International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. , No. (Mon-20..)

http://dx.doi.org/10.12785/ijcds/XXXXXX

Toward a New Similarity Measure Based on Combining Tourist Check-ins and Their Trip Path for a Point-Of-Interest Recommendations in a LBSN

Djelloul BETTACHE[1,2] and Nassim DENNOUNI[3,4]

[1] *LME Laboratory, UHBC, Chlef*
[2] *Computer Science Department, Hassiba Benbouali University, Chlef, Algeria*
[3] *Higher School of Management, Tlemcen*
[4] *LIA Laboratory, UHBC, Chlef*

Received Mon. 20, Revised Mon. 20, Accepted Mon. 20, Published Mon. 20

Abstract: In recent years, tourists have tended to share their travel experiences with friends through location-based social networks (LBSNs). However, these networks accumulate large masses of data, making them ineffective in guiding individual tourists through their journeys. To overcome this drawback, point-of-interest (POI) recommender systems (RS) can provide a beneficial solution by exploiting the potential of LBSNs to suggest places they have never visited to new tourists. These systems can be classified into two categories: the first uses memory-based algorithms, while the second employs algorithms based on machine learning models. Collaborative filtering (CF) is a popular memory-based smart tourism approach commonly used in literature. This approach predicts the probability of POI check-ins by new tourists based on their similarities with other tourists, using measures such as Cosine, Jaccard, Pearson correlation, and Euclidean distance. However, to our knowledge, no formal framework takes POI check-ins and visit paths into account when calculating similarities between tourists. For this reason, in this paper, we propose a novel measure called SPPUR (Similarity of Paths and the Proximity of Users for Recommending POIs) inspired by the term frequency-inverse document Frequency (TF-IDF) method, which uses POI frequentation and geographical proximity between users to calculate similarities that can predict POIs to be visited by new tourists. Our experimental results on Foursquare show that compared with other state-of-the-art measures, this similarity measure significantly improves SR performance regarding PRECISION, RECALL, MAP, and NDCG.

Keywords: LBSN, POI recommendation system, Collaborative filtering, Similarity measures, TF-IDF, Check-ins

1. INTRODUCTION

Today, smartphone users are often interested in diverse areas such as course recommendations [1], self-driving technology [2], car parking detection systems [3], and connecting with friends on location-based social networks (LBSNs) [4]. For these raisons, the rapid urbanization of cities has significantly increased the number of points of interest (POIs) appealing to tourists, including restaurants, hospitals, hotels, museums, and so on. On the other hand, the intensive use of LBSNs like Geolife, Facebook, Gowalla, and Foursquare helps identify their users's behavior and collect their current and future needs. In this context, POI recommender systems (RS) are used to analyze LBSN user behavior and provide personalized services to new tourists during their travels. Collaborative filtering is one of the memory-based methods that provides systematic recommendations that are close to the current user's context [5]. This method first computes similarities between users and then selects only the active user's neighbors to generate the predictions for recommending POIs to visit. In contrast, the model-based method first builds a model that describes the user's behavior and then predicts the POI evaluation. This article uses user-based CF for POI recommendations because of its simplicity and effectiveness. This principle is based on two essential phases : (1) the selection of users most similar to the active user and (2) the prediction of the POIs from the most similar users. Similarity measures can significantly improve the accuracy of prediction algorithms. Several popular similarity measures using the CF principle have been used in POI recommendation algorithms. For this reason, selecting an appropriate similarity measure is considered a crucial issue when implementing an RS, as it can significantly impact its performance [6]. Several similarity methods between users can be used in the literature for POI recommendation in an LBSN. However, only some of these methods incorporate a combination of dimensions such as path similarity and user proximity. However, existing POI recommendation techniques often have problems such as

E-mail address: d.bettache@univ-chlef.dz, n.dennouni@univ-chlef.dz

data sparsity and cold start, which can reduce the quality and accuracy of predictions. For these reasons, we have proposed a novel similarity measure based on the TF-IDF [7] technique adapted to a POI recommendation context in this paper, considering users' proximity to their departure and arrival check-ins. This measure can be used with CF based on the active user to provide accurate recommendations and reduce the problems of cold start and data sparsity. Our findings from experiments on a real dataset obtained from Foursquare show that our proposed method presents a significant performance improvement over other similarity measures. The contributions of this paper can be summarized as follows:

- Proposal of a novel similarity measure called SPPUR (Similarity of Paths and the Proximity of Users for Recommending POIs) inspired by the TF-IDF technique, to be integrated into the POI recommendation process based on collaborative filtering centered on the active user.

- Implement a framework that uses this SPPUR similarity to generate user visit predictions. - Experimentation of the SPPUR model with the Foursquare dataset to evaluate its performance using parameters including PRECISION, RECALL, MAP, and NDCG.

The other sections of the paper are as follows. Section 2 presents a state-of-the-art on current POI recommendation approaches based mainly on POI frequentation in an LBSN, explaining the motivations for this research. Next, section 3 details the mathematical formulas for calculating SPPUR similarity and the algorithm for implementing it to recommend POIs. In Section 4, we describe the design of the SPPUR model and how its main components work. Before concluding, the experimental results are analyzed and discussed in Section 5, comparing the SPPUR model with other existing models. Finally, the last section summarizes the paper's contributions and suggests future perspectives.

2. Literature Review

In a smart tourism context, user profiles refer to a set of characteristics such as travel preferences (preferred destination, travel time, preferred season, available budget, etc.), interests (hiking, museums, shopping, food, etc.), travel history (places previously visited, activities performed, accommodations frequented, means of transport used, etc.), etc. These profiles are generally based on data (ratings and check-ins) collected via digital technologies such as mobile applications, websites, LBSNs, IoT (Internet of Things) devices, etc. [8]. Next, collaborative filtering (CF) techniques will use the preferences and behaviors of similar users to recommend POIs. In this context, Ye et al. developed a geo-social CF model that combines geographical and social influences to improve POI recommendations. This technique combined the geographical proximity of POIs with user check-ins and improved the accuracy of recommendations compared with traditional CF models [9].

Cheng et al., on the other hand, proposed a matrix factorization method that uses spatial distance functions to capture the user preferences required to model geographical influences. This method achieves better recommendation accuracy than the basic CF approaches [10]. In the same context, Lian et al. introduced GeoMF, which combines matrix factorization with a geographic model. This model uses geographic distance and check-in information to improve matrix factorization's recommendation accuracy and outperform several other geographic models [11]. Finally, Zhang et al. developed the Geosoca model, which exploits geographical and social relationships by integrating the proximity of POIs and their categories into the POI recommendation process. This model provides better recommendation performance than geographic CF models [12].

In contrast to previous CF studies, which used only geographical influences (see Table I) to recommend POIs, Liu et al. proposed a check-in-based recommendation model that incorporates contextual factors such as time of day and user category. This model uses a machine learning algorithm to model user preferences using contextual data, significantly improving recommendation accuracy over traditional models [13].

In the same context, Zhao et al. have developed a Geo-Teaser, a model that combines geo-temporal sequential embedding and ranking for POI recommendations. This model integrates geo-temporal sequences of check-ins to capture user preferences and uses embedding techniques to improve recommendation accuracy over methods based solely on geography or time [14]. In the same way, Wang et al. introduced ST-RNN, a spatial-temporal recurrent neural network, to solve the problem of missing check-ins. This model uses RNN to deduce the spatial and temporal dependencies of check-ins, helping to solve the problem of missing check-ins and improving POI recommendations [15]. In contrast to this work, Lian et al. aims to eliminate the dependency between social and contextual data, focusing solely on check-ins. For this reason, they have developed an RS called de LightRec that uses algorithms optimized for processing large volumes of check-in data to achieve good recommendation accuracy while maintaining low computational complexity [16]. Finally, Zheng et al. exploit data such as check-ins, reviews, and social relationships using graphs to model complex user interactions and improve the accuracy of POI recommendations [17].

The research mentioned in Table II above shows the evolution of POI recommendation techniques based on check-ins, thanks to deep learning and graphs. However, more recent research focuses more on the semantics of geographical correlations and the use of encoders. For example, the Bayes-enhanced Multi-view Attention Networks (BayMAN) model seeks to improve POI recommendations by simultaneously exploring semantic correlations, geographical distance dependencies, and personal view preferences. This model constructs multi-view graphs of POIs to understand the relationships between POIs better, thus addressing the

TABLE I. RS uses only check-ins to deduce geographical influencess

Studies	Approach used	Key features	Results
Ye et al. (2011)	Geo-social CF	Integrating geographical and social proximity	Significant improvement over traditional CF
Cheng et al. (2012)	Matrix factorization	Modeling spatial distances and social influences	Outperforms basic and geographic CFs
Lian et al. (2014)	GeoMF	Combined geographic model with matrix factorization	Highest recommendation accuracy
Zhang et al. (2015)	Geosoca	Use of geographical/social correlations and POI categories	The performance is better compared to individual models

TABLE II. RS integrate check-ins and context to deduce spatio-temporal influences

Studies	Approach used	Key features	Results
Liu et al. (2017)	Modeling user preferences	Integration of time and user category	Improved accuracy of recommendations compared with traditional models
Zhao et al. (2017)	Geo-Teaser	Combining geo-temporal sequential embedding and ranking	Improved accuracy compared with geography- or time-based methods
Wang et al. (2018)	ST-RNN	Inferring the spatial and temporal dependencies of check-ins	Solving the problem of missing check-in data
Lian et al. (2020)	LightRec	Processing large volumes of check-ins	Maintain low computational complexity

problem of unreliable check-in data [18]. On the other hand, the Meta-learning Enhanced POI Recommendation (MERec) model uses two encoders to improve the accuracy of predictions of the next POI to be visited by the user. For this reason, the first category-level encoder captures common user behaviors, while the POI-level encoder learns the precise transition patterns of POIs in the target city [19]. In the same context, Zhang et al. proposed a model called Event-Based Probabilistic Embedding (TARE) to model geographical influences in POI recommendations by capturing users' check-in activities in specific regions. This approach is compared to several state-of-the-art methods, demonstrating its effectiveness in leveraging temporal, geographic, and semantic factors to predict user preferences for POIs [20]. Finally, these advances demonstrate the importance of considering several factors to improve POI recommendation in LBSNs; however, to our knowledge, no approach in the literature uses the combination of POI frequentation and user path similarity. In this article, we concentrate on identifying similarities between user profiles derived from their POIs and the check-in sequences during their visits.

3. Method

This section explains how to calculate SPPUR similarity from users' check-ins and POI paths. To achieve this goal, we first describe the formulas that can be used to calculate predictions from this type of similarity. Then, using an example, we explain how to obtain the SPPUR similarity matrix from the User-POI-check-in matrix. Finally, we propose an algorithm that uses this similarity matrix to generate predictions of future POIs to visit based on user check-ins and user paths.

A. The SPPUR similarity-based recommendation process

In this subsection, we calculate SPPUR similarity using three types of similarity. The first type of similarity considers users' preferences through their check-ins of POIs (frequentations) during the journey. In contrast, the second and third types of similarity are based solely on their departure and arrival check-ins. We then combine these three types of similarity to obtain the SPPUR similarity. Finally, this similarity will be used to calculate the predictions for the POI recommendation process.

1) User/user similarity Based on POIs path

In the following, we are interested in the profiles deduced from tourists' check-ins and POI paths. For this reason, we assume that similarity between users can be deduced from the similarity between their POI paths taken during the journey. Each user profile is characterized by a sequence of character strings containing the order of the POIs they visited. Consequently, the similarity between two users can be calculated by using the TF-IDF [21][22] technique to the smart tourism domain, considering that (1) each term corresponds to a POI and (2) all the POIs visited by a user during a journey corresponds to a document. This adaptation makes it possible to determine the importance of POIs (the terms) in the path of visits made by tourists (the documents). To achieve this, we calculate the frequency of each POI in each user's profile (see Equation(1)), then look at the importance of that POI through the number of visitors it received (see Equation(2)), then deduce the score for each pair (user, POI) using (Equation(3)). Finally, based on this score calculation for each user, we can calculate the similarity between each pair of users using (Equation(4)) below.

- TF Calculation: The term frequency (TF) value for each POI P_i in the user U_a profile is defined as the number of times the U_a has visited the P_i. This value is calculated using Equation(1) below:

$$TF(U_a, P_i) = \frac{Freq(U_a, P_i)}{Visits(U_a)} \tag{1}$$

With :
- $Freq(U_a, P_i)$ = Number of check-ins of U_a on P_i.
- $Visits(U_a)$ = Total check-ins of U_a on all POIs.

- IDF Calculation: In order to measure the importance of a given POI_i, we calculate the inverse document frequency (IDF) of each P_i by the logarithm of the ratio of the number of users who have visited this POI to the total number of users as shown in Equation(2) below:

$$IDF(P_i) = \log \times \frac{NUsers}{P_i(Users)} \tag{2}$$

With :
- $NUsers$ = Total number of Users.
- $P_i(Users)$ = Number of Users who visited P_i.

- TF-IDF Calculation: To determine the importance of a P_i for a Usera in the set of POIs and users, we calculate the TF-IDF value of each pair (U_a, P_i) as shown in Equation(3) below:

$$TFIDF(U_a, P_i) = TF(U_a, P_i) \times IDF(P_i) \tag{3}$$

- Sim_{path} calculation: The $TF\text{-}IDF$ values obtained from Equation(3) can be used to calculate the path similarity values (denoted Sim_{path}) between each pair of users, as shown in Equation(4) below:

$$Sim_{path}(U_a, U_b) = \frac{\sum_{i \in I}(A \times B)}{\sqrt{\sum_{i \in I}(A)^2} \times \sqrt{\sum_{i \in I}(B)^2}} \tag{4}$$

Where:
$A = TFIDF(U_a, P_i)$ and $B = TFIDF(U_b, P_i)$

2) User/User similarity Based on point of departure

We consider that two users U_a and U_b are similar if their last visited POIs are close to each other, so the similarity between these users, noted Simend, can be calculated using the Equation(5) below:

$$Sim_{start}(U_a, U_b) = \frac{1}{1 + dis_{start}(U_a, U_b)} \tag{5}$$

Where: $dis_{start}(U_a, U_b)$ represent the distance between the two users U_a and U_b according to their initial locations (first check-in).

3) User/user similarity Based on arrival point

We also consider that two users U_a and U_b are similar if their last visited POIs are close to each other, so the similarity between these users, noted Simend, can be calculated using Equation(6) below:

$$Sim_{end}(U_a, U_b) = \frac{1}{1 + dis_{end}(U_a, U_b)} \tag{6}$$

Where: $dis_{end}(U_a, U_b)$ represents the distance between the two users U_a and U_b according to their final locations (last check-in).

Note that to calculate the distance between two points of interest (POI), we used the Haversine formula above, which is commonly employed in navigation and geographical information systems [23][24]:

$$dis = 2r \cdot \arcsin \sqrt{\sin^2\left(\frac{\Delta l}{2}\right) + \cos(l_1) \cdot \cos(l_2) \cdot \sin^2\left(\frac{\Delta g}{2}\right)} \tag{7}$$

With:

- dis is the distance between the two points,
- r is the earth's radius (6,371 (km)),
- l_1 and l_2 are the latitudes of the POI,
- g_1 and g_2 are the longitudes of the POI,
- $\Delta l = l_2 - l_1$ is the difference in latitude,
- $\Delta g = g_2 - g_1$ is the difference in longitude.

4) SPPUR similarity formula

To obtain the SPPUR similarity value between each pair of users, we combine the Sim_{path}, Sim_{start} and Sim_{end} similarity values with the parameters α, β and γ, as shown in Equation(8) below:

$$SPPUR) = \alpha(Sim_{path}) + \beta(Sim_{start}) + \gamma(Sim_{end})) \tag{8}$$

Where α, β and $\gamma \in [0, 1]$ and $(\alpha + \beta + \gamma = 1)$ are the adjusting values that control the novel similarity.

5) SPPUR prediction formula

After computing the SPPUR similarities between users using Equation(8), the prediction for a target user is calculated using the following Equation[25].

$$Prediction(U_a, P_i) = \frac{\sum_{b \in U} SPPUR(U_a, U_b) \times f_{b,i}}{\sum SPPUR(U_a, U_b)} \tag{9}$$

With:
-U is the set of all users.

TABLE III. An example of User-POI frequency matrix

	POI_1	POI_2	POI_3	POI_4	POI_5
$user_1$	1	2	0	4	1
$user_2$	3	0	1	0	2
$user_3$	0	0	4	1	6
$user_4$	3	0	1	0	0

TABLE IV. TF score matrix

TF	POI_1	POI_2	POI_3	POI_4	POI_5
$user_1$	0.125	0.25	0	0.5	0.125
$user_2$	0.5	0	0.166	0	0.333
$user_3$	0	0	0.363	0.09	0.545
$user_4$	0.75	0	0.25	0	0

-$SPPUR(U_a, U_b)$ is the final similarity ($SPPUR$) between U_a and U_b.
-$f_{b,i}$ is the the visit frequency of U_b on P_i

B. SPPUR similarity example

The proposed SPPUR similarity can be calculated using the example below (see Table III). This example concerns a set of four users and five POIs. The similarity matrix (Sim_{path}) for this set (see Table VII) is calculated using the TF, IDF and TF-IDF value, respectively (Table IV, Table V and Table VI).

1- Calculation of TF value : First we use Equation(1) to calculate the TF value of each peer (U_a, P_i). Let's take the $user_1$, POI_1 pair as an example. The TF value of this pair is calculated as follows:
$TF(user_1, POI_1) = \frac{1}{1+2+4} = 0.1428$

In the same way, we calculate the rest of the TF values, as shown in Table IV:

2- Calculation of IDF value : To calculate the IDF value of each POI, we use Equation(2). For instance, the IDF value of a POI_1 is calculated as follows:
$IDF = \log(\frac{4}{3}) = 0.415$

The IDF values of the other POIs are calculated in the same way, as shown in Table V:

3- Calculation of TF-IDF value : To calculate the TF-IDF value for each pair (U_a, P_i), we use Equation(3). For example, the TF-IDF value of $(user_1, POI_1)$ is calculated as follows:
$TFIDF(user_1, POI_1) = TF(user_1, POI_1) \times IDF(POI_1)$
$TFIDF(user_1, POI_1) = 0.0592$

The other TF values are calculated in the same way, as shown in Table VI:

TABLE V. IDF score matrix

	POI_1	POI_2	POI_3	POI_4	POI_5
IDF	0.415	2	0.415	1	0.415

TABLE VI. TF-IDF score matrix

TF-IDF	POI_1	POI_2	POI_3	POI_4	POI_5
$user_1$	0.051	0.5	0	0.5	0.051
$user_2$	0.207	0	0.069	0	0.138
$user_3$	0	0	0.150	0.090	0.226
$user_4$	0.311	0	0.103	0	0

TABLE VII. Sim_{path} matrix

Sim_{path}	$user_1$	$user_2$	$user_3$	$user_4$
$user_1$	1	0.0975	0.2805	0.0692
$user_2$	0.0975	1	0.5624	0.8452
$user_3$	0.2805	0.5624	1	0.1664
$user_4$	0.0692	1	0.1664	1

4- Calculation of Sim_{path} value: Finally, to calculate the Sim_{path} value between each pair (U_a, U_b), we use Equation(4). For example, the Sim_{path} value between ($user_1$,$user_2$) is:
$Sim_{path}(user_1, user_2) = 0.0694$

We calculate the other Sim_{path} values as shown in Table VII:

5- Sim_{start} and Sim_{end} calculation:

Considering the scenario described in Figure 1 and Figure 2, $user_1$ and $user_2$ have POI_1/POI_3 as starting point and POI_4/POI_1 as end point, respectively

Based on the POI neighborhood matrix (see Table IX), which describes the different distances between POI pairs calculated using Equation(7), we can deduce the Sim_{start} and Sim_{end} similarities between each user pair using Equation(5) and Equation(6), respectively

For example, the Sim_{start} value between $user_1$ and $user_2$ is calculated as shown below:
$Sim_{start}(user_1, user_2) = \frac{1}{1+dist_{start}(user_1,user_2)}$
$Sim_{start}(user_1, user_2) = \frac{1}{1+dist(POI_1,POI_3)}$

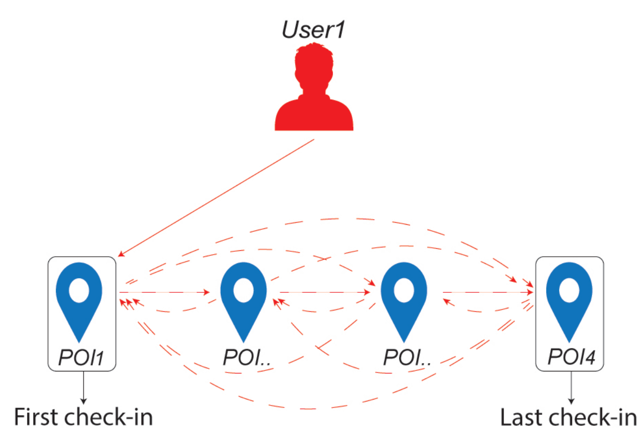

Figure 1. Path taken by user1

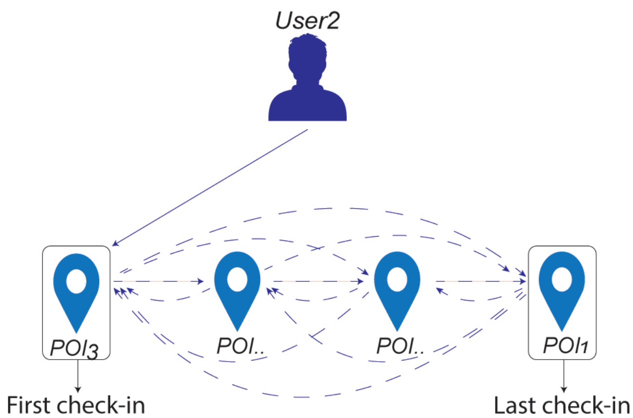

Figure 2. Path taken by user2

TABLE VIII. First and Last POI visited

User	First POI	End POI
$user_1$	POI_1	POI_4
$user_2$	POI_3	POI_1
$user_3$	POI_4	POI_3
$user_4$	POI_1	POI_5

$Sim_{start}(user_1, user_2) = \frac{1}{1+6}$
$Sim_{start}(user_1, user_2) = 0.1428$

Furthermore, the Sim_{end} value between $user_1$ and $user_2$ is calculated as shown below:
$Sim_{end}(user_1, user_2) = \frac{1}{1+dist_{end}(user_1,user_2)}$
$Sim_{end}(user_1, user_2) = \frac{1}{1+dist(POI_4,POI_1)}$
$Sim_{end}(user_1, user_2) = \frac{1}{1+3}$
$Sim_{end}(user_1, user_2) = 0.25$

In the same way, we calculate the rest of Sim_{start} and Sim_{end} values, as shown in Table X and Table XI:

6- Finally, to calculate SPPUR similarity, we use Equation(8). For example, the SPPUR similarity between $user_1$ and $user_2$ is calculated as follows:

TABLE IX. POIs distance

	POI_1	POI_2	POI_3	POI_4	POI_5
POI_1	0	1	6	3	2
POI_2	1	0	2	5	5
POI_3	6	2	0	1	3
POI_4	3	5	1	0	2
POI_5	2	5	3	2	0

TABLE X. Sim_{start} matrix

Sim_{start}	$user_1$	$user_2$	$user_3$	$user_4$
$user_1$	1	0.1428	0.25	1
$user_2$	0.1428	1	0.5	0.1428
$user_3$	0.25	0.5	1	0.25
$user_4$	1	0.1428	0.25	1

TABLE XI. Sim_{end} matrix

Sim_{end}	$user_1$	$user_2$	$user_3$	$user_4$
$user_1$	1	0.25	0.5	0.333
$user_2$	0.25	1	0.1428	0.333
$user_3$	0.5	0.1428	1	0.25
$user_4$	0.333	0.333	0.25	1

TABLE XII. $SPPUR$ similarity matrix

SPPUR	$user_1$	$user_2$	$user_3$	$user_4$
$user_1$	1	0.1469	0.3277	0.3679
$user_2$	0.1469	1	0.4419	0.5416
$user_3$	0.3277	0.4419	1	0.2082
$user_4$	0.3679	0.5416	0.2082	1

$SPPUR(U_1, U_2) = \alpha(Sim_{path}) + \beta(Sim_{start}U_1, U_2)) + \gamma(Sim_{end}U_1, U_2))$
Suppose : $\alpha = 0.5$, $\beta = 0.25$ and $\gamma = 0.25$
$SPPUR(U_1, U_2) = 0.1065$

We also calculate the other $SPPUR$ similarity values as shown in Table XII:

C. the SPPUR similarity-based recommendation algorithm

In this subsection, we propose pseudocode for Algorithm 1 below, which will be used to implement the model named SPPUR (Similarity of Paths and the Proximity of Users for Recommending POIs). This algorithm calculates SPPUR similarities between users and then predicts which POIs to visit, mainly using the user frequentation matrix as input.

4. Proposed model

In this section, we propose the SPPUR (Similarity of Paths and the Proximity of Users for Recommending POIs) model for POI recommendation adapted from the TF-IDF method. The model calculates similarities between POIs and users based on their frequentation and geographical proximity, thereby predicting which POIs new tourists should visit. Below are the main steps of the proposed model:

Step 1: Preprocessing phase: First, from the existing dataset, we construct the User-POI frequency matrix (see Figure 3). Then, we normalize each user's check-in frequency into the range [0, 5]. The process of normalization is described as fellow:

$$P_{Nfreq} = \begin{cases} 5, & \text{if } P_{freq} = Max_{freq}. \\ \frac{5 \times P_{freq}}{Max_{freq}}, & \text{Otherwise} \end{cases} \tag{10}$$

where P_{Nfreq} indicates the normalized frequency value, P_{freq} indicates the real user's check-in number, Max_{freq} indicate the largest frequency of a user.

Step 2: for the target user ($user_a$), we calculate the $TFIDF$ score of all pairs ($user_a, POI_i$), then, we compute

Algorithm 1: SPPUR Algorithm

Input: Users-POIs Frequency Matrix R, Users U, POIs P, First Visited POI FP, Last Visited POI LP, Similar Users N, Recommended POI K, Target User U_a, Setting value α, β, γ

Output: Final Similarity Matrix $SPPUR$, POIS Prediction Matrix PPM, POIs recommended List $ListRecPOI$

Function $Sim_{Start}(U_a, U_b, FP)$:

$distStart(U_a, U_b) = Haversine(FP(U_a), FP(U_b))$;

$Sim_{Start} = \frac{1}{1+distFirst(U_a,U_b)}$;

return Sim_{Start};

FinFunction

Function $Sim_{End}(U_a, U_b, LP)$:

$distEnd(U_a, U_b) = Haversine(LP(U_a), LP(U_b))$;

$Sim_{Start} = \frac{1}{1+distEnd(U_a,U_b)}$;

return Sim_{End};

FinFunction

begin

foreach (U_a) **do**

foreach (U_a, U_b) **do**

$N = 0, D_a = 0, D_b = 0$;

foreach $POI\ p$ **do**

if $(R(U_a, P_p) \neq 0\ Or\ R(U_b, P_p) \neq 0)$ **then**

$D_a = D_a + (TF(a, p) \times IDF(p))^2$;

$D_b = D_b + (TF(b, p) \times IDF(p))^2$;

$N = N + (D_a \times D_b)$;

$Sim_{path}(U_a, U_b) = \frac{N}{\sqrt{(D_a)^2} \times \sqrt{(D_b)^2}}$;

$SPPUR(U_a, U_b) = \alpha \cdot Sim_{path}(U_a, U_b) + \beta \cdot Sim_{Start}(U_a, U_b) + \gamma \cdot Sim_{End}(U_a, U_b)$;

$NSusers = \nabla(SPPUR(U_a, U_b), N)$;

foreach $POI\ p$ **do**

$N = 0, D = 0$;

foreach $U_b \in NSusers$ **do**

$N = N + (SPPUR(U_a, U_b) \times R(U_b, p))$;

$D = D + SPPUR(U_a, U_b)$;

$PPM(U_a, U_b)$

$ListRecPOI = \nabla(PPM, K)$;

Fin

the Sim_{path}, Sim_{start} and Sim_{end} values between target and other users ($user_b$) by using Equation(4).

Step 3: After computing the Sim_{path}, Sim_{start} and Sim_{end}, we use the Equation(8) to compute the final $SPPUR$ similarity between target and other users.

Step 4: Next, a list of N most similar users to the target user is selected.

Step 5: Then a prediction is generated by using the Equation(9).

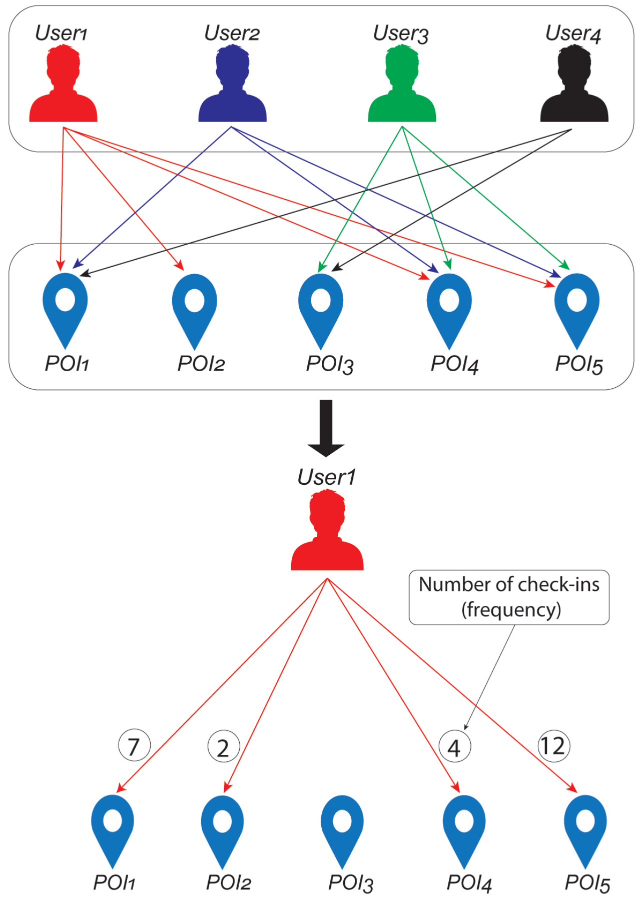

Figure 3. Transition from the user check-in to the POIs frequentation

Step 6: Finally, we recommend a Top K ranked POIs to the target user.

Figure 4 illustrates the main steps of our Framework based on the SPPUR model.

5. Experiments and Results

In this section, we describe the Foursquare dataset used in our experiments. Next, we explain the metrics (PRECISION, RECALL, MAP and NDCG) chosen to evaluate the performance of the SPPUR model. Finally, we define our experimental procedure and the associated hyper-parameters.

A. Data Collection

To implement the SPPUR model, we used the Foursquare data set [26], which contains tourist check-ins associated with POIs (in our work, we studied data from the cities of New York and Tokyo from April 2012 to February 2013). Each row of this dataset contains the user ID, POI ID, POI longitude, POI latitude, POI category and Check-in time. Table XIII describes all the columns in the Foursquare data set used in our experiments.

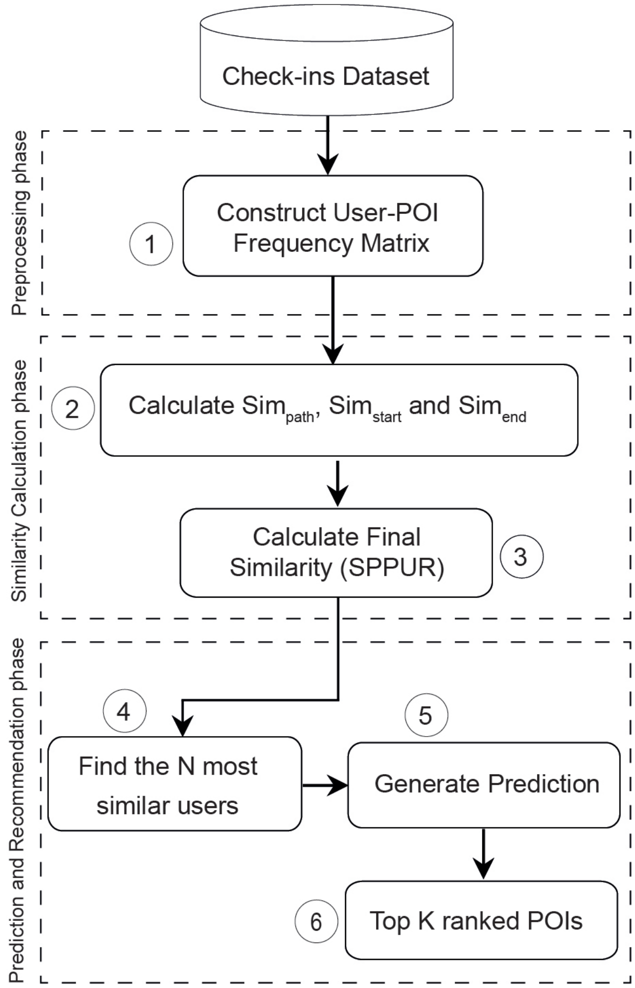

Figure 4. System architecture based on the SPPUR model

TABLE XIII. Dataset used in the experiments

Statistics	New York Dataset	Tokyo Dataset
Users	1083	2293
POIs	38333	61858
Check-ins	227428	573703

B. Evaluation metrics

To compare the SPPUR model with algorithms based on other types of similarity found in the literature, such as Pearson's correlation [27][28], Spearman's correlation [29][30], Euclidean distance [30], cosine [29], adjusted cosine [29], mean square error (MSD) [27] [29] and Jaccard's similarity measure [29][30], we used four evaluation measures: PRECISION, RECALL, MAP and NDCG.

1) PRECISION@K: This measure is calculated using the proportion of relevant results among the first K elements returned by the system, as shown in the following formula [31][32]:

$$PRECISION@K = \frac{1}{K}\sum_{u\in U}\frac{|Rec_u \cap Test_u|}{|Rec_u|} \tag{11}$$

2) RECALL@K: This measure presents the proportion of relevant results among all the relevant elements available in the dataset that are found in the first K elements returned by the system, as indicated in the following formula [33][34]:

$$RECALL@K = \frac{1}{K}\sum_{u\in U}\frac{|Rec_u \cap Test_u|}{|Test_u|} \tag{12}$$

3) MAP@K: (Mean Average PRECISION@K) This parameter measures the average precision of relevant results over several queries, taking into account the first K results returned for each query. For this reason, the average precision is calculated at different points (Average Precision) for each query in order to deduce the average over all queries [34]. In the following, Equation(13) and Equation(14) are used to calculate the $MAP@K$ value:

$$MAP@K = \frac{1}{K}\sum_{j=1}^{M}\frac{1}{r}\sum_{k=1}^{K} PRECISION@k \times rel(k) \tag{13}$$

$$rel(k) = \begin{cases} 0, & \text{if POI at } k_{th} \text{ rank is relevant} \\ 1, & \text{otherwise} \end{cases} \tag{14}$$

4) NDCG@K: (Normalized Discounted Cumulative Gain) This metric evaluates the quality of SRs according to the relevance of the recommended items and their ranking in the results list, assigning greater weight to items at the top of the list [35][36]. We use Equation(15), Equation(16) and Equation(17) to calculate the $NDCG@K$ value:

$$NDCG@K = \frac{DCG@k}{IDCG@k} \tag{15}$$

$$DCG@K = \sum_{i=1}^{k}\frac{rel(i)}{log_2(i+1)} \tag{16}$$

where $IDCG$ is ideal discounted cumulative gain.

$$IDCG@K = \sum_{i=1}^{|REL_p|}\frac{rel(i)}{log_2(i+1)} \tag{17}$$

REL_p represents the list of relevant documents (ranked by relevance) in the corpus up to position p.

C. Hyper Parameters setting

The hyper parameters used of all experiments were defined as shown in Table XIV:

TABLE XIV. parameter configurations

Symbol	Description
K	Number of recommended POIs (Top ranked POI)
N	Number of similar users
%Training set	Dataset used to produce recommendation
%Test set	Dataset used on evaluation phase
α, β and γ	SPUUR similarity adjustment values

To achieve the experiments, 80% of the datasets are utilized as training data, while 20% are used as testing data to determine the accuracy of the approaches. We recommend for each user 5, 10, 15 and 20 top ranked POIs (K=5,10,15,20) by considering 50 similar users (N=50). The optimal values of α, β and γ are: (α = 0.75, β = 0.075 et γ = 0.175). In addition, experiments were conducted using PHP on a Windows 11 64-bit with an Intel Core i5-8th@1.30 GHz processor and 20 GB RAM.

D. Experimental procedure

We evaluated the performance of our SPPUR model by comparing it to other recommendation models using traditional similarity measures such as including cosine, adjusted cosine, Spearman's correlation, Pearson's correlation, mean square error (MSD), Euclidean distance and Jaccard's similarity measure. To do this, we divided the dataset into a training data and the test data in terms of check-ins. We use 80% of the check-ins records generated by each user for the training data and the rest for the test data. After, we use user-based collaborative filtering method to recommend Top@K (K=5, 10, 15, 20) POIs to each user. Finally, to evaluate the performance of the algorithm used, we used PRECISION, RECALL, MAP and NDCG. The evaluation steps are described below:

1) Divide the Dataset into training data and test.
2) Build the User-POI frequency matrix.
3) For each user:
   - Calculate the similarity between the selected user and the other users.
   - Select the N most similar users to the selected user.
   - Generate prediction.
   - Select K Top-ranked POIs to recommend to the selected user.
   - Compute the PRECISION@K, RECALL@K, MAP@K and NDCG@K.
4) Compute the global PRECISION@k, RECALL@k, MAP@k and NDCG@k.

6. Results and Discussion

To evaluate the SPPUR model, we used two data sets from Foursquare: the first data set concerns New York City, and the second data set concerns Tokyo. Firstly, we compared the SPPUR model variants with each other to find the one corresponding to the best values of the PRECISION, RECALL, MAP and NDCG parameters. Next, we compared the best variant of the SPPUR model with various state-of-the-art models that use traditional similarities for POI recommendation. Finally, we analyzed and discussed the results obtained during the SPPUR model evaluation process.

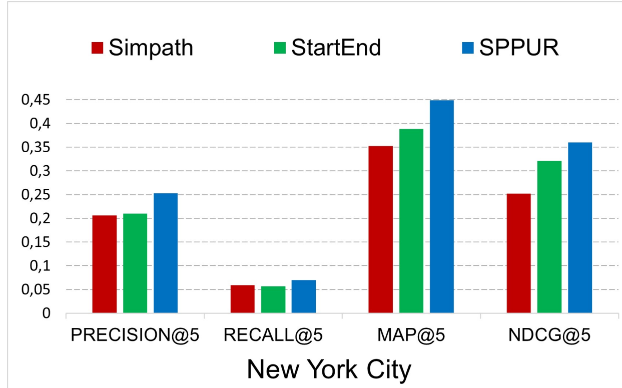

Figure 5. Comparison of different variants of SPPUR on New York dataset

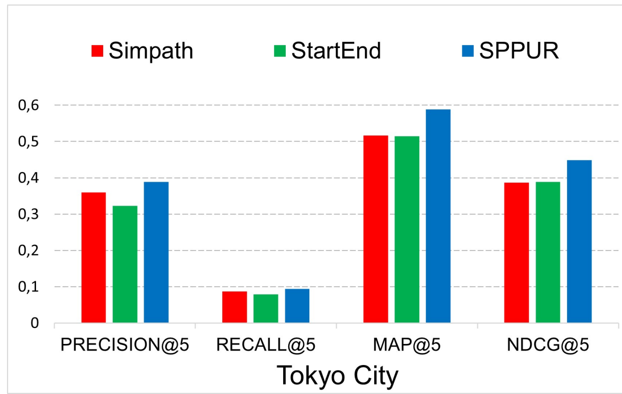

Figure 6. Comparison of different variants of SPPUR on Tokyo dataset

A. Comparison of the three SPPUR model variants

First, we evaluate the performance of each similarity variant in the SPPUR model to determine the optimal similarity measure. For (1) the Sim_{path} based solely on POI paths, (2) the $Sim_{StartEnd}$ using only the start and end points of visits, and finally the SPPUR, which combines the Sim_{path}, Sim_{start}, and Sim_{end} similarities defined in subsection 3-A. Figure 5 and Figure 6 below show that the SPPUR similarity outperforms the Sim_{path} and $Sim_{StartEnd}$ similarities according to PRECISION, RECALL, MAP and NDCG. As a result, we use this SPPUR model similarity because it allows us to adjust the combination of Sim_{path}, Sim_{start}, and Sim_{end} similarities in terms of the hypers parameters α, β and γ.

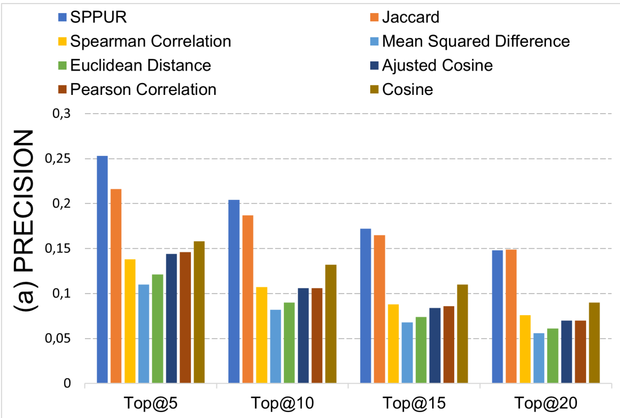

Figure 7. Comparison of SPPUR model with other similarity models using PRECISION on New York dataset

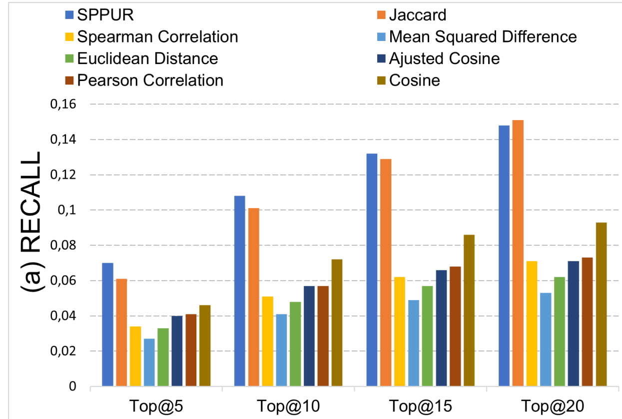

Figure 8. Comparison of SPPUR model with other similarity models using RECALL on New York dataset

B. Comparison of the SPPUR model with other similarity models

In the following, we compare our model based on the SPPUR similarity measure with models in the literature using traditional similarity measures such as Pearson's correlation, Spearman's correlation, Euclidean distance, cosine, adjusted cosine, mean square error (MSD), and Jaccard's similarity measure. This comparison concerns the PRECISION, RECALL, MAP, and NDCG parameters and uses the New York City and Tokyo City data sets.

Figure 7 and Figure 9 below show that the PRECISION of all models using similarity measures decreases as the number of recommended POIs (K) increases. However, Figure 8 and Figure 10 below show that RECALL increases as the number of recommended POIs decreases.

On the other hand, Figure 7, Figure 8, Figure 9 and Figure 10 show that the PRECISION and RECALL parameters of the $SPPUR$ model outperform the other similarity models in the case of the New York data set and the Tokyo dataset.

Figure 11 and Figure 13 below show that the MAP of all models using similarity measures decreases as the number of recommended POIs (K) increases. However, Figure 12 and Figure 14 below show that the NDCG increases as the number of recommended POIs decreases.

On the other hand, Figure 11, Figure 12, Figure 13 and Figure 14 show that the MAP and NDCG parameters of the SPPUR model outperform the other similarity models in the case of the New York data set and the Tokyo dataset.

Finally, the SPPUR similarity measure performs better than other traditional measures according to PRECISION, RECALL, MAP and NDCG. These experiments demon-

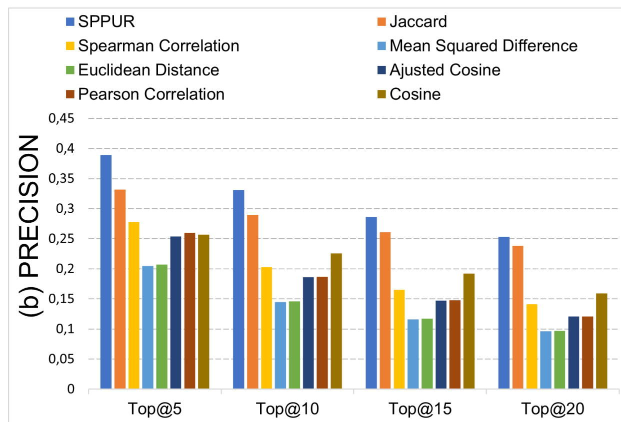

Figure 9. Comparison of SPPUR model with other similarity models using PRECISION on Tokyo dataset

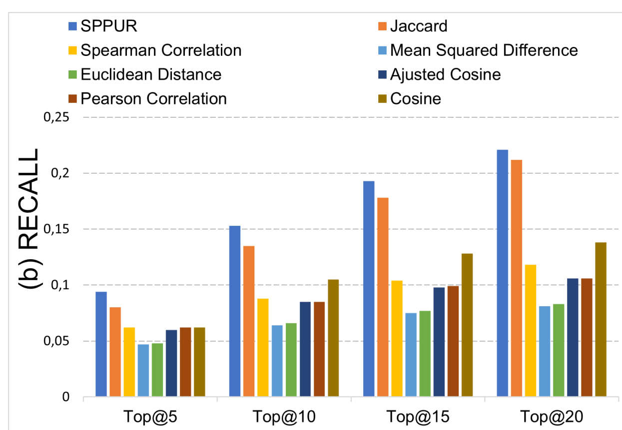

Figure 10. Comparison of SPPUR model with other similarity models using RECALL and Tokyo dataset

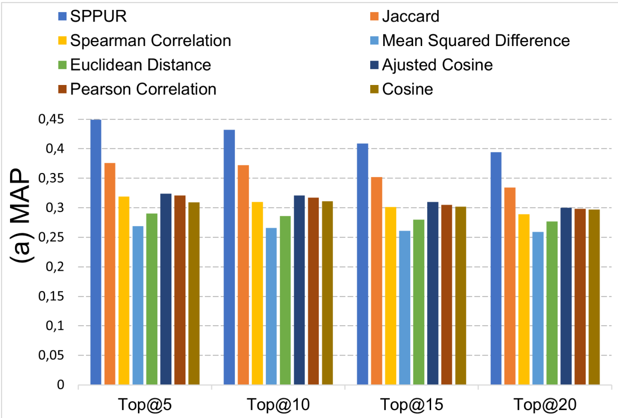

Figure 11. Comparison of SPPUR model with other similarity models using MAP on New York dataset

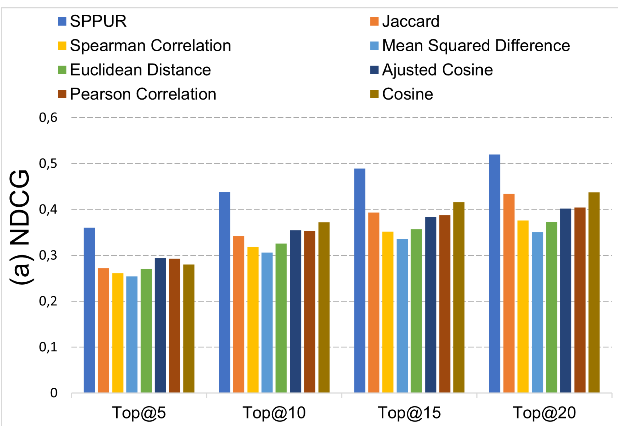

Figure 12. Comparison of SPPUR model with other similarity models using NDCG on New York dataset

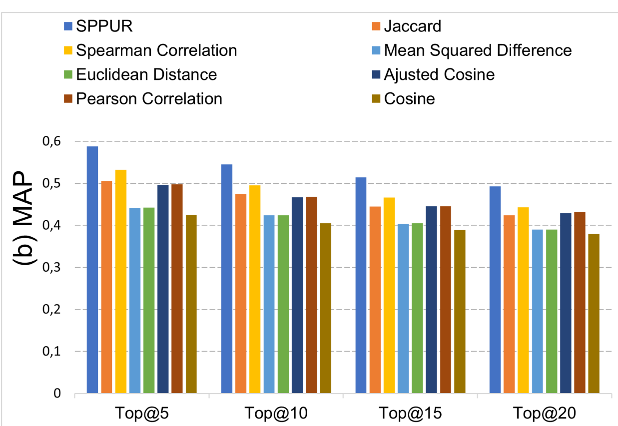

Figure 13. Comparison of SPPUR model with other similarity models using MAP on Tokyo dataset

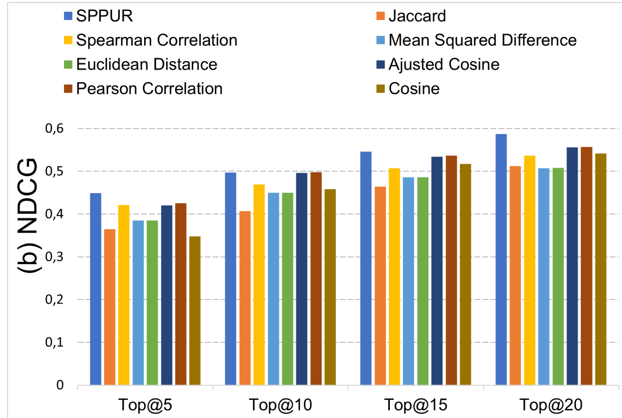

Figure 14. Comparison of SPPUR model with other similarity models using NDCG and Tokyo dataset

TABLE XV. Comparison results on different Number of similar users on New York dataset

N	PRECISION	RECALL	MAP	NDCG
5	0,229	0,06	**0,487**	**0,485**
10	0,249	0,067	0,486	0,457
20	0,255	0,069	0,466	0,406
30	**0,257**	**0,07**	0,461	0,384
40	0,254	0,07	0,451	0,369
50	0,253	0,07	0,449	0,36

strate the effectiveness of this novel similarity measure using two data sets relating to two different cities (New York and Tokyo).

C. Results summary and discussions

The results in subsection 6-A show that the SPPUR similarity outperforms the other variants (Sim_{path} and $Sim_{StartEnd}$) according to PRECISION, RECALL, MAP and NDCG. Therefore, we select this similarity for the SPPUR model. Then, in subsection 6-B, this model is compared with other models in the literature using traditional similarity measures including cosine, adjusted cosine, Spearman's correlation, Pearson's correlation, mean square error (MSD), Euclidean distance and Jaccard's similarity measure. This comparison is conducted using PRECISION, RECALL, MAP and NDCG parameters on the New York and Tokyo datasets. Finally, the results confirm the effectiveness of the SPPUR similarity measure, as it performed better than traditional similarity measures. In addition, to study the effect of the number of neighborhoods (noted N in the algorithm of subsection 3-C) on the performance of the SPPUR model, we recommended the Top@5 POIs for each user by varying the number of neighborhoods used in the user/user similarity calculation. The performance of the SPPUR model is evaluated as a function of different neighborhood sizes (N) according to PRECISION, RECALL, MAP and NDCG. Table XV and Table XVI below present the results of these evaluations.

TABLE XVI. Comparison results on different Number of similar users on Tokyo dataset

N	PRECISION	RECALL	MAP	NDCG
5	**0,419**	0,098	**0,651**	**0,566**
10	0,418	**0,099**	0,636	0,53
20	0,411	0,098	0,62	0,497
30	0,401	0,097	0,61	0,477
40	0,396	0,096	0,599	0,463
50	0,389	0,094	0,588	0,449

The results show that an increase in the number of neighbors leads to a decrease in MAP and NDCG values. This means that the number of neighbors selected plays an important role in the SPPUR similarity calculation process.

7. Conclusions and Future Work

In recent years, the fast development of LBSNs has facilitated tourist activities for people. Similarity measures have an important impact and can considerably improve the accuracy of recommendations for places to visit. In this paper, we propose a novel similarity based on the TF-IDF technique for user-based POI recommendation systems. This measure takes advantage of the TF-IDF technique's effectiveness in the analysis and the exploitation of historical check-ins generated by tourists. Experimental results show that, compared to traditional methods, the proposed similarity measure significantly improves the accuracy of POI recommendation systems that use user-based collaborative filtering. Finally, in terms of perspective, we aim to improve our approach by considering more contexts, such as the semantic characteristics of POIs, regions, weather, and season.

8. Data Availability Statement

The Foursquare Dataset is available online: (https://sites.google.com/site/yangdingqi/home) as open benchmark datasets.

9. Acknowledgment

This study was conducted as part of the PRFU project (Grant: COOLO7UN020120220001) at the Computer Science Department of Hassiba Benbouali University of Chlef (Algeria) and supported by the Ministry of Higher Education and Scientific Research.

References

[1] M. Mustafeez ul Haque, B. Kotaiah, and J. Ahmed, "A comprehensive review of course recommendation systems for moocs," *International Journal of Computing and Digital Systems*, vol. 16, no. 1, pp. 1–15, 2024.

[2] W. Farag, "A comprehensive real-time road-lanes tracking technique for autonomous driving," *International Journal of Computing and Digital Systems*, vol. 9, no. 03, 2020.

[3] F. Hasan Yusuf and M. A Mangoud, "Real-time car parking detection with deep learning in different lighting scenarios," *International Journal of Computing and Digital Systems*, vol. 15, no. 1, pp. 1–9, 2024.

[4] S. Medjroud, N. Dennouni, and M. Loukam, "Towards a systematic point-of-interest recommendations based on trust between users deduced from their ratings and check-ins in a lbsn," *International Journal of Computing and Digital Systems*, vol. 16, no. 1, pp. 189–203, 2024.

[5] M. H. Henni, N. Dennouni, Z. Slama, S. Medjroud, and D. Bettache, "Towards an approach for online evaluation of new variants of content-based poi recommender systems by mobile tourists," in *2022 First International Conference on Big Data, IoT, Web Intelligence and Applications (BIWA)*. IEEE, 2022, pp. 89–94.

[6] M. HADJHENNI, N. DENNOUNI, and Z. SLAMA, "Toward a systematic evaluation approach of point-of-interest recommendation algorithms of a novel smart tourism tool," *International Journal of Computing and Digital Systems*, vol. 15, no. 1, pp. 653–670, 2024.

[7] S. Qaiser and R. Ali, "Text mining: use of tf-idf to examine the relevance of words to documents," *International Journal of Computer Applications*, vol. 181, no. 1, pp. 25–29, 2018.

[8] U. Gretzel, M. Sigala, Z. Xiang, and C. Koo, "Smart tourism: foundations and developments," *Electronic markets*, vol. 25, pp. 179–188, 2015.

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

[10] C. Cheng, H. Yang, I. King, and M. Lyu, "Fused matrix factorization with geographical and social influence in location-based social networks," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 26, no. 1, 2012, pp. 17–23.

[11] 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 *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014, pp. 831–840.

[12] J.-D. Zhang and C.-Y. Chow, "Geosoca: Exploiting geographical, social and categorical correlations for point-of-interest recommendations," in *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*, 2015, pp. 443–452.

[13] Y. Liu, T.-A. N. Pham, G. Cong, and Q. Yuan, "An experimental evaluation of point-of-interest recommendation in location-based social networks," 2017.

[14] S. Zhao, M. R. Lyu, I. King, S. Zhao, M. R. Lyu, and I. King, "Geo-teaser: Geo-temporal sequential embedding rank for poi recommendation," *Point-of-Interest Recommendation in Location-Based Social Networks*, pp. 57–78, 2018.

[15] T. Zhang, W. Zheng, Z. Cui, Y. Zong, and Y. Li, "Spatial–temporal recurrent neural network for emotion recognition," *IEEE transactions on cybernetics*, vol. 49, no. 3, pp. 839–847, 2018.

[16] D. Lian, H. Wang, Z. Liu, J. Lian, E. Chen, and X. Xie, "Lightrec: A memory and search-efficient recommender system," in *Proceedings of The Web Conference 2020*, 2020, pp. 695–705.

[17] Z. Huang, J. Ma, Y. Dong, N. Z. Foutz, and J. Li, "Empowering next poi recommendation with multi-relational modeling," in *Proceed-*

ings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2022, pp. 2034–2038.

[18] J. Xia, Y. Yang, S. Wang, H. Yin, J. Cao, and S. Y. Philip, “Bayes-enhanced multi-view attention networks for robust poi recommendation,” *IEEE Transactions on Knowledge and Data Engineering*, 2023.

[19] J. Wang, L. Zhang, Z. Sun, and Y.-S. Ong, “Meta-learning enhanced next poi recommendation by leveraging check-ins from auxiliary cities,” in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 2023, pp. 322–334.

[20] T. Zhang, H. Liu, X. Geng, and G. Yu, “Event-based probabilistic embedding for poi recommendation,” *Applied Sciences*, vol. 13, no. 3, p. 1236, 2023.

[21] M. Chiny, M. Chihab, O. Bencharef, and Y. Chihab, “Netflix recommendation system based on tf-idf and cosine similarity algorithms,” *no. Bml*, pp. 15–20, 2022.

[22] J. Ni, Y. Cai, G. Tang, and Y. Xie, “Collaborative filtering recommendation algorithm based on tf-idf and user characteristics,” *Applied Sciences*, vol. 11, no. 20, p. 9554, 2021.

[23] P. Dauni, M. Firdaus, R. Asfariani, M. Saputra, A. Hidayat, and W. Zulfikar, “Implementation of haversine formula for school location tracking,” in *Journal of Physics: Conference Series*, vol. 1402, no. 7. IOP Publishing, 2019, p. 077028.

[24] R. Purnomo, T. D. Putra, H. Kusmara, W. Priatna, and F. Mukharom, “Haversine formula to find the nearest petshop. jatisi (jurnal teknik informatika dan sistem informasi), 9 (3), 2205–2221,” 2022.

[25] N. Idrissi and A. Zellou, “A systematic literature review of sparsity issues in recommender systems,” *Social Network Analysis and Mining*, vol. 10, no. 1, p. 15, 2020.

[26] “Dingqi yang foursquare dataset. dingqi yang’s homepage; china. available online,” https://sites.google.com/site/yangdingqi/home, accessed on 1 Feb 2024.

[27] A. A. Amer, H. I. Abdalla, and L. Nguyen, “Enhancing recommendation systems performance using highly-effective similarity measures,” *Knowledge-Based Systems*, vol. 217, p. 106842, 2021.

[28] A. Almu and Z. Bello, “An experimental study on the accuracy and efficiency of some similarity measures for collaborative filtering recommender systems,” 2021.

[29] H. Khojamli and J. Razmara, “Survey of similarity functions on neighborhood-based collaborative filtering,” *Expert Systems with Applications*, vol. 185, p. 115482, 2021.

[30] F. Fkih, “Similarity measures for collaborative filtering-based recommender systems: Review and experimental comparison,” *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 9, pp. 7645–7669, 2022.

[31] S. R. Bashir, S. Raza, and V. B. Misic, “Bert4loc: Bert for location—poi recommender system,” *Future Internet*, vol. 15, no. 6, p. 213, 2023.

[32] M. Singh, “Scalability and sparsity issues in recommender datasets: a survey,” *Knowledge and Information Systems*, vol. 62, no. 1, pp. 1–43, 2020.

[33] D. Valcarce, A. Bellogín, J. Parapar, and P. Castells, “Assessing ranking metrics in top-n recommendation,” *Information Retrieval Journal*, vol. 23, pp. 411–448, 2020.

[34] A. Vineela, G. Lavanya Devi, N. Nelaturi, and G. Dasavatara Yadav, “A comprehensive study and evaluation of recommender systems,” in *Microelectronics, Electromagnetics and Telecommunications: Proceedings of the Fifth ICMEET 2019*. Springer, 2021, pp. 45–53.

[35] K. V. Rodpysh, S. J. Mirabedini, and T. Banirostam, “Employing singular value decomposition and similarity criteria for alleviating cold start and sparse data in context-aware recommender systems,” *Electronic Commerce Research*, vol. 23, no. 2, pp. 681–707, 2023.

[36] C. Xu, A. S. Ding, and K. Zhao, “A novel poi recommendation method based on trust relationship and spatial–temporal factors,” *Electronic Commerce Research and Applications*, vol. 48, p. 101060, 2021.

Djelloul BETTACHE is a PhD student in the computer science department at Hassiba Benbouali University of Chlef (Algeria). He earned a Master’s degree in computer science in 2015 from Hassiba Benbouali University of Chlef . His primary research domains are artificial intelligence and POI recommendation in Location-Based Social Networking (LBSN).

Nassim DENNOUNI is an associate professor at the Higher School of Management at Tlemcen since October 2022. He is the ICAR team leader of the LIA Laboratory of the Hassiba BENBOUALI University of Chlef (Algeria) and he is also a member of the ISIBA team of EEDIS laboratory of Computer Science Department of Djillali Liabes University of Sidi Bel Abbes (Algeria). In 2018, he obtained his university accreditation from Djillali Liabes University of Sidi Bel Abbes. In 2016, he received a PhD in computer science from Djillali Liabes University of Sidi Bel Abbes. His primary research domains are artificial intelligence, ubiquitous computing, mobile learning, orchestration of activities, and POI recommendation.