Received: November 11, 2021. Revised: March 22, 2022.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

An Overlapping Community Framework for Personalized Location Recommendation System

Khaled Soliman[1]* **Mahmood Mahmood[1,3]** **Hesham Hefny[2]** **Nagy Darwish[1]**

[1]*Department of Information Systems and Technologies, Faculty of Graduate Studies for Statistical Research, Cairo University, Egypt*
[2]*Department of Computer Science, Faculty of Graduate Studies for Statistical Research, Cairo University, Egypt*
[3]*Department of Information Systems, College of Computer Science and Information, Jouf University, Saudi Arabia*
* Corresponding author's Email: khaled.mohamed@pg.cu.edu.eg

Abstract: This paper proposes a novel location recommendation method that provides the most interesting locations based on personal preferences. The paper provides a framework for overlapping community-based personalized location recommendation system (CMPR). The framework has two main phases: offline clustering and online recommendation. In the offline clustering phase, the proposed framework incorporates geographical, categorical, and timing preferences to construct overlapping user's communities through which popular and interesting locations are introduced based on Geo-Communities location that uses spatial group analysis as a technique for cluster construction. In the online recommendation phase, similar users, based on the overlapping user's communities, are provided with candidate locations in geo-community with higher popularity. In addition, the framework uses k-nearest neighborhood (KNN) to classify the overlapping between similar users based on candidate location participation. Experiments show that the proposed framework outperforms state-of-the-art POI recommendation approaches in terms of quality, efficiency, and performance of the recommendation process and achieve precision 87% in Foursquare dataset, 84% in Gowalla dataset where N =10 and achieve precision 79% in Foursquare dataset, 77% in Gowalla dataset where N =20.

Keywords: Spatial data mining, Location recommendation, Hybrid method, Community recommendation system, Categorical hierarchy, Location popularity.

1. Introduction

Nowadays, with the growing use of the Internet and mobile devices, different web platforms such as Twitter, Facebook, Foursquare, and Gowalla have implemented social network environments. Social networks (SN) present different services and facilitate the connection among users sharing behaviors and similar interests. SN can provide location-based services for people to check-in to different places. Hence, they are called 'location-based social networks' (LBSNs) [1, 2]. LBSN, also referred to as geo-social network, is defined as a specific type of social networking platforms that complements traditional social networks with some geographical services [3].

Recommender systems (RS) are bridging users and relevant products, services, and peers on the Web by using user preferences and other information (e.g., social friendships). So far, the personalized recommendation has been the most effective solution to information overload problem which resulted from complex growth of the internet and social networks [4, 5]. LBSNs collect user check-in information including geographical details and visiting places to be shared by other users, they can share their experiences and information about visiting locations, like museums, restaurants, and cinemas …etc., which are called Points-Of-Interest (POIs).

Fig. 1 illustrates statistics check-in in two publicly datasets [1]. Most users only have fewer check-in record and sparsity of row data confines the precision and accuracy in gathering personalized

Received: November 11, 2021. Revised: March 22, 2022.

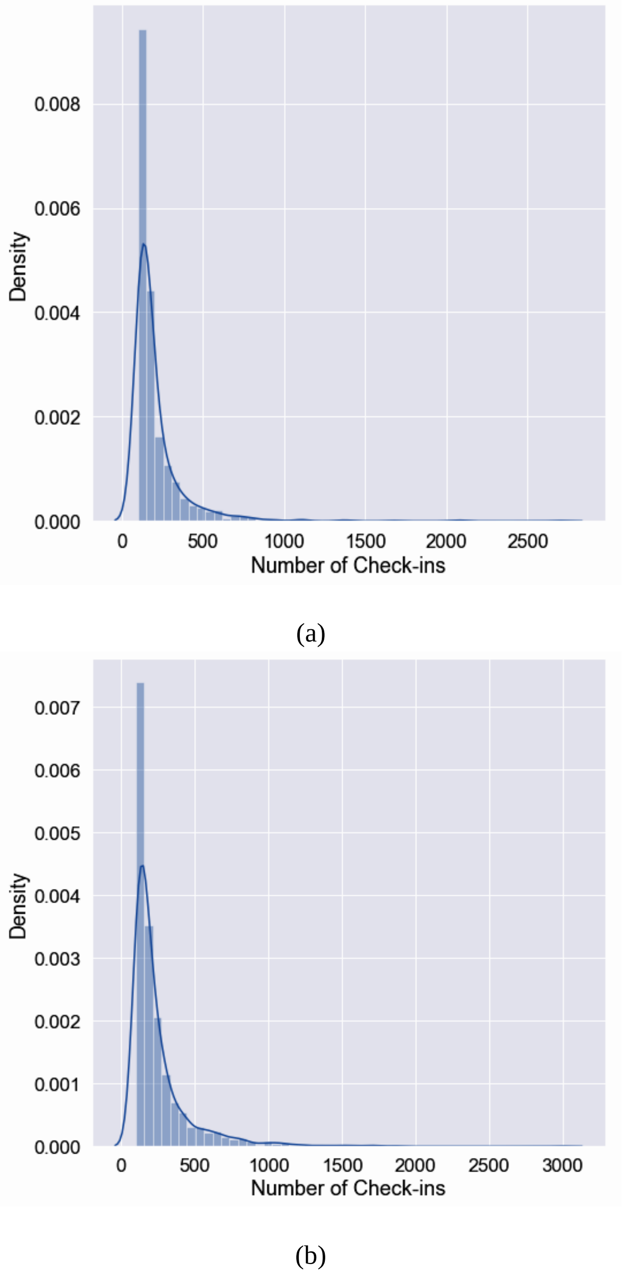


Figure. 1 Statistics of check-ins in: (a) NYC and (b) TKY

preferences of users. Subsequently, the critical issue resolved by personalized location recommendation is to properly model preferences from a few check-ins.

Detecting user groups in the LBSN has become a research challenge, known as 'user-cluster' or 'user-community' problem, which indicate to the network users cluster which they are matching in item preferences [6]. Therefore, in the same community to improve the information access of online users could be used social associations formed among members as an information source, also used for users who are engaged in various communities to enhance personalized recommendations [7]. Fig. 2 shows a user-community and overlapping user-community

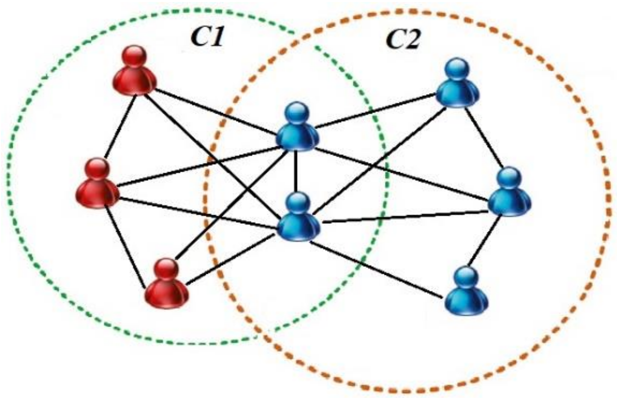


Figure. 2 Overlapping community membership

where the user may have a membership in multiple user-community.

To generate a personalized location recommendation system on LBSN, both quality and efficiency should be considered, and to achieve high performance it requires to devise an effective structure and efficient algorithms, and to achieve high quality it need to consider the following factors: user preference, users comment on location, social connection. For utilizing this recent research get the following challenges:

- Geographical influence with different communities and user historical preferences, it is complex to detecting users' community and recommend new POI to match with this factor.
- There is extensive information of user check-ins, which can have a considerable role in clustering.
- New user, they have not known his/her preferences and it is hard to recommend (also known as, cold start problem).
- Data sparseness problem, which leads to user location matrix very large but sparse.

In this paper, we propose a new approach of location recommendation that inferring personalized preferences of users to locations by categorical, geographical, and timing integrations. Throughout this paper, we try to give answers to the following questions: "What are overlapping communities?" (i.e. user similarity problem), "What is the performance measure?" (i.e. a time complexity problem), "How does the new user affect the recommendation?", (i.e. a cold start problem), also "How data sparseness affects inferring personalized user preferences" (i.e. data sparseness problem).

Our main contribution of this paper is:

- Introducing a new framework of location recommendation that associate categorical, geographical, and timing preferences to construct overlapping users' communities.
- Popularity of location is used to inferring users' visited probability history of candidate locations to introduce spatial geo-communities of locations based spatial clustering method.
- Using user preferences to extract candidate users with a high similarity score based on overlapping user communities.
- Exploiting location popularity to extract candidate locations based on geo-community, (unlike existing studies).
- Evaluating the proposed recommendation approach through experiments on two datasets with large-scale, that collected from Foursquare and Gowalla; therefore, it is compared with different location recommendation approaches.

The remainder of this paper is organized as follows: section 2 presents the related work then section 3 presents the proposed framework, section 4 illustrates the experimental result and discussions and finally, section 5 concludes the paper.

2. Related work

Geographical information is the most exploited to model user preference based on mining check-in patterns, that explore the types of locations which users preferred. It has been generally utilized in numerous applications of location-based, containing geo-specific tag recommendation [8], similar user discovery [9], friendship prediction between users [3, 10], and travel recommendation [11], but they don't consider users preference and temporal factor which enhancing recommendation result.

Categorical information that indicates the visited locations is introduced to profile user preference [12], and it is used for calculating the similarity between users [13]. [14] propose a weighted category hierarchy using 'Term frequency-inverse term document frequency' (TF-IDF) to recognize the weight of categories. categorical influences can be collaborating with a geographical role to alleviate the data sparsity issue in POI recommendation [15], but authors don't consider timing and geographical influence with user preference.

Timing information is the most exploited to model user preference based on mining check-in patterns, it can be divided 24 hours into some periods or matrix of hours following the time law of people's work and life, so that user similarity calculated by such periods will be more accurate [10, 16]. The effective POI recommendation on LBSNs generally combines crucial factors. [17] from historical check-in datasets on LBSNs used a probabilistic statistical analysis method to extract three crucial factors of user activity time-based, POI popularity, and distance features. [18] proposed a personalized recommendation framework with time constraints for the LBSN by exploiting geographical features and social relationships that satisfy user preferencesn, but this research lacking to extract user similarity-based POI category and geographical effectiveness between POI and users.

Community or grouping activities have usually contributed topic relevant information. The ability to construct communities has emerged as one of the most popular features in many types of recommendation systems [7], author improve personalized recommendation-based construct communities for user, but they don't consider timing and geographical influence. [19] developed a unique personalized recommender engine that collects user information's from user location records and is suitable for service discovery in a smart community, but this paper lacking to venues category and check-in timing. [20] it is used to extract users' interests using a community detection approach based on location interest mining, with the spatial-temporal topic, but they don't consider user similar based POI category. POI categorical preference is the most exploited to construct personalized POI groups [21], but it is also important to extract geographical preference for users. Therefore, community approach can use geographical limits and overlapping interest community information to improve POI recommendation accuracy [22], but the author construct communities based geographical factor and don't considering users similarity-based POI category and timing.

There are several methods to detect overlapping communities, [6] proposed an intelligent method based on fuzzy clustering concerning user check-ins of user and his behavior, and the features of location in LBSNs. [23] proposed a new location recommendation method to cluster users into several communities based on geographical distance and location popularity. POI communities [24] also proposed an overlapping community detection method based on 'Nonnegative Matrix Factorization' (NMF) is utilized for POI heterogeneous connection detection process, but they don't consider user-location similarity and timing of check-in for constructing communities.

Received: November 11, 2021. Revised: March 22, 2022.

In summary, combining categorical, social, and geographical information, as well as popularity can improve process of location recommendation. [25] this study is most related to ours, which proposed adoption of location popularity as a public adjustment factor, that help estimating preference of user-categorical by taking into consideration the category tags which hierarchy structure of it. Unlike their study, our study adopts a proposed framework for integrating categorical, geographical, and timing preferences to construct an overlapping user community and geo-community location, which provides the most interesting location based on their personal preferences.

3. The proposed CMPR:

In this section, we propose a community based personalized location recommendation framework (CMPR) by fully esteeming categorical and geographical factors of locations, and check-ins timing among users, with popular and interesting locations based on geo-community location.

3.1 System architecture:

As shown in Fig. 3, the architecture of the proposed system consists of two-part: community modelling (off-line clustering), and community-based personalized location recommendation (on-line recommendation). The first part (off-line clustering) it's divided into five steps: (a) Preference extraction modelling in which an extract user preference by location hierarchy category, this hierarchy capture user preference each category. (b) User geographical preference, which reflects the range of their activities and detects the geographical correlation between candidate locations and visited locations. (c) Time dimension, this step extract users who different check-in timing at different periods. (d) Constructing overlapping user communities, apply the fuzzy C-means algorithm for weighted category users, and they geographical preference and users timing preference to create overlapping similar weight user's communities. (e) Spatial community model, and create different Geo-community based (X, Y) coordinates.

Second part: Extract similar users with higher matching preference with requested user, also extract candidate locations with higher popularity in geo-community. Then create a user-location matrix from previous results and apply a KNN classifier for classifying users and ranking top N locations. Finally recommend top-N locations from the candidate locations.

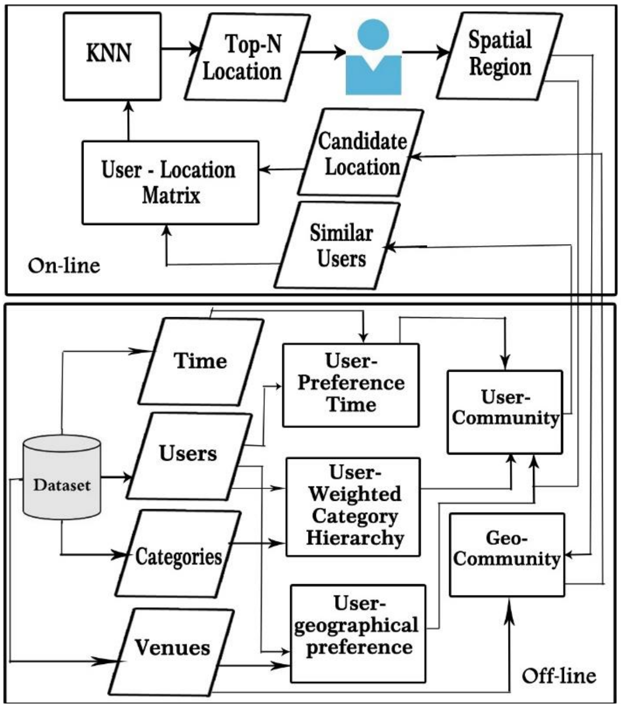


Figure. 3 System architecture

3.2 Community modeling (off-line clustering)

Community modelling has three stages

Stage 1: Preference extraction model (Feature engineering model):

In Preference extraction model have 3 steps:

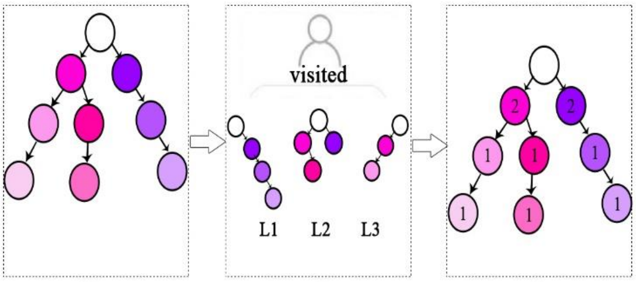
- *User categorical preference:*

Locations are usually described with categories regarding the activities in the places. The various levelled structure of classifications can be utilized to quantify the being related among areas [25].

First project a user's location history across all the cities onto a predefined category hierarchy, where nodes occurring on a deeper layer denote the categories of finer granularity. A personal category hierarchy tree is generated based on check-in history that represents the interests of a user, as shown in Fig. 4 (a). After recompensing the categories sets of all venues and struct category hierarchy framework, each location can be defined as a subtree of this framework. where each location category is represented as node and each category-subcategory relationship are parent-child edge. Finally, can be obtained location trees from multiple history of the user's check-in.

For instance, if any user (u) checked-in three venues, then create for this user three trees of location, as shown in Fig. 4 (b). Form category hierarchy tree,

Received: November 11, 2021. Revised: March 22, 2022.



(a) (b) (c)

Figure. 4 Category hierarchy tree of user u: (a) personal tree, (b) check-in tree, and (c)aggregate tree

this trees which u has checked-in can be connected through the root node, it used to infer the u personalized category preferences. In the same process, categories will be merged if different trees have the same layer category, and the number in the category is used to present the number of times of their appearance, as shown in Fig. 4 (c). There are many ways to do this like using hierarchical clustering algorithms or extract the category from POIs or a simple way by using existing knowledge bases to create a hierarchy. Overall, preference weight for user's (W (u, c)) is calculated by Eq. (3), first, the information category (IC) by [25], the value of each category C is measured using Eq. (1) (Lines 1–3 in Algorithm 1)

$$IC_{(c)} = -\log(P_{(c)}) \tag{1}$$

"Where $IC_{(c)}$ is the *IC* values of category *c*, and $P_{(C)}$ the probability that category c appears in the categories sets of the locations".

Then, calculate the score S for user u in category c by Eq. (2)

$$S(u,c) = \frac{|\{u.l_i : l_i.c = c\}|}{|u.l|} . \log\left(1 \left(\frac{|u|}{|\{u_j : c \in u_j.C\}|}\right)\right). \quad (log\ (1 + \frac{IC(c)}{P})) \tag{2}$$

"Where *{u.li : li.c = c}* is the number of times that user *u* visits category *c*, |*u.L*| is the size of the *u's* location histories, *{uj : c ∈ uj.C}* is the number of users who have visited *c*, |*U*| is the number of users and *P* is a probability that *U* appears in the category".

Finally, the weighted categorical preference W(u,c) is a calculation by Eq. (14)

$$W_{(u,c)} = \log(1 + S) \tag{3}$$

Algorithm 1 shows the pseudo-code of estimating categorical preference.

Algorithm 1: The pseudo-code of calculating W (u, c)	
Input:	The set of all categories C
	The set of users U
	Begin
	For each category in C do
1:	Calculate its information category value by Eq. (1).
2:	end
3:	For each ui in U do
4:	Calculate the score between ui and ci by Eq. (2).
5:	end
6:	Obtain the weighted categorical preference W (u, c) to all users by Eq. (3).
7:	End

The benefits of extract user's category preferences to present personalized recommendations is:

1. Not no need to share any physical location histories when computing similarity between users.
2. Inferring user preferences and recommend interesting locations.
3. reduced different data scales of different users and replace the user-location matrix with a category matrix.

- *User geographical preference:*

Users are preferring to check-in venues in regions they commonly visit. Modelling user's geographical preference can be used personalized check-in history for users [25]. The average distance between unvisited candidates and visited locations is represented as geographical proximity of candidate locations and user activity range. Therefore, to extract the similarity between users and locations are calculated the geographical correlation between candidate locations and visited locations from the geographical perspective. Therefore, by using a power-law function with geographical proximity we can inferring the geographical preference. The pseudo-code is shown in Algorithm 2.

First estimate the geographical weighted center user u check-in history locations by Eqs. (4) and (5) (Line 1in Algorithm 2) then infer the geographic distance between user activity range and candidate unvisited location by calculating the average distance between all unvisited locations and the geographically weighted center user u, which will be used as a public factor of geographical preference, we measure this distance by the haversine formula, it calculates the geographic distance between two (X,Y) coordinates [26] as Eq. (6) (Lines 2–4 in Algorithm 2).

Algorithm 2: The pseudo-code of calculating geographical preference.	
Input:	The set of locations L
	The set of locations that user u visited Lu
	Begin
	Obtain the geo weighted center for u by Eq. (4,5).
1:	**For** each *li* in *L* **do**
2:	Calculate the geographic distance between *lj* and L_w by Eq. (6).
3:	**end**
4:	Obtain the average distance between the locations in *L* and L_w by Eq. (7).
5:	Obtain the geographical preference ($GeoP_{(u,l)}$) to all candidate locations by Eq. (8).
6:	End

$$\varphi_w = \frac{\sum_{i=1}(lat\ .\ T)}{\sum_{i=1} T} \quad (4)$$

$$\lambda_w = \frac{\sum_{i=1}(long\ .\ T)}{\sum_{i=1} T} \quad (5)$$

"Where φ_w is the weighted latitude and λ_w is weighted longitude, otherwise, T is the number of times the user was checked-in this venue".

$$Dis_{(l_i,l_w)} = 2.r.arcsin\left(\sqrt{sin^2\left(\frac{\varphi_w-\varphi_i}{2}\right)+\cos\varphi_i\cos\varphi_w\, sin^2\left(\frac{\lambda_w-\lambda_i}{2}\right)}\right), l_i \epsilon L, l_w \epsilon L_u \quad (6)$$

where r is the radius of the sphere; φ_i and φ_w are the latitudes of li and lw, respectively; λi and λw are the longitudes of li and lw respectively. Algorithm 2 shows the pseudo-code of estimating geographical preference.

The average distance between the locations in L and Lu is calculated using Eq. (7). It measures the geographical proximity of location l to the weighted center of user activity.

$$\bar{d}_u = \frac{\sum_{li\ \epsilon\ L}\sum_{lw\ \epsilon\ lu} Dis_{(li,lw)}}{|L|} \quad (7)$$

The average distance between candidate location l and weighted locations that u has visited is calculated by $Dis_{(l_i,l_w)}$, and the geographical preference of user u to location l is calculating using Eq. (8) (Lines 5 in Algorithm 2):

$$GeoP_{(u,l)} = (1 + Dis_{(l_i,l_w)})^{-\delta} \quad (8)$$

where δ is estimated using Eq. (20).

$$\delta = \ln(1 + \bar{d}_u)^{-1} \quad (9)$$

- *User preference time:*

Given user's different check-ins at various venues during the day, so it's important to take check-ins time as a criterion to extract similar users, Therefore, the time dimension is divided following the law of most people's lifetime and work. The distribution of periods is shown in Fig. 5. Hence a matrix with 7 columns is created based on several periods, the numbers of row depend on users' number and each cell it represents the number of weights of a user in this period. Eqs. (10) and (11).

$$S_{(u,r)} = \frac{F_{ru}}{\sum_{i=1} F_u} \quad (10)$$

"Where *S(u,r)* the score for user *u* in period *r*, F_{ru} is the frequency of user *u* in this period *r*, and F_u is present the total user *u* check-ins over all periods".

Then calculate the weight of the timing score W(u,t) by next. Eq. (11)

$$W_{(u,t)} = \frac{S_{(u,r)}}{\sum_{i=1} S_r} \quad (11)$$

"Where *Sr* is the total scoring in period *r"*.

Stage 2: Constructing user community:

Clustering is the unsupervised approach and it's a method of grouping similar data [27]. The construction of user's communities depends on their preferences. It is based on geographical and categorical aspects, and timing preference. Furthermore, each user is present as a vector. The FCM algorithm is applied to cluster users into the K-community by using cosine similarity measure to compute the similarity between each user and other users from geographical, categorical, and user-time matrices output, the FCM algorithm can extract the best result for overlapped data set and comparatively better than k-means algorithm. it more than focused on the grouped node and one user on any community can be founded in other communities but with

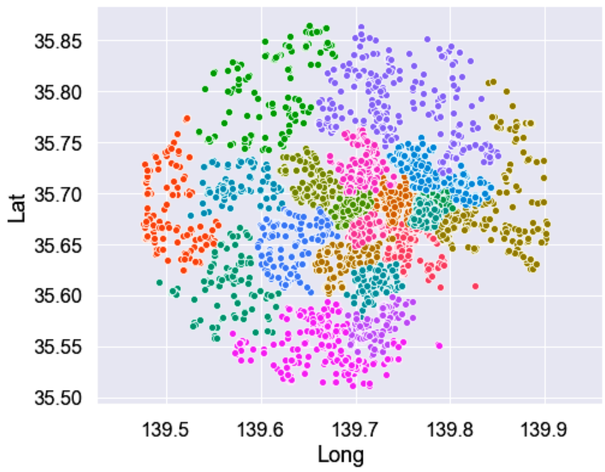


Figure. 5 Geo-community

different value, these communities are named overlapping communities. Complexity and run time of system one of major criteria and give it important so that we compare the run time between FCM and KM on both datasets to approve CMPR it can process and handle a large amount of data without taking the complex time.

Stage 3: Spatial community model

Spatial data mining is a process of discovering potentially useful spatial interests. There are many techniques that can be applied to mining spatial data, but more recent work used clustering techniques [28].

The spatial clustering method is utilizing (X, Y) coordinates to construct a geo-community and convert a city from one large community to different small communities, which increases real-time location and facilities to dedicate location performance and efficiency to generate candidate location that aid to increase recommendation process see Fig. 5.

3.3 Community based personalized location recommendation (on-line recommendation).

Community based personalized location recommendation (on-line recommendation) divide into three stages

Stage1 Candidate locations

Usually, users prefer to visit locations with high check-ins frequency because these locations provide better services and with high popularity. The popularity of locations can be estimate by the number of users visited these locations. Score popularity of location follows the power-law distribution, also

Algorithm 3: The pseudo-code of calculating candidate location in geo-community.

Input: The set of locations in the geo-community that *u* visited *L*
The set of check-ins for all locations in *L*
The recommended period *r*
Begin
For each *li* in *L* **do**
1: Calculate the weighted location frequency by Eq. (12 and 13).
2: **end**
3: Obtain the location popularity (*Popl*) of all candidate locations by Eq. (14).
4: End

users' check-in possibility should rise with this score, so we calculate popularity preference for geo-community was requested and for the user every time a recommendation is to be made.

First, calculate the weighted location frequency for each location in each geo-community like the two Eqs (10) and (11) in user preference time, but we replace the user u and calculate for locations L Eqs. (12) and (13). (Lines 1–3 in Algorithm 3).

$$S_{(l,r)} = \frac{F_{rl}}{\sum_{i=1} F_l} \tag{12}$$

"Where *S(l,r)* the score for location l in period *r*, F_{rl} is the frequency of location *l* in this period r and F_l is present the total location *l* check-ins over all periods". Then calculate the weight of the timing W(l,t) score by Eq. (13)

$$W_{(l,t)} = \frac{S_{(l,r)}}{\sum_{i=1} S_r} \tag{13}$$

Algorithm 3 shows the pseudo-code of estimating geographical preference.

Then the location popularity is calculated by Eq. (14) (Line 3 in Algorithm 3)

$$Pop_l = \int_0^n f_{pop}(\mathcal{Z})d\mathcal{Z} = 1 - (1 + \mathcal{S})^{1-\gamma} \tag{14}$$

where fpop(x) is used to estimate the distribution of location popularity using Eq. (15), and γ is estimated by Eq. (16):

$$f_{pop}(\mathrm{x}) = (\gamma - 1)(1 + x)^{\gamma}\,, x \geq 0, \gamma > 1 \quad (15)$$

$$\gamma = 1 + \ln(1 + W)^{-1} \quad (16)$$

Stage 2 User Similarity computing:

FCM extract the correlation (corr) score between user based these factors, which will be used for calculation users similarity in online modelling, also extracts overlapping users communities with membership score for users overlapping more one communities, otherwise by using the Eq. (18) find the most relevant member on the community for this user after submitting a request of recommendation.

$$SP_{(u,c)} = -\sum_{c \epsilon C} (u.pc) * \log u.p(c) \quad (17)$$

"Where *SP (u, c)* is the scoring probability user u in community c, $\sum_{c\epsilon C}(u.pc)$ is the summation probability that u visited other communities and $u.p(c)$ is the probability that u visited community c"

$$Sim_{(u,u')} = 1 + \log((\frac{corr_{(u,u')}}{1 + |SP_{(u,c)} - SP_{(u',c)}|}) .F_{(u,c)} \quad (18)$$

"Where F is the fuzzy membership of user u in community c"

Stage 3: Recommending locations:

This stage using the previous result from candidate locations and user similarity computing and create a user-location matrix, the value m[i][j] of this matrix is the frequency value. Finally running KNN- classifier to classify the users using user-location matrix. KNN works by finding the distances between a user and all the similar users, then selecting the specified number N closest to the user, then recommends the top N-locations.

4. Experimental results and discussion

4.1 Datasets description

The proposed framework is tested using two publicly available datasets. The first one New york city (NYC) within Gowalla check-ins and the second one Tokyo city (TKY) within Foursquare check-ins [1]. We delete the POIs was checked-in by less than 5 users and user he has less than 5 POIs in his record. Table 1 shown the basic statistics of the two datasets after cleaning.

Table 1: Basic Statistics of datasets

Dataset	#Users	#POIs	#Check-in
Foursquare	1821	6402	8774
Gowalla	917	3610	4598

4.2 Experiments comparative methods

In the experiments, the proposed CMPR is compared with several state-of-the-art algorithms for geographical recommendation. They are listed:

- **MGMMF**: This is the state-of-the-art method is a multi-center Gaussian model fused with matrix factorization, considering incorporating multi-center geographical influence and social influence into the fused framework. [31].
- **WRMF:** Extract a scalable optimization procedure to obtain vector representations for POIs and users, with based WMF that is state-of-the-art recommendation method [29].
- **BPR-MF:** This is the state-of-the-art method to deal, combines the popular ranking-based MF in the learning process with bayesian personalized ranking criterion [30]
- **GeoMF:** Improve recommendation performance, by combining the spatial clustering phenomenon into MF [32].
- **Rank-GeoFM:** Preference of users is integrating with geographical influence to obtain the latent representations for users and POIs in a weighted scheme [33].
- **ARMF:** Exploits geographical and categorical correlations between POIs and users are also integrating in the recommendation process [34].
- **RecNet:** Therefor using the deep neural network to extract joint influence on user behavior. RecNet invests both content and collaborative information [15].
- **L-WMF:** Further improves the WMF method by integrating location relationships as regularization terms [35].
- **LGLMF:** Fused geographical model into the logistic matrix factorization approach [36].
- **APRA:** Using tensor factorization-based approach and a Voronoi diagram-based approach to model the impact of temporal and spatial features on users' preference [37].
- **FCCF:** Clustering method is utilized to place the user to appropriate similar user clusters considering the geographical feature and time feature of user's the check-ins [38]

Received: November 11, 2021. Revised: March 22, 2022.

Table 2. Performance comparison between CMPR and method on the Gowalla and Foursquare datasets

	Gowalla					
	Precision		Recall		F-Measure	
CMPR	Top10	Top20	Top10	Top20	Top10	Top20
MGMMF	0.084	0.077	0.121	0.16	0.099	0.104
BPRMF	0.033	0.026	0.0991	0.148	0.050	0.044
WRMF	0.049	0.04	0.088	0.131	0.063	0.061
GEOMF	0.05	0.039	0.09	0.129	0.064	0.060
RankGeoFM	0.058	0.04	0.101	0.139	0.074	0.062
ARMF	0.069	0.05	0.11	0.15	0.085	0.075
RecNet	0.06	0.045	0.099	0.14	0.075	0.068
L-WMF	0.075	0.055	0.119	0.155	0.092	0.081
LGLMF	0.04	0.031	0.05	0.09	0.044	0.046
APRA	0.018	0.015	0.025	0.04	0.021	0.022
FCCF	0.025	0.021	0.035	0.049	0.029	0.029
CMPR	0.013	0.009	0.09	0.13	0.023	0.017
	Foursquare					
	Precision		Recall		F-Measure	
CMPR	Top10	Top20	Top10	Top20	Top10	Top20
MGMMF	0.087	0.079	0.099	0.15	0.093	0.103
BPRMF	0.031	0.024	0.061	0.093	0.041	0.038
WRMF	0.058	0.043	0.057	0.085	0.057	0.057
GEOMF	0.06	0.049	0.062	0.095	0.061	0.065
RankGeoFM	0.069	0.052	0.073	0.109	0.071	0.070
ARMF	0.064	0.047	0.119	0.15	0.083	0.072
RecNet	0.075	0.055	0.071	0.105	0.073	0.072
L-WMF	0.083	0.071	0.095	0.14	0.089	0.094
LGLMF	0.037	0.031	0.062	0.087	0.046	0.046
APRA	0.027	0.02	0.02	0.032	0.023	0.025
FCCF	0.033	0.027	0.03	0.04	0.031	0.032
CMPR	0.013	0.01	0.085	0.14	0.023	0.019

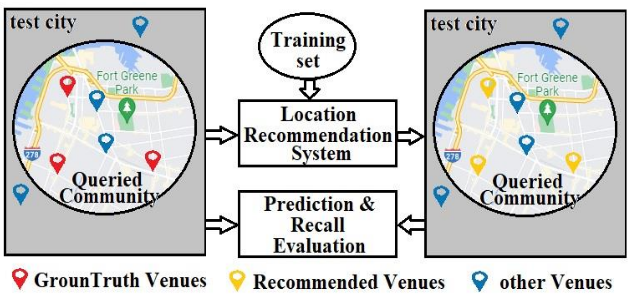


Figure. 6 minimum bounding boxes

4.3 Evaluation metrics

We use the minimum bounding box of locations to model the user-community that evaluates the performance of our proposed framework. The training set is obtained by randomly selection of 30% of each user's visited POIs. The remaining 70% of user's visited POIs are taken as the test set for evaluation.

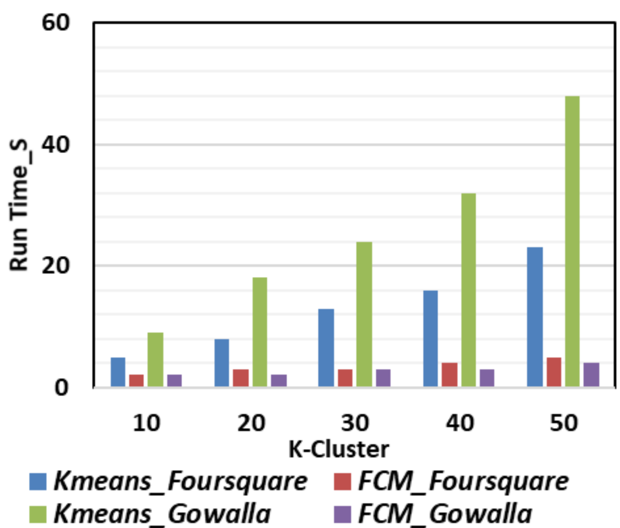


Figure. 7 Comparative FCM, KM results on both datasets

 DOI: 10.22266/ijies2022.0630.36

The test set is represented by black filled circles and recommendation community represented by dotted line is defined as minimum boundary box;

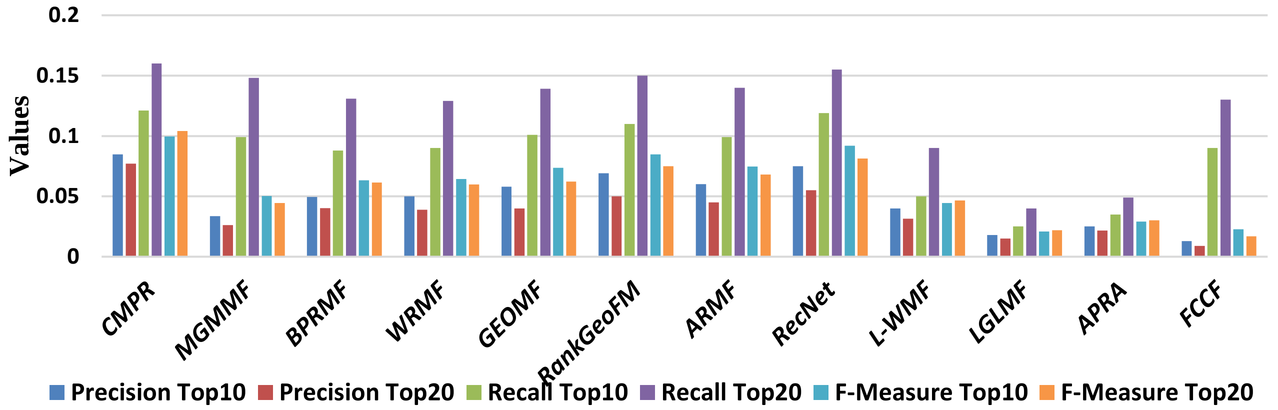


Figure. 8 Comparative results with other traditional recommendation approaches in Gowalla dataset

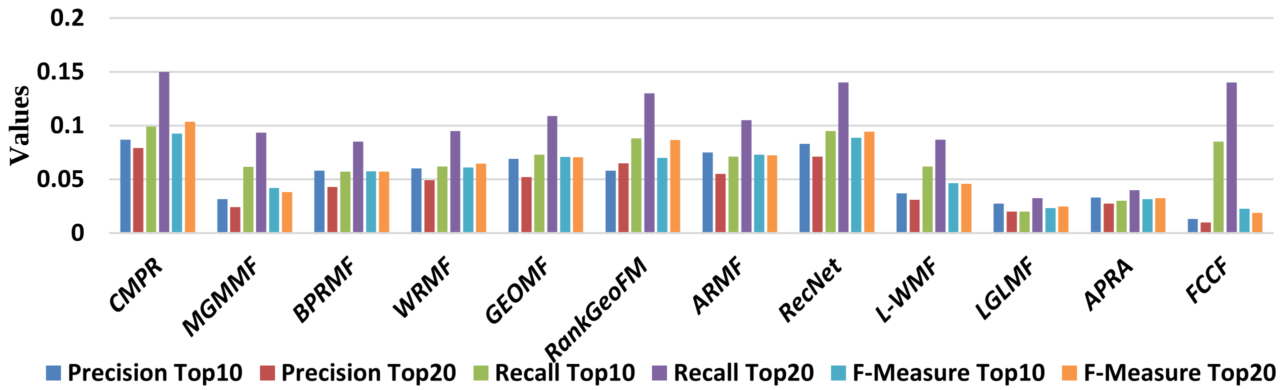


Figure. 9 Comparative results with other traditional recommendation approaches in Foursquare dataset

recommended locations is represented by slash circles [35]. We compute recommendation algorithms performance by averaging above metrics to all test users over all the communities.

4.4 Recommendation performance

Tables 2 illustrate the performance of CMPR in comparison to the existing state-of-the-art POI recommendation models for top-K POI recommendation on Gowalla, Foursquare.

In this subsection, first discuss the performance between two algorithms on both dataset FCM, KM due process user community cluster, with running two algorithms in both dataset and increase number of K-cluster from 10 to 50 Fig. 7 shows that with the increase of training dataset, the execution time of FCM algorithm increases slowly, but the execution

time of KM algorithm records higher increases, also the execution time of FCM algorithm is less than KM algorithm. Secondly, discuss the CMPR performance and other comparative methods. Figs. 8 and 9 shows the comparative results for the performance of all algorithms for the case of N=10 and N=20. Sum of points are extracted from the results: (1) general MF-based models, such as WRMF and BPRMF, achieve poor performance on two datasets, because they only simply factorize the user–POI matrix and ignore the context information, e.g.,

Categorical preference and geographical constraints. Meanwhile, simply incorporating geographical clustering phenomena of check-ins (e.g., MGMMF) does not perform well, since it fails to overlook the fine-grained POI-level context. In contrast, geographical MF-based implicit ranking methods, such as GeoMF and RankGeoFM, perform relatively well, which indicates that modeling user check-ins as implicit feedback is more appropriate in POI recommendation and that geographical influence is the most important factor for POI recommendation for POI recommendation (2) Among the deep recommendation models, RecNet does exhibit the performance as expected, because it only learns the shallow embedding of users and POIs, while the collaborative filtering signals are not fully exploited, although RecNet can incorporate various features for POI recommendation but still prove less effective than CMPR. (3) ARMF employs both information of categorical and geographical in LBSNs but didn't used for grouping similar users with value that facilitate recommendation. (4) FCCF based fuzzy clustering method with extract timing feature without respect users' preference like CMPR, otherwise CMPR construct geo-communities that increase recommendation performance. (5) Both LWMF and LGLMF, they only introduce factorize the user–POI matrix, with geographical characteristics, but they don't consider context information or other features to model users' check-in behavior, they are less efficient than CMPR. (6) on both datasets CMPR outcompetes all comparative methods, showing the advantage of overlapping community structure to combine features in LBSNs and consider the user behavior influence.

5. Conclusion

Solving the overload information problem and increasing the intelligibility of locations represent the main goal of users' location recommendation systems. Such systems deal with different types of information that have been exploited from meager check-in records. In this paper, we propose a CMPR framework that overcomes the limitations faced by other existing personalized location recommendation systems. The proposed CMPR framework integrates categorical, geographical, and timing preferences for constructing an overlapping users' community and geo-community location which provide most the interesting location based on their personal preferences. Hence, the CMPR framework can identify similar users and extract candidate location with higher popularity to improve recommendation performance. Finally, it allows prediction of top candidate N-locations to requested user. The experimental results show significant improvement on Foursquare and Gowalla datasets compared with other algorithms, approved that the proposed framework outcompetes other traditional recommendation approaches in terms of quality, efficiency, and performance for the recommendation process and achieve precision 87% in Foursquare dataset, 84% in Gowalla dataset where N =10 and achieve precision 79% in Foursquare dataset, 77% in Gowalla dataset where N =20. The power of CMPR is construction overlapping communities which integrated geographical, categorical, and timing user preference however used membership of users with popularity of candidate locations based Geo-communities while comparative methods based on traditional collaborative filtering not able to combine these features effectively, CMPR can learn high-order interactions of features by spatial data mining and create user community preference.

Appendix A:

Table A1. List of notations.

Symbol	Descriptions
IC(c)	information category values of category c
P(c)	probability category c appears in the category's sets
{u.li : li.c = c}	number of times that user u visits category c
u.L	the size of the u's location histories
{uj : c ∈ uj.C}	the number of users who have visited c
|U|	the number of users
P	probability that U appears in the category
S(u,c)	score for user u in category c
W(u,c)	weighted categorical preference
φ_w	weighted latitude
λ_w	weighted longitude
T	number of times the user was checked-in this venue
Dis (li,lw)	The distance between Li, Lw
r	the radius of the sphere
φ_i	are the latitudes of li and lw, respectively
φ_w	
λi	are the longitudes of li and lw respectively
λw	
$\bar{d}_u$	average distance between the locations in L and Lu
$(GeoP_{(u,l)})$	geographical preference of user u to location l
S(u,r)	the score for user u in period r
F_{ru}	the frequency of user u in this period r
F_u	total user u check-ins over all periods
W(u,t)	weight the timing of user
S_r	total scoring in period
S(l,r)	score for location l in period r
F_{rl}	frequency of location l in this period r
F_l	total location l check-ins over all periods
W(l,t)	weight the timing of location
Popl	location popularity
fpop(x)	estimate the distribution of location popularity
SP(u,c)	scoring probability user u in community c
$\sum_{c \epsilon C}(u.pc)$	summation probability that u visited other communities
$u.p(c)$	probability that u visited community c
$Sim_{(u,u')}$	similarity between two users

$\boldsymbol{corr}_{(\boldsymbol{u},\boldsymbol{u}')}$	correlation between two users
$\boldsymbol{F}_{(\boldsymbol{u},\boldsymbol{c})}$	fuzzy membership of user u in community c

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

Conceptualization, Khaled Soliman. and Mahmood Mahmood; methodology, Khaled Soliman, Mahmood Mahmood, and Hesham Hefny; software, Khaled Soliman, and Mahmood Mahmood; validation, Khaled Soliman and Hesham Hefny; formal analysis, Khaled Soliman, and Nagy Darwesh; writing—original draft preparation, Khaled Soliman; writing— review and editing Khaled Soliman, Mahmood Mahmood, Hesham Hefny, And Nagy Darwesh; visualization, Khaled Soliman.

References

[1] D. Yang, D. Zhang, V. Zheng, and Z. Yu. "Modeling user activity preference by leveraging user spatial temporal characteristics in LBSNs", *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 45, Vol. 1, pp. 129-142, 2014.

[2] K. Soliman, M. Mahmood, A. E. Azab, and H. Hefny. "A survey of recommender systems and geographical recommendation techniques", *GIS Applications in the Tourism and Hospitality Industry*, IGI Global, pp. 249-274, 2018.

[3] J. V. Rebaza, M. Roche, P. Poncelet, and A. A. Lopes. "The role of location and social strength for friendship prediction in location-based social networks", *Information Processing & Management*, Vol. 54, No. 4, pp. 475-489, 2018.

[4] W. Zhou and W. Han. "Personalized recommendation via user preference matching", *Information Processing & Management*, Vol. 56, No. 3, pp. 955-968, 2019.

[5] O. Kaššák, M. Kompan, and M. Bieliková, "Personalized hybrid recommendation for group of users: Top-N multimedia recommender", *Information Processing & Management*, Vol. 52, No. 3, pp. 459-477, 2016.

[6] M. G. Ostad, H. V. Nejad, and M. A. Nezhad, "Detecting overlapping communities in LBSNs by fuzzy subtractive clustering", *Social Network Analysis and Mining*, Vol. 8, No. 1, pp. 1-11, 2018.

[7] D. Lee, and P. Brusilovsky, "Improving personalized recommendations using community membership information", *Information Processing & Management*, Vol. 53, No. 5, pp. 1201-1214, 2017.

[8] J. Liu, Z. Li, J. Tang, Y. Jiang, and H. Lu, "Personalized geo-specific tag recommendation for photos on social websites", *IEEE Transactions on Multimedia*, Vol. 16, No. 3, pp. 588–600, 2014.

[9] M. Celik, and A. Dokuz, "Discovering socially similar users in social media datasets based on their socially important locations", *Information Processing and Management*, Vol. 54, No. 6, pp. 1154–1168, 2018.

[10] X. Mingjun, and W. Lijun. "Using multi-features to partition users for friends recommendation in location based social network", *Information Processing & Management*, Vol. 57, No. 1, pp. 102–125, 2020.

[11] A. Majid, L. Chen, G. Chen, H. Mirza, I. Hussain, and J. Woodward, "A context-aware personalized travel recommendation system based on geotagged social media data mining", *International Journal of Geographical Information Science*, Vol. 27, No. 4, pp. 662–684, 2013.

[12] J. Ying, H. Chen, K. Lin, E. Lu, V. Tseng, and H. Tsai, "Semantic trajectory-based high utility item recommendation system", *Expert Systems with Applications*, Vol. 41, No. 10, pp. 4762–4776, 2014.

[13] L. Spinsanti, F. Celli, and C. Renso, "Where you stop is who you are: Understanding people's activities", In: *Proc. of the 5Th WOrkshop on Behaviour Monitoring and INTERpretation User Modelling*, 2010.

[14] J. Bao, Y. Zheng, and M. Mokbel, "Location-based and preference-aware recommendation using sparse geo-social networking data", In: *Proc. of the 20th International Conference on Advances in Geographic Information Systems*, pp. 199–208, 2012.

[15] 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*, Vol. 32, No. 8, pp. 1631–1648, 2018.

[16] M. Rahman, W. Liu, S. Suhaim, S. Thirumuruganathan, N. Zhang, and G. Das, "Density based clustering over location based services", In: *Proc. of IEEE 33rd International Conference on Data Engineering (ICDE)*, pp. 461-472, 2017.

[17] Y. Si, F. Zhang, and W. Liu. "An adaptive point-of-interest recommendation method for location-based social networks based on user

activity and spatial features", *Knowledge-Based Systems*, Vol. 163, pp. 267-282, 2019.

[18] Y. Hsueh, and H. Huang. "The personalized itinerary recommendation with time constraints using GPS datasets", *Knowledge and Information Systems*, Vol. 60, No. 1, pp. 523-544, 2019.

[19] G. Lye, W. Cheng, T. Tan, C. Hung, and Y. Chen, "Creating Personalized Recommendations in a Smart Community by Performing User Trajectory Analysis through Social Internet of Things Deployment", *Sensors*, Vol. 20, No. 7, p. 2098, 2020.

[20] M. Chen, W. Li, S. Lu, and D. Chen, "Location-Interest-Aware Community Detection for Mobile Social Networks Based on Auto Encoder", In: *Proc. of International Conference on Knowledge Science, Engineering and Management Springer, Cham*, pp. 176-184, 2019.

[21] F. Yu, Z. Li, S. Jiang, and X. Yang. "Personalized POI Groups Recommendation in Location-Based Social Networks", In: *Proc. of Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint Conference on Web and Big Data*, *Springer*, *Cham*, pp. 114-123, 2017.

[22] M. Li, L. Huang, and C. Wang, "Geographical and overlapping community modeling based on business circles for POI recommendation", In: *Proc. of International Conference on Intelligent Science and Big Data Engineering, Springer*, *Cham*, pp. 665-675, 2017.

[23] W. Cai, Y. Wang, R. Lv, and Q. Jin, "An efficient location recommendation scheme based on clustering and data fusion", *Computers & Electrical Engineering*, Vol. 77, pp. 289-299, 2019.

[24] Z. Zhang, C. Zou, R. Ding, and Z. Chen, "VCG: Exploiting visual contents and geographical influence for Point-of-Interest recommendation", *Neurocomputing*, Vol. 357, pp. 53-65, 2019.

[25] Y. Ma, J. Mao, Z. Ba, and G. Li, "Location recommendation by combining geographical, categorical, and social preferences with location popularity", *Information Processing & Management*, Vol. 57, No. 4, p. 102251, 2020.

[26] F. Mehmood, S. Ahmad, and D. Kim, "Design and development of a real-time optimal route recommendation system using big data for tourists in jeju island", *Electronics*, Vol. 8, No. 5, p. 506, 2019.

[27] S. Panda, S. Sahu, P. Jena, and S. Chattopadhyay, "Comparing fuzzy-C means and K-means clustering techniques: a comprehensive study", In: *Proc. of Advances in Computer Science, Engineering & Applications, Springer, Berlin, Heidelberg*, pp. 451-460, 2012.

[28] M. Parimala, D. Lopez, and N. Senthilkumar, "A survey on density-based clustering algorithms for mining large spatial databases", *International Journal of Advanced Science and Technology*, Vol. 31, No. 1, pp. 59-66, 2011.

[29] Y. Hu, Y. Koren, and C. Volinsky, "Collaborative filtering for implicit feedback datasets", In: *Proc. of 2008 Eighth IEEE International Conference on Data Mining Ieee*, pp. 263-272, 2008.

[30] S. Rendle, C. Freudenthaler, Z. Gantner, and L. S. Thieme, "BPR: Bayesian personalized ranking from implicit feedback", In: *Proc. of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence AUAI Press*, pp. 452-461, 2009.

[31] C. Cheng, H. Yang, I. King, and M. Lyu, "Fused matrix factorization with geographical and social influence in location-based social networks", In: *Proc. of the AAAI Conference on Artificial Intelligence*, Vol. 26, No. 1, 2012.

[32] D. Lian, C. Zhao, X. Xie, G. Sun, E. Chen, and Y. Rui, "GeoMF: joint geographical modeling and matrix factorization for point-of-interest recommendation", In: *Proc. of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 831-840, 2014.

[33] X. Li, G. Cong, X. Li, T. Pham, and S. Krishnaswamy, "Rank-geofm: A ranking based geographical factorization method for point of interest recommendation", In: *Proc. of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 433-442, 2015.

[34] H. Li, Y. Ge, R. Hong, and H. Zhu. "Point-of-interest recommendations: Learning potential check-ins from friends", In: *Proc. of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 975-984, 2016.

[35] L. Guo, Y. Wen, and F. Liu, "Location perspective-based neighborhood-aware POI recommendation in location-based social networks", *Soft Computing*, Vol. 23, No. 22, pp. 11953-11945, 2019.

[36] M. Baratchi, M. Afsharchi, and F. Crestani, "LGLMF: Local Geographical Based Logistic Matrix Factorization Model for POI Recommendation", In*: Proc. of Information Retrieval Technology: 15th Asia Information Retrieval Societies Conference, AIRS, Springer*

Nature, Hong Kong, China, Vol. 12004, p. 66, 2020.

[37] M. Li, W. Zheng, Y. Xiao, and X. Jiao. "An Adaptive POI Recommendation Algorithm by Integrating User's Temporal and Spatial Features in LBSNs", In: *Proc. of the 3rd International Conference on Data Science and Information Technology*, pp. 135-139, 2020.

[38] M. Yin, Y. Liu, X. Zhou, and G. Sun, "A Fuzzy Clustering Based Collaborative Filtering Algorithm for Time-aware POI Recommendation", *Journal of Physics: Conference Series, IOP Publishing*, Vol. 1746, No. 1, p. 012037, 2021.