportion of recent check-in data being used. For example, 10% means we only use the last 10% recent check-ins to recommend and 100% means all recent check-ins are used. Note that ADPR achieves better results than baselines, with different amounts of historical data. We observe that the performance ($Precision@10$) of all methods rises when we increase the amount of recent check-ins, which indicates the more data can improve the performance of models.

4) Effects of embedding dimensions: This experiment is to evaluate the effect of the number of embedding dimensions D. We also evaluate the baselines with this parameter. As shown in Figure 5(a), the performance ($Precision@10$) of all methods improves with the increase of D. We set $D = 200$ in our experiments by considering the trade-off between effectiveness and efficiency.

5) Effects of the number of output channels and the time window: This experiment is to evaluate the effects of the number of channels Ω in the convolution layer and the time window τ. We set $\Omega = 8, 16, 32, 64, 128$ and $\tau = 6\,hours, 12\,hours, 24\,hours$. Figure 5(b) shows the performance ($Precision@10$) of ADPR with different Ω and different τ. The performance first improves as we enlarge Ω, and then it decreases slightly. The more channels means that we achieve the more features from users' recent check-ins, which can reflect the user preference better. However, it brings in worse performance to generate too many features because of the limited data. Note that ADPR with $\tau = 12\,hours$ achieves better results than ADPR with $\tau = 6\,hours$ and ADPR with $\tau = 24\,hours$. On the hand, if τ is small, the context of check-ins is too little to capture the relationships between POIs. On the other hand, if τ is large, the context of check-ins is too much and contains much noise. Therefore, we need to choose an appropriate τ. The best performance is acquired at $\Omega = 16$ and $\tau = 12\,hours$, so we use the same parameters in other experiments.



(a) Effects of dimension D



(b) Effects of Ω and τ

Fig. 5. Effects of Parameters

V. Conclusions and Future Work

In this paper, we consider the problem to mine the actual user preference from the recent check-ins. To address the problem, we propose an attention-based deep learning framework (ADPR) which consists of a latent representation method and an attention-based deep convolutional neural network. Experiments on a real-world LBSN dataset show that our framework outperforms the state-of-art methods on the task of the POI recommendation.

There are several interesting and promising directions in which this work could be extended. First, the movements of users are influenced by many factors and we can consider other information such as the temporal influence for the POI recommendation. Second, because it is of great value to identify the potential visitors for locations, we would like to investigate more advanced framework to generate the bi-directional recommendation for both users and POIs. Third, we would like to improve our framework to address the cold start problem better.

VI. Acknowledgment

This work was supported by the National Natural Science Foundation of China under Grant (No.61802205, 61872186, 61772284), the Natural Science Research Project of Jiangsu Province under Grant (18KJB520037,17KJB520024), the Jiangsu Key Laboratory of Big Data Security & Intelligent Processing of NJUPT under Grant (BDSIP1803), the research funds of NJUPT under Grant (NY218116), Jiangsu Provincial Natural Science Foundation of China under Grant (BK20171447).

References

[1] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: user movement in location-based social networks," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2011, pp. 1082–1090.

[2] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee, "Exploiting geographical influence for collaborative point-of-interest recommendation," in *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. ACM, 2011, pp. 325–334.

[3] C. Cheng, H. Yang, M. R. Lyu, and I. King, "Where you like to go next: Successive point-of-interest recommendation." in *IJCAI*, vol. 13, 2013, pp. 2605–2611.

[4] B. Xia, Z. Ni, T. Li, Q. Li, and Q. Zhou, "Vrer: context-based venue recommendation using embedded space ranking svm in location-based social network," *Expert Systems with Applications*, vol. 83, pp. 18–29, 2017.

[5] B. Xia, T. Li, Q.-F. Zhou, Q. Li, and H. Zhang, "An effective classification-based framework for predicting cloud capacity demand in cloud services," *IEEE Transactions on Services Computing*, 2018.

[6] B. Xia, T. Li, Q. Li, and H. Zhang, "Noise-tolerance matrix completion for location recommendation," *Data Mining and Knowledge Discovery*, vol. 32, no. 1, pp. 1–24, 2018.

[7] S. Feng, X. Li, Y. Zeng, G. Cong, Y. M. Chee, and Q. Yuan, "Personalized ranking metric embedding for next new poi recommendation." in *IJCAI*, vol. 15, 2015, pp. 2069–2075.

[8] J. He, X. Li, L. Liao, D. Song, and W. K. Cheung, "Inferring a personalized next point-of-interest recommendation model with latent behavior patterns." in *AAAI*, 2016, pp. 137–143.

[9] H. Wang, H. Shen, W. Ouyang, and X. Cheng, "Exploiting poi-specific geographical influence for point-of-interest recommendation." in *IJCAI*, 2018, pp. 3877–3883.

[10] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in neural information processing systems*, 2013, pp. 3111–3119.

[11] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[12] X. Liu, Y. Liu, and X. Li, "Exploring the context of locations for personalized location recommendations." in *IJCAI*, 2016, pp. 1188–1194.

[13] S. Feng, G. Cong, B. An, and Y. M. Chee, "Poi2vec: Geographical latent representation for predicting future visitors." in *AAAI*, 2017, pp. 102–108.

[14] F. Morin and Y. Bengio, "Hierarchical probabilistic neural network language model." in *Aistats*, vol. 5. Citeseer, 2005, pp. 246–252.

[15] R. Paulin, "Human-robot motion: an attention-based approach," Ph.D. dissertation, Université Grenoble Alpes, 2018.

[16] H. Han, J. Liu, and G. Liu, "Attention-based memory network for text sentiment classification," *IEEE Access*, vol. 6, pp. 68 302–68 310, 2018.

[17] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, 2017, pp. 5998–6008.

[18] Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. M. Thalmann, "Time-aware point-of-interest recommendation," in *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2013, pp. 363–372.

[19] C. Cheng, H. Yang, I. King, and M. R. Lyu, "Fused matrix factorization with geographical and social influence in location-based social networks." in *Aaai*, vol. 12, 2012, pp. 17–23.

[20] 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 *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 2015, pp. 433–442.

[21] R. Ding and Z. Chen, "Recnet: a deep neural network for personalized poi recommendation in location-based social networks," *International Journal of Geographical Information Science*, pp. 1–18, 2018.

[22] S. Xing, F. Liu, X. Zhao, and T. Li, "Points-of-interest recommendation based on convolution matrix factorization," *Applied Intelligence*, vol. 48, no. 8, pp. 2458–2469, 2018.

[23] J. Gehring, M. Auli, D. Grangier, D. Yarats, and Y. N. Dauphin, "Convolutional sequence to sequence learning," *arXiv preprint arXiv:1705.03122*, 2017.

[24] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[25] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural language processing (almost) from scratch," *Journal of Machine Learning Research*, vol. 12, no. Aug, pp. 2493–2537, 2011.

[26] P. Blunsom, E. Grefenstette, and N. Kalchbrenner, "A convolutional neural network for modelling sentences," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, 2014.

[27] J. Gao, P. Pantel, M. Gamon, X. He, and L. Deng, "Modeling interestingness with deep neural networks," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 2–13.

[28] Y. Shen, X. He, J. Gao, L. Deng, and G. Mesnil, "A latent semantic model with convolutional-pooling structure for information retrieval," in *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*. ACM, 2014, pp. 101–110.

[29] B. Xia, Y. Li, Q. Li, and T. Li, "Attention-based recurrent neural network for location recommendation," in *Intelligent Systems and Knowledge Engineering (ISKE), 2017 12th International Conference on*. IEEE, 2017, pp. 1–6.

[30] Y. Liu, T.-A. N. Pham, G. Cong, and Q. Yuan, "An experimental evaluation of point-of-interest recommendation in location-based social networks," *Proceedings of the VLDB Endowment*, vol. 10, no. 10, pp. 1010–1021, 2017.