Article

A Point-of-Interest Recommendation Method Exploiting Sequential, Category and Geographical Influence

Xican Wang ¹ , Yanheng Liu ^{2,3}, Xu Zhou ^{3,4,*} , Xueying Wang ² and Zhaoqi Leng ¹

¹ College of Software, Jilin University, Changchun 130012, China; xicang20@mails.jlu.edu.cn (X.W.); lengzq20@mails.jlu.edu.cn (Z.L.)

² College of Computer Science and Technology, Jilin University, Changchun 130012, China; yhliu@jlu.edu.cn (Y.L.); xueyingw20@mails.jlu.edu.cn (X.W.)

³ Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun 130012, China

⁴ Center for Computer Fundamental Education, Jilin University, Changchun 130012, China

* Correspondence: zhox16@jlu.edu.cn

Abstract: Point of interest (POI) recommendation as an important service in location-based social networks has developed rapidly, which can help users find more interesting unknown locations and facilitate service providers to provide users with more accurate notifications or advertisements. Some existing work has addressed the data sparsity problem of collaborative filtering by incorporating contextual information into the model. However, they ignore the sequence relationship contained in the user's historical check-in records, which makes it difficult to accurately model the user's preference and affects the final recommendation results. To acquire users' preference for a location more accurately, this paper proposes a new POI recommendation framework exploiting sequential, category, and geographical influence. Firstly, we obtain the latent vector of POI and the latent vector of the user's preference for POI from the user's check-in sequence based on the word embedding model. Next, a virtual common access sequence for users is constructed according to the user's check-ins, a new similarity computation method is present combining category differentiation and POI latent vector. Then, we apply it to the collaborative filtering framework to get the user's behavioral preference probability of POI. In addition, the kernel density estimation method is employed to get the user's geographical preference probability of POI by considering the geographical influence. Finally, the POI recommendation list is obtained by the weighted fusion of the two users' preference probability to improve the performance of the POI recommendation. Experimental results on two datasets indicate that the proposed method has better performance in terms of three evaluation metrics than the other five POI recommendation methods.

Keywords: POI recommendation; sequence similarity computation; word embedding; location-based social network

Citation: Wang, X.; Liu, Y.; Zhou, X.; Wang, X.; Leng, Z. A Point-of-Interest Recommendation Method Exploiting Sequential, Category and Geographical Influence. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 80. <https://doi.org/10.3390/ijgi11020080>

Academic Editors: Wolfgang Kainz and Norbert Bartelme

Received: 2 December 2021

Accepted: 19 January 2022

Published: 20 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the rapid development of mobile internet technology and the growth of built-in GPS smart devices recently, location-based social networks (LBSN) such as Foursquare and Gowalla are growing quickly and becoming popular. People are accustomed to sharing their comments on the places they visit in LBSN. The check-in behavior of users is constrained and influenced by multiple contextual information (time factor, category and geographical factor, etc.) in LBSN. Point of interest (POI) recommendation is one of the most important applications, which is to predict a list of unvisited POIs users are interested in. With the rise of the online tourism industry, the recommendation algorithm can provide personalized services and help travelers find interesting places—such as recommending restaurants that meet their tastes when users travel. It not only reduces decision-making time for users

while traveling but also provides more customized services and brings economic value to businesses [1].

Collaborative filtering (CF) is commonly adopted for POI recommendation by leveraging user's check-ins due to its simple model. Furthermore, model-based CF methods such as matrix factorization (MF) and its variants' probability matrix factorization (PMF) is often used for POI recommendation since it can obtain implicit feedback data of a user-POI matrix [2,3]. It takes the influence of hidden factors on users' preferences and features of POI into account and uses the inner product of low-dimensional latent vectors of user and POI to predict the user's access probability to POI. However, CF-based techniques often suffer from the data sparsity problem.

To improve recommendation performance, much contextual information (time factor, category, geographical factor) is incorporated into the CF and MF method, which alleviates the matrix sparsity problem. For example, Yuan developed a CF-based POI recommendation model combined with time information, which shows that the user check-in behavior has periodicity and is influenced by the time factor [4]. However, they omit considering the user's continuous trajectory for obtaining user similarity. Xu et al. [5] proposed a POI recommendation method based on the continuous trajectories of users, and they constructed virtual common access sequence for users to find users with the same preferences. Since the categories of POI convey important information about users' interests and habits [6], He et al. proposed a new POI recommendation method by predicting firstly users' preferences for categories and getting ranked POI candidates according to category preferences [7]. It achieves a good prediction effect on sparse check-in data due to the number of categories of POI being much smaller than that of POI. Lyu utilized a category hierarchical tree to model user's preferences [8]. The geographical influence of locations on users' movement behavior is studied and widely used for POI recommendation. It is found that the user check-in places show an obvious clustering phenomenon [9]. Some studies express the distribution of historical check-ins of all users in the same model, such as power-law distribution [10] or polycentric Gaussian distribution, and explore the geographical influence of location. However, due to the different interests and lifestyles of users, it is necessary to model user spatial distribution characteristics. The present research is not able to get better performance with considering a single factor or two factors, and more influential features should be integrated to improve the effectiveness of POI recommendations.

Inspired by the above views, to take more factors into account for improving the performance of POI recommendations, this paper proposes a new personalized POI recommendation method by exploring the sequential, category, and geographical influence. Firstly, we propose an improved virtual common access sequence construction method and find users with similar mobility habits to the target users. Meanwhile, a new user similarity calculation method is proposed according to the check-in sequence and category differentiation computation. In particular, the preference vector representation of the user for POIs is learned based on the continuous bag of words (CBOW) model. Next, the user's preference vector representation is combined with a CF method to obtain the behavior preference probability of target users for POI. In addition, a kernel density estimation (KDE) model is employed to get the user's geographical preference probability of POIs. Finally, a recommendation list of top-k POIs is generated according to the user behavioral preference probability and geographical preference probability computed as a weighted sum and sorted in descending order.

The main contributions of our work are summarized as follows:

1. In this paper, we propose a new POI recommendation framework for exploiting sequential, category, and geographical influence (named SCGM) and get the user's preference probability computed from the linear combination of CF and KDE.
2. A new user similarity computation method is proposed based on the constructed virtual common access sequence of users and category differentiation of POI.
3. Specifically, we introduce CBOW to capture the contextual influence of POI in the sequence, and obtain the users' preference for POI.

4. A large number of experiments performed on two LBSN datasets show that our proposed method performs significantly better than other methods in terms of precision, recall, and F1 score.

The rest of the paper is organized as follows: we review the POI recommendation methods in Section 2. In Section 3, some preliminary works are described. Section 4 details the proposed POI recommendation approach. Section 5 gives the experimental results and the corresponding parameter analysis. Finally, Section 6 presents the conclusions.

2. Related Work

POI recommendations have received much more attention from academia and have become an important application direction in location-based social network services. To better predict user preferences, people try to incorporate multiple contextual factors that influence user check-in behavior into the recommendation model.

2.1. Temporal and Sequential Information

User behavior and interests change over time in LBSN. Researchers try to improve the performance of the POI recommendation system by combining the time information [11,12]. Some researchers mainly divide users' check-in time into multiple periods and learn users' preferences for location in each period. Since people have different visiting behavior on weekdays and non-weekdays, Hosseini et al. [13] proposed a recommendation framework based on users' weekly time preferences. They explore the preference characteristics of people's check-in records on weekdays and weekends. Gao et al. [14] proposed a spatiotemporal sensing social collaborative ranking model considering the sparsity of data for recommendation. It is based on the tensor factorization framework, and each dimension corresponds to the potential characteristics of the user, check-in time, check-in location, and social information. Gan et al. [15] integrated memory-based preference into the POI recommendation scheme based on Ebbinghaus' memory theory. They create a memory-based attenuation model for handling user POI preference and calculate the POI preference similarity between the users through their check-in records in LBSN. Zhang et al. [16] introduced the concept of POI stickiness and proposed a CF framework combining memory preference and POI stickiness.

2.2. Category Information

POI usually has one or more category attributes, which also significantly influence the user's access behavior. The category information of the POI plays an important role in modeling the specific preferences of users. In reality, user's POI decision is influenced by their category preference, and users with the same category preferences tend to exhibit the same check-in habits. For example, travelers tend to check-in at places like hotels, while students check-in at places like the library. Most studies are based on the relationship between users and POIs, but few studies focus on the relationship between users and categories. Bao et al. [17] calculated user similarity by computing user category deviation in the user-based CF recommendation method. Liu et al. [18] clustered POI by category and constructed user-category matrix to replace user-POI matrix according to the historical check-in data of users; then, matrix decomposition technology is employed to explore the top-k category that users wanted to visit next. Zhou et al. [19] described users' check-in behavior of different types of POI at different times as time curves in their study. They proposed a dynamic time warping-based location recommendation algorithm and the best curve coupling-based location recommendation algorithm. Rahmani et al. [6] proposed a POI recommendation model based on category awareness, which captured the category information features of POIs. Li et al. [20] applied a virtual trajectory to the CF recommendation framework and calculated the similarity of trajectories by decomposing the user-category matrix and constructing a Voronoi diagram. They only highlight the category feature vector in category information but ignore the category differentiation, which can more intuitively reflect the similarity between users.

2.3. Geographic Information

Since the check-in behavior of users is essentially a physical interaction between users and POIs, the user prefers to access the POI close to them. Therefore, geographical factors affect users' check-in behavior, and mining geographical influence can improve the performance of POI recommendations. Wang et al. [21] found that the influence between POIs are asymmetric and different, and they analyzed geographical sensitivity, and physical distance to simulate the Poisson geographical influence between two POIs. They improve recommendation performance by integrating specific POI geographic effects into the recommendation model. Rahmani et al. [22] incorporated geographic models into the logical matrix decomposition method and proposed a POI recommendation model LGLMF. It improves the recommendation effect by considering the geographical information from the user aspect and POI aspect, respectively. Li et al. [23] proposed a hierarchical geographic decomposition model rank-GEOMF, which learns the embedding of users and POIs based on the user check-in frequency. Liu et al. [24] modeled user mobility patterns by capturing the geographical impact on users' check-in behavior and proposed a geographical probability factorization framework that includes user preferences, geographical influence, and user behavior patterns. Yin et al. [25] studied the influence of geographical regions on user preference. They construct user preference model based on the collective preference of the public in the target region and the personal preference of users in the adjacent region. It is worthwhile to model the personalized geographical impact of location as a separate distance distribution for each user when recommending POIs for users.

3. Preliminaries

The key definitions of POI recommendation, and a brief introduction of the word embedding is described in this section.

3.1. Definitions

Definition 1 (Check-in record). Let $U = \{u_1, u_2, \dots, u_n\}$ be the set of users in LBSN, $V = \{v_1, v_2, \dots, v_n\}$ be the set of POIs and $C = \{c_1, c_2, \dots, c_n\}$ be the set of POI categories. V is represented by $\{lat, lon\}$, where lat is the latitude of a POI, and lon is the longitude. All check-in records sorted by check-in time are defined as $CH = \{ch_1, ch_2, \dots, ch_n\}$, where ch_i represents all check-in records of u_i , and each check-in record is denoted as $ch = (u, v, t)$, which indicates that POI v was visited by user u at the time t .

Definition 2 (Check-in sequence). A check-in sequence of u_i can be defined as $seq_i^r = \{v_1, v_2, \dots, v_n\}$, where r represents the r th period in a day, and POIs are sorted by check-in time. Moreover, all check-in sequences of u_i are denoted as $S_{u_i} = (seq_i^1, seq_i^2, \dots, seq_i^n)$.

Definition 3 (Contextual POI). In the check-in sequence seq_i^r , the target v_i and its contextual $\{v_{i-w} : v_{i+w}\}$ are POIs with different check-in times, where w represents the contextual window size.

Definition 4 (POI recommendation). POI recommendation is to recommend a list of unvisited POIs to a user by mining the user's check-in records. Given all check-in records CH , a ranked list of POIs $top-k = \{v_1, v_2, \dots, v_k\}$ is returned for a user.

3.2. Word Embedding

Recently, word embedding technology has been extended to trajectory data mining and sequential recommendation systems [26]. It depends on the CBOW or Skip-Grams to set up neural learning and the learned word vectors that effectively capture significant relationships in contextual information within the training dataset. In the POI recommendation system, the target user's check-in sequence is treated as a sentence, and POI is the word in the sentence, which is represented as a one-hot vector. In this paper, the CBOW model is

used to train the POI vector. It predicts the central POI according to the surrounding POIs, which has the advantages of high efficiency.

During estimating the probability of generating a central POI, each POI has two different representations. One is the center POI vector, and the other is the contextual POI vector. It is necessary to take the average of the contextual POIs' vectors and make a softmax operation on the inner product of the vectors to generate the probability of the central POI. As shown in Figure 1, when given window size is two, the contextual POIs are $v_{i-2}, v_{i-1}, v_{i+1}, v_{i+2}$. Then, the probability of generating v_i by softmax operation on the inner product of vector is as shown in Equation (1):

$$P(v_i | v_{i-2}, v_{i-1}, v_{i+1}, v_{i+2}) = \frac{\exp[p_i^T (q_{i-2} + q_{i-1} + q_{i+1} + q_{i+2})]}{\sum_{j=1}^D \exp[p_j^T (q_{i-2} + q_{i-1} + q_{i+1} + q_{i+2})]} \quad (1)$$

where $D = \{0, 1, \dots, |D| - 1\}$ is the index set, $|D|$ is the length of the dictionary. In addition, p_i and q_i represent POI vectors when the POI indexed by i is used as the center POI and contextual POI, respectively. By performing the optimization of the objective function, the POI's one-hot vector is embedded in the low-dimensional vector representation and its formula is:

$$O = -\log \sum_{i=1}^D P(v_i | v_{i-2}, v_{i-1}, v_{i+1}, v_{i+2}) \quad (2)$$

Figure 1. CBOW model.

4. Proposed Method

4.1. Framework of SCGM

In this paper, we propose a POI recommendation method exploiting sequential, category, and geographical influence (named as SCGM). SCGM is a hybrid POI recommendation model of CF and kernel density estimation. It is described in Algorithm 1.

We first construct a virtual common access sequence based on Algorithm 2 and calculate the similarity between user trajectories according to Algorithm 3. Next, SCGM gets the user behavioral preference probability according to Algorithm 4, and it is based on CF with trajectory similarity calculation. In addition, considering that geographical proximity significantly affects users' check-in behavior, SCGM gets geographical preference probability based on KDE. Finally, the user behavioral preference probability and geographical preference probability are combined in a linear model and the recommendation list of top-k POIs is generated. Formally, α is a geographical weighting coefficient, the probability that user u_i visits POI v_j can be expressed as:

$$\text{score}(u_i, v_j) = \text{pscore}(u_i, v_j) + \alpha \cdot p^{\text{geo}}(v_j | ch_i) \quad (3)$$

Algorithm 1 SCGM method.**Input:** The target user u_i , all check-in records CH and parameter α **Output:** Top-k list of POIs

```

1:  $ch_i \leftarrow CH$ 
2: for each  $u_j \in U$  do
3:    $ch_j \leftarrow CH$ 
4:   Construct a virtual common access sequence  $seq_{ij}$  according to Algorithm 2
5:   Obtain  $sim(u_i, u_j)$  according to Algorithm 3
6: end for
7: for each  $v_j \in V$  do
8:   Obtain user behavioral preference probability  $pscore(u_i, v_j)$  according to Algorithm 4
9:   Obtain user geographical preference probability  $p^{geo}(v_j|ch_i)$  based on KDE
10:  Calculate user preference probability  $score(u_i, v_j)$  according to Equation (3)
11: end for
12: Select the top-k POIs that descending sort by user preference probability.
13: return Top-k list.

```

Algorithm 2 The method of constructing a virtual common access sequence.**Input:** check-in records ch_i, ch_j **Output:** $S_{u_i}, S_{u_j}, seq_{ij}$

```

1: divide a day into four time periods  $T = \{T_1, T_2, T_3, T_4\}$ 
2:  $S_{u_i} \leftarrow \{\phi\}, S_{u_j} \leftarrow \{\phi\}$ 
3: for each  $T_r$  in  $T$  do
4:   generate check-in sequences  $seq_i^r$  and  $seq_j^r$ 
5:    $S_{u_i} \leftarrow S_{u_i} \cup \{seq_i^r\}$ 
6:    $S_{u_j} \leftarrow S_{u_j} \cup \{seq_j^r\}$ 
7:   for each  $v_k \in seq_i^r$  do
8:      $t_{v_1} \leftarrow seq$ 
9:      $t_{v_k} \leftarrow t_{v_k} - t_{v_1}$ 
10:     $t_{v_1} \leftarrow 0$ 
11:   end for
12:   for each  $v_k \in seq_j^r$  do
13:      $t_{v_1} \leftarrow seq$ 
14:      $t_{v_k} \leftarrow t_{v_k} - t_{v_1}$ 
15:      $t_{v_1} \leftarrow 0$ 
16:   end for
17:   for each  $v_n \in seq_i^r$  do
18:      $set_n \leftarrow \{\phi\}$ 
19:   end for
20:   for each  $v_m \in seq_j^r$  do
21:     for each  $v_n \in seq_i^r$  do
22:        $t_{mn} \leftarrow |t_m - t_n|$ 
23:     end for
24:     find the minimum value of  $t_{mn}$ 
25:      $set_n \leftarrow set_n \cup \{v_m\}$ 
26:   end for
27: end for

```


Algorithm 3 The similarity computation of two users.**Input:** $S_{u_i}, S_{u_j}, seq_{ij}$ **Output:** $sim(u_i, u_j)$

- 1: **for** each $seq_i^r \in S_{u_i}$ **do**
- 2: **for** each $v_n \in seq_i^r$ **do**
- 3: $sim(v_n, set_n) = \frac{\sum_{v_m \in set_n} \cos(v_n, v_m)}{|set_n|}$
- 4: **end for**
- 5: Calculate the category differentiation $F(seq_{ij}^r)$ according to Equation (5)
- 6: Calculate $sim(seq_i^r, seq_j^r)$ according to Equation (6)
- 7: **end for**
- 8: Calculate $sim(u_i, u_j)$ according to Equation (7)

Algorithm 4 User behavioral preference calculation based on the CF algorithm.**Input:** the target user u_i , the target POI v_j , all check-in records CH **Output:** user behavioral preference probability $pscore(u_i, v_j)$

- 1: **for** $u_j \in U$ **do**
- 2: $ch_j \leftarrow CH$
- 3: **for** each $v_k \in ch_j$ **do**
- 4: Initialize latent vector v_k by embedding model
- 5: **end for**
- 6: Calculate u_j by aggregating $v_k \in ch_j$
- 7: $u_j = \frac{\sum_{v_k \in ch_j} v_k}{|ch_j|}$
- 8: $p_{u_j, v_j}^{context} \leftarrow \cosine(u_j, v_j)$
- 9: **end for**
- 10: Calculate $pscore(u_i, v_j)$ according to Equation (8)

4.2. Construct Virtual Common Access Sequence

User trajectory similarity is used to represent the similarity of access preferences among users in this paper. To find users with high similarity to the target user, we explore the similar behavior of users in the historical check-in sequence. Algorithm 2 shows the method of constructing a virtual common access sequence $seq_{ij} = \{seq_{ij}^1, \dots, seq_{ij}^r, \dots, seq_{ij}^n\}$ based on the check-in history of two users u_i and u_j , and seq_{ij}^r denotes the virtual common sequence during any period of time.

Firstly, the check-in sequences of users are generated in different periods. Affected by time factors, users at different time of the day will show different check-in behavior. Therefore, dividing a day into different periods can better express the user's check-in behavior. Specifically, one day is divided into four time periods $T = \{T_1, T_2, T_3, T_4\}$, where $T_1 = \{0 : 00 - 6 : 00\}$, $T_2 = \{6 : 00 - 12 : 00\}$, $T_3 = \{12 : 00 - 18 : 00\}$, and $T_4 = \{18 : 00 - 24 : 00\}$. The similarity between users is calculated by analyzing the historical behavior of target users similar to other users in each period. As shown in Figure 2, the target user u_i visits $v_1, v_2, \dots, v_n$ at a period T_r , then the check-in sequence can be expressed as $seq_i^r = (v_1, v_2, \dots, v_n)$. In the initialization, the corresponding set_k of POI v_k visited by the target user is set to empty. As a result, a check-in sequence has n sets, where n denotes the number of POIs contained in the check-in sequence.

Secondly, we adjust the POI check-in time to explore the behavior pattern between users in each period (Lines 7–11 in Algorithm 2). For a given period T_r , the similarity between users is mainly manifested by similar historical behavior between users. However, the constructed check-in sequence cannot accurately describe the similar check-in behavior patterns of the two users. For instance, u_i visited the park, library, and company from 6:00 a.m. to 9:00 a.m. u_j visited the park, library, and company from 9:00 a.m. to 12:00 p.m. Although u_i and u_j have similar check-in behavior patterns, they are regarded as users

with different preferences due to the difference in check-in time. Therefore, to accurately describe the similarity between users, we try to construct a virtual common access sequence by processing the check-in time of each POI. Specifically, for each check-in sequence, set t_{v_1} to be the check-in time of the first record in the sequence, and change the check-in time of other records as to maintain the time offset from the timestamp of the first record. After adjusting the time, the POIs of user's sequence remain the same as before.

Figure 2. An example of construction of virtual common access sequence in T_r .

Finally, in a certain period, we construct the virtual common access sequence of two users seq_i^r and seq_j^r . Initializing a set set_n for each POI $v_n \in seq_i^r$, and, for each POI $v_m \in seq_j^r$, we calculate the check-in time interval between v_n and v_m —then, to find the minimum check-in time interval and divide v_m into the set_n corresponding to v_n . As shown in Figure 2, the virtual common access sequence of user u_i and user u_j is constructed in the divided period T_r , and the visited POI of user u_j is assigned to the corresponding set. The empty set can be obtained by the construction method. For example, if v_3^j and v_2^j has the minimum check-in time interval to v_2^i , then they are divided into set_2 .

4.3. New User Similarity Calculation

After constructing the virtual common access sequence, we propose a novel method to calculate the similarity between users. It is shown in Algorithm 3, and the similarity can be calculated by contextual information and category differentiation.

The check-in behavior of users is often affected by contextual information, through which users' behavior habits and movement patterns can be fully explored. We introduce the word embedding model in this paper, a single POI is treated as a word, and each POI is converted into a latent vector. By constructing the virtual common sequence, we can transform the trajectories' similarity calculation of two users into the calculation of the similarity between each $v_n \in seq_i^r$ and its corresponding set set_n . The calculation is as follows:

$$sim(v_n, set_n) = \frac{\sum_{v_m \in set_n} \cos(v_n, v_m)}{len(set_n)} \quad (4)$$

Due to the sparsity of check-in records, there are few common POIs accessed by users. Thus, it is difficult to find users with the same access preference. We introduce the concept of category differentiation to find users with semblable access preferences. The similarity between users can be effectively calculated through users' access information to the category. Category differentiation evaluates the degree of similarity among users based on their preference for accessing categories. Inspired by the inverse document frequency (IDF), a category accessed by most users does not reflect the user's personalized access preference, while a category accessed by only a few users can reflect the user's preference,

and users accessing this category often have high similarity. The category differentiation $F(seq_{ij}^r)$ is shown as follows:

$$F(seq_{ij}^r) = \begin{cases} \frac{1}{|sc_{ij}|} \sum_{q=1}^{|sc_{ij}|} \log \frac{|U|}{n_q} & \text{if } sc_{ij} \neq \phi \\ 0 & \text{if } sc_{ij} = \phi \end{cases} \quad (5)$$

The similarity of the check-in sequence in any period can be obtained by calculating $sim(v_n, set_n)$ and $F(seq_{ij}^r)$, and the similarity between users can be obtained by measuring the results of each time period. The specific computation formula is given in (6) and (7):

$$sim(seq_i^r, seq_j^r) = \frac{\sum_{n=1}^{len(seq_i^r)} sim(v_n, set_n)}{len(seq_i^r)} \cdot F(seq_{ij}^r) \quad (6)$$

$$sim(u_i, u_j) = \frac{\sum_{r=1}^4 sim(seq_i^r, seq_j^r)}{4} \quad (7)$$

Specifically, seq_{ij}^r indicates the set of common categories accessed by two users during given period T_r , and n_q denotes the total number of users visiting the q th category in seq_{ij}^r , and $len(\cdot)$ is used to compute the length of the user's sequence.

4.3.1. POI Recommendation Based on CF

The user-based CF algorithm aims to recommend the POI visited by users with similar preferences. It only needs to analyze the similarity between the target user and other users according to the historical check-in records, and then predicts the check-in probability of the target user at a certain POI based on the access POIs of similar users and generates a recommendation list. It is shown in Algorithm 4, user similarity is obtained based on Algorithm 3 introduced before, and the formula of getting the user's behavioral preference probability of POI can be expressed as:

$$pscore(u_i, v_j) = \frac{1}{\sum_{u_j \in |F_{u_i}|} sim(u_i, u_j)} \cdot \sum_{u_j \in |F_{u_i}|} sim(u_i, u_j) \cdot c_{u_j, v_j} \cdot p_{u_j, v_j}^{context} \quad (8)$$

where $|F_{u_i}|$ is the user similarity set, c_{u_j, v_j} denotes whether a user u_j visits a POI v_j , and its value is 0 or 1, $p_{u_j, v_j}^{context}$ is the user's visiting preference probability for POI obtained according to Equation (10).

Since a user's POI check-in sequence reflects the user's preferences, the embeddings of POI v_j in their check-in sequence can be used to model the user's preferences. In particular, we calculate u_j with average aggregation, and it can maintain the integrity and smoothness of input embedding with linear transformation:

$$u_j = \frac{\sum_{v_k \in ch_j} v_k}{|ch_j|} \quad (9)$$

The preference probability of user accessing to POI can be expressed as:

$$p_{u_j, v_j}^{context} = \cos(u_j, v_j) = \frac{u_j \cdot v_j}{\sqrt{|u_j|^2 \cdot |v_j|^2}} \quad (10)$$

4.3.2. POI Recommendation Based on KDE

As Tobler's first geographical law points out, geographical objects are interrelated in spatial distribution, with clustering, randomness, and regularity. The closer the distance is, the closer the relationship will be. In practice, the geographical impact should be different for each user. For example, some people prefer to usually visit close POIs, and some

people who prefer to travel by vehicle usually explore far POIs. Users' mobile behavior will be affected by geographical distance, and their check-in records have certain spatial distribution characteristics for POI recommendations. We randomly selected three users from the dataset. Figure 3 describes the check-in distribution on the distance between each pair of POIs on their check-in records. It can be seen that the geographical impact is different for each user. User 1 likes to access the short-range POI, user 2 likes to access the POI beyond a certain distance, and user 3 has the same access frequency within a certain range. Therefore, the impact of geographical information on users should not be modeled as a general distribution, and it should be modeled as personalized. Thus, it is necessary to study the personalized geographical impact for user check-in behaviors.

Figure 3. Personal check-in probabilities over geographical distances.(a–c) user 1–user 3.

It is found that KDE can obtain the personalized distribution characteristics of POIs based on the users' historical check-in records, which brings great convenience to the POI recommendation. Thus, we adopt KDE to model user personalized preference for POI from a geographical aspect. The accuracy of kernel density estimation depends largely on the selection of $K(\cdot)$ and h , where $K(\cdot)$ is the kernel function and σ is standard deviation. According to the characteristics of check-in data, the Gaussian kernel function is used as the kernel function in this paper. h is the bandwidth obtained by calculating the standard deviation of POI according to the historical check-in records of users:

$$K(\cdot) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} \quad (11)$$

$$h = \left(\frac{4\sigma^5}{3n} \right)^{\frac{1}{5}} \quad (12)$$

Based on the influence of geographical factors, the probability of user u_i visiting a POI is as follows:

$$p^{geo}(v_j|ch_i) = \frac{1}{nh^2} \sum_{l=1}^n K \frac{(lat_{v_{l+1}} - lat_{v_l}) - (lon_{v_{l+1}} - lon_{v_l})}{h} \quad (13)$$

5. Results

5.1. Datasets

In the experiment, we use two datasets provided by Foursquare [27], which includes the check-ins of New York City and Tokyo from 12 April 2012 to 16 February 2013. The dataset consists of five attributes with the user identifier (ID), check-in place, location longitude, location latitude, and check-in time. In Ref. [20], Li et al. take 10 as the threshold to delete inactive users and inactive POIs. Inspired by this, we make some preprocessing on the dataset to eliminate the users who has visited less than 10 different POIs and those

POIs which are being visited by users less than 10 times. The basic statistics of two datasets are shown in Table 1. In addition, 80% of the check-ins are randomly selected as training data and 20% of the check-ins are for testing.

Table 1. Statistics of two datasets.

DATASETS	New York	Tokyo
Users	825	2214
Venus	1086	2877
Check-ins	40,197	329,744
Sparsity	0.9552	0.9483

5.2. Evaluation Metrics

The performance of the model proposed in this paper is evaluated using Precision@k, Recall@k, and F1 score.

Precision rate is the ratio of correctly predicted POI to the total number of recommended POI:

$$Precision@k = \frac{1}{|U|} \sum_{u \in U} \frac{|L_k \cap L_{visited}|}{|L_k|} \quad (14)$$

Recall rate is the ratio of the correctly predicted POI to the total number of POI actually visited:

$$Recall@k = \frac{1}{|U|} \sum_{u \in U} \frac{|L_k \cap L_{visited}|}{|L_{visited}|} \quad (15)$$

F1 score is based on the accuracy and recall rate of comprehensive evaluation index:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (16)$$

where L_k is the recommended list of Top-k POIs that user u_i would like to visit, and $L_{visited}$ denotes the list of POIs that user u_i has visited in the testing.

5.3. Comparative Method

Five POI recommendation methods are chosen as the baseline methods for comparison.

PMF [3]: It is a POI recommendation model based on matrix factorization, which can predict the access probability of users to POI by decomposing the information obtained by users' POI access matrix.

LRT [12]: It is a POI recommendation model integrating time information based on matrix decomposition, which considers the correlation between check-in location and check-in time.

PFMMGM [2]: It is a recommendation system based on matrix factorization, and it captures geographic influence through a multicenter Gaussian model and integrates social information and geographic influence into the matrix factorization framework.

CPAM [26]: It is a context and preference perception model, through a skip-Gram-based POI embedding model to calculate users' preferences for target POIs, and combined with the logical matrix decomposition algorithm to mine users' preferences for POIs.

Li [20]: It is a context-aware POI recommendation model based on the CF framework. The recommendation list is generated by studying the influence of time and space characteristics on users.

5.4. Results Analysis

The proposed method is compared with other methods, and the influence of parameters on the recommendation is discussed.

5.4.1. Performance Comparison

The proposed model SCGM is compared with the five baseline methods on the New York dataset and Tokyo dataset. The SCGM model obviously has better performance than other algorithms in precision, recall, and F1 score.

Figure 4 shows the precision of each algorithm on two datasets. It can be seen that the precision of the six algorithms on the Tokyo dataset is obviously higher than that on the New York dataset. On the other hand, it can clearly found that the precision of the algorithm decreases with the increase of POI recommended list length. Our algorithm has better performance than others when the list of recommended length is 5, 10, and 20, respectively. In addition, taking the list length of 5 as an example, the precision of LRT and PMF is not high, and neither of the two algorithms takes into account the influence of geographical information, resulting in poor recommendation performance. The precision of PFMMGM, CPAM, and Li's method are close. However, the precision of our algorithm is higher than all the baseline algorithms. On the Tokyo and New York two datasets, the Li's method has the best performance among the five baseline algorithms, and the precision@5 of Li's method reaches 0.054 and 0.024, respectively. SCGM gets the precision@5 value of 0.069 and 0.032, which realizes an improvement of nearly 28% and 33% than Li's method.

Figure 4. Precision results of six methods on two datasets. (a) precision@5, (b) precision@10, (c) precision@20.

Figure 5 illustrates the recall of the six algorithms on the datasets. The recall of the algorithm increases with the length of the POI recommendation list varying from 5 to 20. The recall performance of each algorithm is the best when $k = 20$. LRT and PMF algorithms have the lowest recall@20 value. On the New York dataset, the recall@20 value of PFMMGM and CPAM are both below 0.04, which are significantly lower than that of Li's method and SCGM. The recall@20 value of Li's method and SCGM are 0.05 and 0.06 on the New York dataset. In addition, the recall@20 of PFMMGM, CPAM, and Li's method are around 0.045 on the Tokyo dataset. They are still lower than that of the SCGM algorithm. Therefore, SCGM outperforms other algorithms and has good performance on sparse datasets.

The F1 score reflects the overall performance of the algorithm. As is shown from Figure 6, PMF and LRT algorithms perform poorly, and F1 scores are much lower than other algorithms. The performance of PMFMGM and CPAM algorithms is better than PMF and LRT algorithms. Among them, PMFMGM models the geographical impact based on user's check-in records and achieves better performance, and CPAM considers the implicit feedback and complex contextual impact on the check-in records. Moreover, Li's method performs better than PMFMGM, CPAM, and PMF. It comprehensively considers the influence of time and space factors on user check-in. Furthermore, the proposed SCGM algorithm achieves the best performance. On the New York dataset ($k = 20$), the F1 score of PMFMGM, CPAM, Li's method, and SCGM reaches 0.014, 0.017, 0.024, and 0.029, respectively. On the Tokyo dataset ($k = 10$), compared with PMFMGM, CPAM, and Li's method,

the F1 score of our method SCGM improved by 66%, 56%, and 31% separately. The results show that SCGM, taking the advantage of sequence, time information, category information, and geographic information, can significantly improve the overall performance of POI recommendations.

Figure 5. Recall results of six methods on two datasets. (a) recall@5, (b) recall@10, (c) recall@20.

Figure 6. F1 score results of six methods on two datasets. (a) F1 score@5, (b) F1 score@10, (c) F1 score@20.

5.4.2. Effect of Parameter

It is necessary to study the parameter α , and it indicates the importance of geographical influence in user decision-making. Figure 7 illustrates that precision@5, recall@5, and F1 score with different weight α of two datasets. The parameter α varies from 0.1 to 0.9. The figure shows that the results reach the peak when $\alpha = 0.2$ and $\alpha = 0.4$ on the two datasets of Tokyo and New York separately. It can be inferred that people in different regions have different living habits, and their geographical influence is also different. Therefore, it is necessary to adjust parameter α on different datasets. Thus, we choose $\alpha = 0.2$ as the parameter of Tokyo dataset and $\alpha = 0.4$ as the parameter of the New York dataset in the experimental evaluation.

Figure 7. Effect of parameter α on two datasets. (a) precision, (b) recall, (c) F1 score.

6. Conclusions

This paper proposes a POI recommendation model (SCGM), which integrates sequential, category, and geographical factor to generate the POI recommendation list. According to the CBOW model, the latent vectors of user preference are computed for POI from the user's check-in sequence. Then, we construct a virtual common access sequence for users and design a new user similarity computation method via combining category differentiation and POI latent vector, and apply it to the CF recommendation framework. Furthermore, the kernel density estimation method is employed to model user's personalized check-in behavior. In the end, a list of recommended POIs is obtained based on the user's preference probability of POI computed with the combination of CF and KDE. Experiments on two LBSN datasets show that SCGM is superior to other POI recommendation algorithms in terms of precision, recall, and F1 score. In addition, the proposed POI recommendation algorithm can be applied to online tourism service area. It can provide a user with the hotel or suitable scenic spot based on his personal consumption habit and travel preference to reduce the complexity of tourism planning for users. In the future, we will further optimize the trajectory similarity calculation method to improve the performance of POI recommendations and explore efficient ways to protect users' private information. Furthermore, we would like to extend the algorithm and make a recommendation system for some application scenarios, such as marketing and aviation.

Author Contributions: Conceptualization, Xican Wang and Xu Zhou; Methodology, Xican Wang; Validation, Xican Wang, Yanheng Liu, Xueying Wang, and Zhaoqi Leng; Formal analysis, Xican Wang; Writing—original draft preparation, Xican Wang; Writing—review and editing, Xican Wang and Xu Zhou; Supervision, Yanheng Liu and Xu Zhou; Funding acquisition, Xu Zhou and Yanheng Liu. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China under Grant Nos. 61806083, 61872158, and 62172186) and the Fundamental Research Funds for the Central Universities, JLU under Grant No. 93K172021Z02.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on the website: <https://sites.google.com/site/yangdingqi/home/foursquare-dataset>, <https://www.kaggle.com/chetanism/foursquare-nyc-and-tokyo-checkin-dataset/version/1> accessed on 1 December 2021.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Liu, X.; Yang, Y.; Xu, Y.; Yang, F.; Huang, Q.; Wang, H. Real-time POI recommendation via modeling long- and short-term user preferences. *Neurocomputing* **2022**, *467*, 454–464. [\[CrossRef\]](#)
2. Cheng, C.; Yang, H.; King, I.; Lyu, M.R. Fused matrix factorization with geographical and social influence in location-based social networks. In Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence, Toronto, ON, Canada, 22–26 July 2012; Association for the Advancement of Artificial Intelligence: Menlo Park, CA, USA, 2012; Volume 26, p. 1.
3. Salakhutdinov, R.; Mnih, A. Probabilistic Matrix Factorization. In *Advances in Neural Information Processing Systems 20, Proceedings of the Twenty-First Annual Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 3–6 December 2007*; Neural Information Processing Systems Foundation: La Jolla, CA, USA, 2007; pp. 1257–1264.
4. Yuan, Q.; Cong, G.; Sun, A. Graph-based point-of-interest recommendation with geographical and temporal influences. In Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, CIKM 2014, Shanghai, China, 3–7 November 2014; ACM, Inc.: Tipp City, OH, USA, 2014; pp. 659–668.
5. Jiao, X.; Xiao, Y.; Zheng, W.; Wang, H.; Hsu, C. A novel next new point-of-interest recommendation system based on simulated user travel decision-making process. *Future Gener. Comput. Syst.* **2019**, *100*, 982–993. [\[CrossRef\]](#)
6. Rahmani, H.A.; Aliannejadi, M.; Zadeh, R.M.; Baratchi, M.; Afsharchi, M.; Crestani, F. Category-aware location embedding for point-of-interest recommendation. In Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval, ICTIR 2019, Santa Clara, CA, USA, 2–5 October 2019; ACM, Inc.: Tipp City, OH, USA, 2019; pp. 173–176.
7. He, J.; Li, X.; Liao, L. Category-aware next point-of-interest recommendation via listwise bayesian personalized ranking. In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, 19–25 August 2017; ACM Digital Library: New York, NY, USA, 2017; Volume 17, pp. 1837–1843.
8. Lyu, Y.; Chow, C.; Wang, R.; Lee, V.C.S. iMCR: A multi-criteria framework for personalized point-of-interest recommendations. *Inf. Sci.* **2019**, *483*, 294–312. [\[CrossRef\]](#)
9. Zhao, S.; Zhao, T.; King, I.; Lyu, M.R. Geo-teaser: Geo-temporal sequential embedding rank for point-of-interest recommendation. In Proceedings of the 26th International Conference on World Wide Web Companion, Perth, Australia, 3–7 April 2017; ACM, Inc.: Tipp City, OH, USA, 2017; pp. 153–162.
10. Ye, M.; Yin, P.; Lee, W.; Lee, D.L. 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, SIGIR 2011, Beijing, China, 25–29 July 2011; ACM, Inc.: Tipp City, OH, USA, 2011; pp. 325–334.
11. Chen, J.; Zhang, W.; Zhang, P.; Ying, P.; Niu, K.; Zou, M. Exploiting Spatial and Temporal for Point of Interest Recommendation. *Complexity* **2018**, *2018*, 6928605:1–6928605:16. [\[CrossRef\]](#)
12. Gao, H.; Tang, J.; Hu, X.; Liu, H. Exploring temporal effects for location recommendation on location-based social networks. In Proceedings of the Seventh ACM Conference on Recommender Systems, Hong Kong, China, 12–16 October 2013; ACM, Inc.: Tipp City, OH, USA, 2013; pp. 93–100.
13. Hosseini, S.; Li, L.T. Point-of-interest recommendation using temporal orientations of users and locations. In Proceedings of the Database Systems for Advanced Applications—21st International Conference, DASFAA 2016, Dallas, TX, USA, 16–19 April 2016; Proceedings, Part I; Springer: Berlin, Germany, 2016; Volume 9642, pp. 330–347.
14. Gao, R.; Li, J.; Li, X.; Song, C.; Chang, J.; Liu, D.; Wang, C. STSCR: Exploring spatial-temporal sequential influence and social information for location recommendation. *Neurocomputing* **2018**, *319*, 118–133. [\[CrossRef\]](#)
15. Gan, M.; Gao, L. Discovering memory-based preferences for POI recommendation in location-based social networks. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 279. [\[CrossRef\]](#)
16. Zhang, H.; Gan, M.; Sun, X. Incorporating memory-based preferences and point-of-Interest stickiness into recommendations in location-based social networks. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 36. [\[CrossRef\]](#)
17. Bao, J.; Zheng, Y.; Mokbel, M.F. Location-based and preference-aware recommendation using sparse geo-social networking data. In Proceedings of the SIGSPATIAL 2012 International Conference on Advances in Geographic Information Systems (Formerly Known as GIS), Redondo Beach, CA, USA, 7–9 November 2012; ACM, Inc.: Tipp City, OH, USA, 2012; pp. 199–208.
18. Liu, X.; Liu, Y.; Aberer, K.; Miao, C. Personalized point-of-interest recommendation by mining users' preference transition. In Proceedings of the 22nd ACM International Conference on Information and Knowledge Management, San Francisco, CA, USA, 27 October–1 November 2013; ACM, Inc.: Tipp City, OH, USA, 2013; pp. 733–738.
19. Zhou, D.; Rahimi, S.M.; Wang, X. Similarity-based probabilistic category-based location recommendation utilizing temporal and geographical influence. *Int. J. Data Sci. Anal.* **2016**, *1*, 111–121. [\[CrossRef\]](#)
20. Li, M.; Zheng, W.; Xiao, Y.; Zhu, K.; Huang, W. Exploring temporal and spatial features for next POI recommendation in LBSNs. *IEEE Access* **2021**, *9*, 35997–36007. [\[CrossRef\]](#)
21. Wang, H.; Shen, H.; Ouyang, W.; Cheng, X. Exploiting POI-specific geographical influence for Point-of-interest recommendation. In Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, Stockholm, Sweden, 13–19 July 2018; ACM, Inc.: Tipp City, OH, USA, 2018; pp. 3877–3883.
22. Rahmani, H.A.; Aliannejadi, M.; Ahmadian, S.; Baratchi, M.; Afsharchi, M.; Crestani, F. LGLMF: Local geographical based logistic matrix factorization model for POI recommendation. In Proceedings of the Information Retrieval Technology: 15th Asia Information Retrieval Societies Conference, AIRS 2019, Hong Kong, China, 7–9 November 2019, Proceedings; Springer: Berlin, Germany, 2020; Volume 12004, pp. 66–78.

23. Li, X.; Cong, G.; Li, X.; Pham, T.N.; Krishnaswamy, S. 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, Santiago, Chile, 9–13 August 2015; ACM, Inc.: Tipp City, OH, USA, 2015; pp. 433–442.
24. Liu, B.; Fu, Y.; Yao, Z.; Xiong, H. Learning geographical preferences for point-of-interest recommendation. In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2013, Chicago, IL, USA, 11–14 August 2013; ACM, Inc.: Tipp City, OH, USA, 2013; pp. 1043–1051.
25. Yin, H.; Wang, W.; Wang, H.; Chen, L.; Zhou, X. Spatial-aware hierarchical collaborative deep learning for POI recommendation. *IEEE Trans. Knowl. Data Eng.* **2017**, *29*, 2537–2551. [[CrossRef](#)]
26. Yu, D.; Wanyan, W.; Wang, D. Leveraging contextual influence and user preferences for point-of-interest recommendation. *Multimed. Tools Appl.* **2021**, *80*, 1487–1501. [[CrossRef](#)]
27. Yang, D.; Zhang, D.; Zheng, V.W.; Yu, Z. Modeling user activity preference by leveraging user spatial temporal characteristics in LBSNs. *IEEE Trans. Syst. Man Cybern. Syst.* **2015**, *45*, 129–142. [[CrossRef](#)]