Unifying Spatial, Temporal and Semantic Features for an Effective GPS Trajectory-Based Location Recommendation

Hamidu Abdel-Fatao(✉), Jiuyong Li, and Jixue Liu

School of Information Technology & Mathematical Sciences, University of South Australia, Adelaide, Australia
hamidu.abdel-fatao@mymail.unisa.edu.au

Abstract. Location recommendation aims at providing personalized suggestions of a set of new and potentially interesting locations to a target user. The underlying principle of this problem is to predict the *Degree of Relevance* of candidate locations to the user and make recommendations accordingly. Enormous attention has been devoted to this problem by research and industrial community lately due to its applicability in numerous applications. In this work we develop an effective GPS trajectory-based location recommendation framework for *Location Based Social Networks*. We propose an algorithm, *STS Location Recommender*, to leverage unique properties of GPS trajectories namely spatial, temporal and semantic features for recommendation. Our algorithm specifically exploits temporal and semantic influence on users' mobility fused with spatial properties of locations to model relevance of locations to users. Prior to our work, no existing studies based on GPS trajectories simultaneously used all of these features for location recommendation. We experiment on real-world GPS datasets to show that our approach provides more precise recommendations compared with baseline approaches.

Keywords: GPS trajectory data mining · Collaborative filtering · Location recommendation systems · Location-based social networks

1 Introduction

Location Based Social Networks (LBSNs) are new wave of interactive platforms which allow users to share their geospatial locations alongside other location-related contents such as comments, experiences etc. The advent of LBSNs is a direct consequence of recent advances and ubiquity of Web and mobile technologies [1–3]. Since their evolution, LBSNs have received considerable attention from research and industry because they play crucial role in the development of many important applications [2,4]. Popular among these are location recommendation systems, urban planning, counter-terrorism etc.

The object of this study is to develop a GPS trajectory-based Location Recommendation System (LRS) for LBSNs. This problem fundamentally aims at

M.A. Sharaf et al. (Eds.): ADC 2015, LNCS 9093, pp. 41–53, 2015.
DOI: 10.1007/978-3-319-19548-3_4

exploiting historical GPS trajectory data to provide personalized suggestions of previously unvisited locations to users in LBSNs. The core issue is to predict, with a reasonable level of accuracy, the *Degree of Relevance (DoR)* of candidate locations to users and to make recommendations accordingly. For example if a user highly favours a particular type of restaurant e.g. Asian restaurants, LRS can infer this and make recommendations accordingly.

The major challenges that confront LRSs entail how to effectively model the DoR of locations to users that is - (i) able to leverage spatial features, temporal features and semantic features of locations for recommendation, (ii) personalized and diverse i.e. capable of recommending truly relevant locations which represent all possible interests to users. This is challenging considering the number of factors that must by taken into account.

In a bid to address these challenges a number of techniques have been proposed. For example Ye et al [5] exploit three factors namely user preference, social influence and geographical influence for *PoI* recommendations based on user check-in datasets. Yuan et al [3] is similar to [5] except that, [3] considers temporal factor as an additional constraint for time-specific PoI recommendation. Zheng et al al. [6] proposed a matrix decomposition technique using GPS trajectories combined with activity information, for global location and activity recommendations. As an extension to [6], Zheng et al [1] proposed *UCLAF* that incorporates a user dimension as an additional entity for personalized location and activity recommendations using Tensor decomposition technique.

These existing works are riddled with either one or both of two major flaws. These flaws include the fact that (i) *they are geographically constrained* $-$ cannot make recommendations when there exist no geographical overlap between a target user's location history and candidate locations. (ii) *they are time-unaware* $-$ only recommend locations globally and cannot tell where a user will like to be at a specific time. This is because majority of the works [2,5,7] only rely on spatial features for recommendation. Few others [1,6,8] consider semantic features in addition to spatial features. Even fewer works [3,4] consider temporal features in addition to spatial features. To the best of our knowledge, no existing work based on GPS trajectories simultaneously used all the three features for location recommendation. Our work bridges these knowledge gaps by considering all three features simultaneously to provide more precise time-aware and semantically meaningful recommendations.

We summarize the main contributions of this paper as follows.

– We develop a novel approach for modeling DoR of locations to users that simultaneously exploits semantic, temporal and spatial features of locations.
– We develop an effective algorithm *STS Location Recommender*, the provides more precise, time-aware and semantically meaningful locations recommendations based on historical GPS trajectory data.

We evaluate our proposed location recommendation framework by conducting experiments using real-world GPS dataset. Experimental results show that our proposed LRS is more precise compared with baseline approaches.

The rest of this work is organized as follows. We explain relevant concepts and our problem statement in Section 2. In Section 3, we detail our proposed technique, present our experiments in Section 4 and conclude in Section 5.

2 Statement of Problem

In this section, we clarify relevant concepts and present formal statement of the problem addressed in this paper.

2.1 Preliminaries

Historical human mobility behaviour can be reconstructed using traces of geographic points known as trajectory points. A *Trajectory Point* p is a spatial point associated with a timestamp, denoted by the triple $p = (x, y, t)$ where x and y are the latitude and longitude respectively of point p at timestamp t.

UserID	Date	Time	Latitude	Longitude
1509	2008-02-06	16:29:31	123.5322	42.30713
1509	2008-02-06	16:29:36	123.53218	42.30713
1509	2008-02-06	16:29:41	123.53217	42.30713
1509	2008-02-06	16:29:46	123.53217	42.30714
1509	2008-02-06	16:29:51	123.53217	42.30713

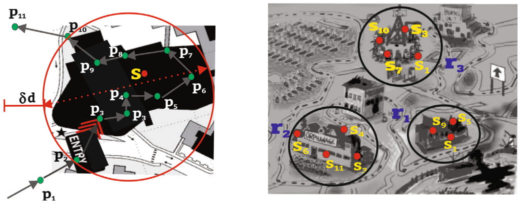
Table 1. Sample Trajectory Dataset

Definition 1. *Trajectory* denoted by $P = \langle p_1, p_2, ..., p_z\rangle$ is a sequence of trajectory points organised in ascending order of timestamps, where $\{p_i \in P : p_i = (x_i, y_i, t_i)\}$ is a trajectory point and $t_i < t_{i+1}$, $\forall i \in [1, z]$.

Typically, raw trajectory datasets (see sample in Table 1) are available as very large volumes of geographically close points. Analysing such datasets directly can introduce significant computational overheads. To curb this problem, we follow the intuition that when people visit places of interest, they stay within nearby areas for significant periods of time. For example, in a cinema people usually stay within the Cinema Hall for considerably periods of time watching movies. We therefore extract such significant areas called *Stay Points* from trajectories.

Definition 2. *Stay Point* denoted by $s = [(x, y), t_a, t_s]$ is a geographical area characterized by a maximum distance threshold δ_d where a user stayed for at least a minimum threshold period of time δ_t, and x, y, t_a and $(t_s \geq \delta_t)$ are respectively latitude, longitude, arrival time and stay time of s.

To extract stay points, we implement an existing algorithm by Zheng et al [2]. We chose this approach for its intuitive and consistency with our definition. Figure 1a illustrates the approach diagrammatically (please refer to [2] for depths).



(a) Stay Point (b) Reference Points

Fig. 1. Trajectory Transformation

Transformation of trajectory points to stay points drastically reduces the trajectory data. However, the volume of stay points can still be reduced by clustering close and adjacent stay points representing the same location but having slightly different coordinates. To achieve this we employ a density-based clustering algorithm *OPTICS* [9] to cluster geographically close stay points into non-overlapping clusters. We represent each discovered cluster with a single point called *Reference Point* (see Figure 1b) defined below.

Definition 3. *A Reference Point* denoted by $r = [(x_r, y_r), t_a, t_s]$, is a representative of a cluster of stay points $S_c=\{s'_1, s'_2, ..., s'_q\}$, where (x_r, y_r) is the average coordinate of the stay points $s'_i \in S_c$, t_a and t_s are respectively the earliest arrival time and mean stay time of the stay points in S_c.

Since our core objective in this study is to recommend semantically meaningful locations to users, it is necessary to enrich reference points discovered with their underlying semantic tags such as gym, restaurant, park etc. Fortunately, many places in LBSNs have been labeled with semantic tags [8]. We therefore take steps to annotate each reference point with its corresponding semantic tag and call it a *Semantic Location.*

Definition 4. *Semantic Location L,* denoted by $L = [(x_r, y_r) : l_f, t_a, t_s]$ symbolizes a reference point annotated with a semantic tag where l_f represents semantic tag, and $(x_r, y_r), t_a, t_s$ have their usual meanings.

To date, there exists no universally accepted standard for assigning precise semantic tags to geographic points even though there has been attempts by some studies [10,11]. In this work, we use of a *Point of Interest (PoI)* database for this task. A PoI database is a corpus of PoIs such as restaurant, shop, cinema, etc associated with geographic coordinates. Specifically, we use Foursquare[1] category database because it's PoI data is highly reliable [12] and used by popular search engines such as Microsoft Bing.

Having obtained semantic locations from reference points we model users' mobilities as sequences of semantic locations called semantic trajectories.

Definition 5. *Semantic Trajectory* denoted by $T_s = \langle L_1, L_2, ..., L_m\rangle$ is a time-ordered sequence of semantic locations where $L_i \in T_s$ $(1 \leq i \leq m)$ is a semantic location.

As an example, a semantic trajectory of a user can be represented by the sequence ⟨[(39.9993,116.3269) : Hotel, 13:30, 72 min] → [(39.9952,116.3272) : Park, 14:43, 54 min] → [(39.8201,116.2766) : Restaurant, 17:02, 32 min]⟩.

2.2 Semantic Trajectory Pattern Mining

A *Semantic Trajectory Pattern* represents routine mobility behaviour of a user discovered from his/her semantic trajectories. For example, if on most days a user goes to work at 9 am, restaurant at 12 pm and gym at 6 pm we call such a mobility behaviour a semantic trajectory pattern.

[1] www.foursquare.com

A number of works [8,13] have tackled the problem of mining mobility patterns. These studies, inspired by the fact that user mobility is typically sequential in nature, perform sequential pattern mining on preprocessed mobility histories to discover frequent mobility behaviours. However, most of these works neglect temporal information in their approaches. For example Ying et al [8] represent a semantic trajectory pattern in the form ⟨{Hospital}{Park}{Bank}⟩, which clearly does not contain temporal information. We argue that, temporal information is crucial in understanding mobility behaviours because users' mobilities typically exhibit temporal patterns. For example, a user might visit a restaurant at 12 pm everyday, but he/she may typically visit other places at nighttime.

In our work, we take temporal information into account in mining semantic trajectory patterns. To facilitate this, we encode each semantic location with an equivalent numerical representation in a three-step process. Firstly, we assign the semantic tag associated with each semantic location a unique integer identifier. For example (Restaurant ⇒ 103) denotes a transformation of *Restaurant* to an integer value 103. Secondly we split each day into six equal four-hourly non-overlapping time slots e.g. [00:00 − 03:59]⇒ 1, [04:00 − 7:59]⇒ 2 etc. Finally we identify each stay time by integers $1, 2$ or 3 denoting short, medium and long stay times respectively. As an example, the semantic location [(39.8201, 116.2766) : Restaurant, 17:02, 32 mins] is transformed to $[103, 4, 1]$ meaning the user visited semantic feature 103 during time slot 4 and stayed for short time period. We then perform sequential pattern mining on the transformed dataset to extract frequent semantic trajectory patterns.

2.3 Problem Formalization

Having explained basic concepts, we now detail our problem definition.

Problem Definition Given a target user $u_i \in U$ such that $U = \{u_1, u_2, ..., u_n\}$ is a set of users in a city, the problem is to recommend *top k previously unvisited and semantically meaningful locations* that u_i might be interested in at time t, based on his/her location preferences and current location in the city.

To address this problem there is a need to : (i) precisely model u_i's preferences for past locations using his/her location histories, (ii) accurately estimates the DoR of each previously unvisited location to u_i based on his/her preferences. We propose an algorithm, ***S**patio-**T**emporal and **S**emantic-Aware (STS) Location Recommender* to solve the problem. The novelty of our algorithm lies in its ability to leverage semantic, temporal and spatial feature for time-aware and semantically meaningful location recommendation even where there exist no geographical overlap between unvisited locations and users' location histories.

3 The STS Location Recommender

In this section, we elaborate on the STS Location Recommender. The recommender takes a three-step approach summarized as follows: (i) Location preference estimation (ii) User similarity estimation (iii) User-based collaborative location recommendation. We explain each step in the following subsections.

3.1 User Preference Estimation

A User's preference for a semantic location is a measure of interestingness of the location to the user. We express this measure mathematically in terms of *Preference Score* defined and formulated as follows.

Definition 6. *Preference Score* of a location with respect to a target user during a specified time slot is a numerical estimate of the likelihood of the user visiting the location during the given time slot.

Preference score comprises two component probabilities viz *Visit Probability* in terms of *(i) Semantic Influence*; *(ii) Temporal Influence*, derived as follows.

Suppose $T_s(u) = \{L_1, L_2, ..., L_m\}$ denotes a set of semantic locations visited by a user u and let $\tau(u) = \{l_1, l_2, ..., l_m\}$ be corresponding set of semantic tags of $T_s(u)$, where each $l_i \in \tau(u)$ is the semantic tag of $L_i \in T_s(u)$ $(1 \leq i \leq m)$. Let ϕ_{l_i} be a binary variable denoting a visit to a location tagged by l_i. *Visit Probability in terms of semantic influence* $P(l_i)$ of l_i is expressed as $P(l_i) = w_s \times {}^{c(\phi_{l_i})}/_{|\tau(u)|}$, where $c(\phi_{l_i})$ is total count of ϕ_{l_i} and w_s is a weight expressed in terms of popularity of l_i. That is, $w_s = 1 + log({}^{N_u}/_{N_{l_i}})$ where N_u is total number of users, and N_{l_i} is the number of users who visited a location tagged by l_i.

Also, suppose $T_s^t(u) = \{L_1^t, L_2^t, ..., L_k^t\}$ denotes a set of semantic locations visited by u during a specific time slot t and let $\tau^t(u) = \{l_1^t, l_2^t, ..., l_k^t\}$ be corresponding set of semantic tags of $T_s^t(u)$, where each $l_i^t \in \tau^t(u)$ is the semantic tag of $L_i^t \in T_s^t(u)$ $(1 \leq i \leq m)$. Let $\phi_{l_i^t}$ be a binary variable denoting a visit to a location tagged by l_i by u during t. *Visit probability in terms of temporal influence* $P(l_i^t)$ of l_i is expressed as $P(l_i^t) = w_t \times {}^{\sum \phi_{l_i^t}}/_{|T_s^t(u)|}$, where $w_t = 2^{\overline{\delta_t(l_i^t)}/\sum \overline{\delta_t(l_j^t \in \tau^t(u))}}$ such that $\overline{\delta_t(l_i^t)}$ is the mean stay time at l_i during t and $\sum \overline{\delta_t(l_j^t \in \tau^t(u))}$ is the sum of mean stay times for all locations visited during t.

Finally, *Preference Score* for a semantic location with semantic tag l_i with respect u during time slot t expressed in terms of $P(l_i)$ and $P(l_i^t)$ is given by

$$P_u(l_i, t) = \lambda P(l_i) + (1 - \lambda) P(l_i^t) \tag{1}$$

where λ in Equation 1 is a tuning parameter that controls the influence of semantic and temporal factors on $P_u(l_i, t)$.

3.2 User Similarity Estimation

User similarity measures the extent to which two users share preferences for semantic locations visited during specified time slots. Intuitively, users who share similar lifestyles will exhibit similar preferences for semantic locations during specific time slots. We estimate similarity between any two users using Pearson's correlation coefficient similarity metric explained as follows.

Let $\tau^t = \tau^t(u) \cup \tau^t(v)$ be a set of semantic tags corresponding to semantic locations visited by users u and v during a given time slot t. For u, there exists a preference score $P_u(l_i, t)$ for $l_i \in \tau^t$ if u has visited l_i. Similarly for v, there exist a preference score $P_v(l_i, t)$ if v has visited $l_i \in \tau^t$. Given vectors of preference scores for u and v corresponding to semantic locations visited during t, the similarity between u and v during t is given by

$$sim^t(u,v) = \frac{\sum_{l_i \in \tau^t} [P_u(l_i,t) - \overline{P_u(\tau^t(u))}][P_v(l_i,t) - \overline{P_v(\tau^t(v))}]}{\sqrt{\sum_{l_i \in \tau^t} [P_u(l_i,t) - \overline{P_u(\tau^t(u))}]^2}\sqrt{\sum_{l_i \in \tau^t} [P_v(l_i,t) - \overline{P_v(\tau^t(v))}]^2}} \tag{2}$$

where $\overline{P_u(\tau^t(u))}$ and $\overline{P_v(\tau^t(v))}$ are the mean preference scores of locations visited by u and v during t respectively.

Given a set of time slots T, the overall similarity between the two users measured over all time slots is given by

$$Sim(u,v) = \frac{\sum_{t \in T} sim^t(u,v)}{|T|} \tag{3}$$

3.3 Location Recommendation

In this subsection we present our algorithm for location recommendation.

Our idea follows user-based Collaborative Filtering (CF) [14] model. We utilize the intuition that like-minded individuals who share similar lifestyles are most likely to visit similar semantic locations at similar times. For example, two users who exhibit similar levels of preferences for nightlife are likely to visit say, a cinema for movies at night time.

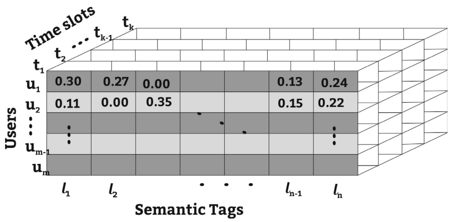


Fig. 2. Aggregated Users' Location Histories

Specifically, we treat each semantic location visited at a specific time as an "item" and users' preference scores for these locations as their implicit ratings on the locations. We then build user-location matrix for each time slot, where entries are preference scores for corresponding user-location pairs. For all time slots the user-location matrices are aggregated into a three dimensional representation (see Figure 2). Based on this model STS Location Recommender algorithm 3 is executed for recommendation. We summarize the steps in the algorithm as follows.

Algorithm 1: STS Location Recommender

Input: Target user u, set of users U, user-location matrices M for predefined time slots T
Output: *Top-N* recommended locations for u at $t \in T$
foreach $v \in U$ **do**
  $Sim(u,v) \leftarrow$ compute similarity between u and v using Equation 3
$U_K \leftarrow$ select K users with the largest $Sim(u,v)$ values
foreach $v_K \in U_K$ **do**
  foreach $l_j \in M$ *for which u has no preference for but visited by v_K at t* **do**
    $R^t(u,l_j) \leftarrow$ compute DoR for l_j using to Equation 4
    $\omega_{l_j} \leftarrow$ compute willingness measure according to Equation 5
    $\hat{R}^t_{u,l_j} \leftarrow$ compute u's recommendation score for l_j
return *Top-N* $l_j \in M$ with the highest $\hat{R}^t_{u,l_j}$ for u at t

Fig. 3. STS Location Recommender

Firstly, given a target user u during a specific time slot $t \in T$ and a set of LBSN users U, the similarity $Sim(u, v)$ between u and each user v ($v \in U$) is computed. Secondly, a set of K users U_K, with the highest similarity score is selected. For each $v_K \in U_K$, a set of semantic locations L_r, previously unvisited by u but visited by v_K are extracted as candidates. For each semantic location in this set, the DoR of its semantic tag l_j to u given by

$$R^t(u, l_j) = \overline{P_u(\tau^t(u))} + \frac{\Sigma_{v_K \in U_K} Sim(u, v_K).P_{v_K}(l_j, t)}{\Sigma_{v_K \in U_K} Sim(u, v_K)} \tag{4}$$

is computed as u's implicit preference for the l_j, where $\overline{P_u(\tau^t(u))}$, $P_{v_K}(l_j, t)$ and $Sim(u, v_K)$ have their usual meanings. Finally, the recommendation score for the semantic location tagged by l_j during time slot t is obtained by taking into account the geospatial distance of u's current location to the semantic location corresponding to l_j. We use a *Proximity Measure* ω_{l_j} defined in terms of geospatial distance to implicitly determine u's willingness to visit the candidate locations. That is $\omega_{l_j} = \frac{1}{log(dist(l_c, l_j))}$, where $dist(l_c, l_j)$ is the geospatial distance between the current semantic location of u having semantic tag l_c and semantic location having semantic tags l_j. Using the proximity measure, u's recommendation score for l_j is computed as

$$\hat{R}^t_{u,l_j} = \omega_{l_j} \times R^t(u, l_j) \tag{5}$$

The *top-N* semantic locations with the highest $\hat{R}^t_{u,l_j}$ are returned as recommendations for the user u during time slot t.

4 Experiments

In this section, we evaluate the effectiveness of STS Location Recommender through experiments. We present a description of our dataset, then discuss metrics employed for evaluation and compare our approach with baseline models.

4.1 Description of Dataset and Experimental Settings

In this work, we utilized *GeoLife*[2] real-world GPS trajectory dataset collected from 182 individuals over a period of 5 years (April 2007 to August 2012). We chose individuals with sufficiently large number trajectories (i.e. having trajectories spanning a period of at least one week) in order to increase our chances of finding trajectories which exhibit routine mobility cycles. We found that trajectories of 149 users satisfied this requirement and processed their datasets accordingly. To evaluate our approach we compared with the methods in [15] herein abbreviated as $UBCF$ and [1] abbreviated as $UCLAF$. We describe how we adapt our dataset to perform these comparisons as follows.

[2] http://research.microsoft.com/en-us/downloads/b16d359d-d164-469e-9fd4-daa38f2b2e13/

UBCF by Herlocker et al [15] is a benchmark for most conventional user-based CF approaches. The idea behind UBCF is to perform location recommendations based on users' implicit ratings on locations estimated from their location histories, regardless of temporal information. For each user we used visit probability in terms of semantic influence as his/her implicit rating on corresponding locations. We then construct a user-location matrix to perform UBCF.

In $UCLAF$ [1], Zheng et al employed a *User-Location-Activity tensor* in addition to *User-User, Location-Feature, User-Location* and *Activity-Activity* matrices for collaborative location and activity recommendation using a tensor decomposition technique. Since our dataset lacks activity information and we did not have access to their dataset, we adapt our dataset to conform to their settings. To do this, we utilized activity information mentioned in their experiments namely *Food* & *Drink*, *Shopping*, *Movies* & *Shows*, *Sports* & *Exercise* and *Tourism* & *Amusement* in addition to our user-location information to generate User-Location-Activity tensor. Further we considered the first five closest semantic features to each reference point to construct location-feature matrix. We utilized Pearson correlation similarity metric to compute similarity between users to construct user-user matrix. For activity-activity matrix, we employed the web in the same manner as their work, using activity information mentioned earlier to get the entries. Finally, we use frequency of visit of each semantic location to generate entries for user-location matrix .

4.2 Evaluation Methodology

Our STS Location Recommender under investigation estimates a recommendation score for each candidate location and returns the *top-N* highest ranked locations to a target user as recommendations, given his/her current location and time of the day. To study the effectiveness and prediction accuracy of our proposed approach, we evaluate our LRS in terms of (i) Precision and Recall; (ii) Root Mean Square Error *(RMSE)*.

Precision and Recall investigate how many locations marked off in our test dataset during the processing step are recovered in the returned recommended locations. More specifically, we examine

1. Precision@N: how many locations in the top-N recommended locations during timeslot t correspond to the hold-off locations in the testing data.
2. Recall@N: how many locations in the hold-off locations in the testing set are returned as top-N recommended locations during timeslot t.

We tested the performance when N = 5, 10, 15 with 5 as default value.

Root Mean Squared Error (RMSE) measures deviation of generated recommendation scores $\hat{R}_{u,l_j}.predicted$, for each user-location pair (u, l_j) from the actual values $\hat{R}_{u,l_j}.actual$. $RMSE$ between predicted and actual scores is given by

$$RMSE = \sqrt{\frac{1}{|Test_Set|} \sum_{(u,l_j)\in Set} (\hat{R}_{u,l_j}.predicted - \hat{R}_{u,l_j}.actual)^2} \qquad (6)$$

Note that low values of $RMSE$ indicate a good quality of prediction.

Since we do not have ground truth evaluation of our work due to the source of our dataset, we only rely on these metrics for evaluations. For each user, we randomly mark-off 30%, 40% and 50% of all locations visited for testing.

4.3 Experimental Results and Comparison

Firstly, we investigate the impact of semantic and temporal information on the performance of STS Location Recommender. We then compare the effectiveness of our algorithm with a conventional approaches namely *UBCF* [15]. Finally, we compare our accuracy with *UCLAF* [6] and *UBCF* [15] in terms of RMSE.

Influence of Semantic and Temporal Information. We study influence of temporal and semantic features on recommendations by varying tuning parameter λ in Equation 1. A high value of λ indicates semantic information has a higher impact on the accuracy of recommendation. On the other hand, a low value of λ means accuracy of recommendation improves with increase in temporal information. Our findings are shown in Figure 4.

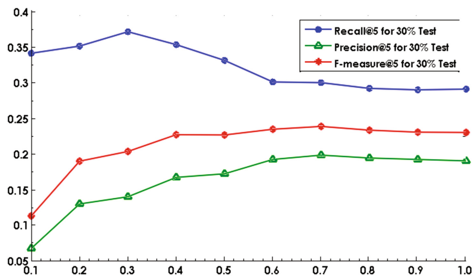
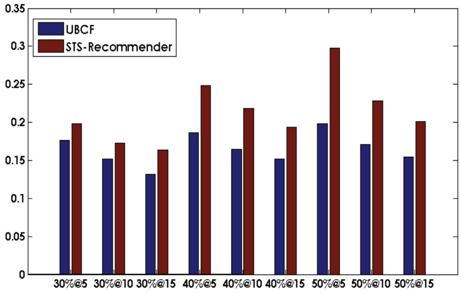


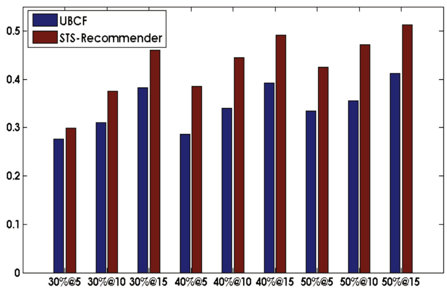
Fig. 4. Variations in Tuning Parameter

As shown in the Figure 4, the best precision was achieved at $\lambda = 0.7$ and the best recall at $\lambda = 0.3$. This shows that semantic information has more effect on determining precision whiles recall is predominantly determined by temporal factor. Since the best value of λ is not the same for precision and recall, we take the harmonic mean of precision and recall. In particular, for each value of λ, we find the F-measure given by $F\text{-}measure = 2 * \frac{precision * recall}{precision + recall}$. We found that λ value of 0.7 gave the best value for F-measure as shown in Figure 4. We therefore set our $\lambda = 0.7$ in our experiments to measure precision and recall.

Comparison of Precision and Recall with Conventional Approach. We investigate the impact of temporal information on the effectiveness of STS Location Recommender in comparison with a conventional approach. Specifically, we compare *Precision* and *Recall* for various percentages of test dataset using our algorithm in comparison with UBCF. UBCF is purely based on conventional user-based collaborative filtering technique which does not take temporal information of users' movement into account. Our algorithm on the other hand uses temporal information in terms of time of visit weighted by stay time at visited locations. Figure 5 shows the results obtained. From the results, both STS Location Recommender and UBCF show similar trends in performances in terms of both precision and recall. That is, both precision and recall generally increase with increase in percentage test of datasets using the two approaches.



(a) Precision@N



(b) Recall@N

Fig. 5. Comparison of Precision and Recall

Also increase in the number of top recommended locations leads to decrease in precision but increase in recall in both cases. However, in spite of the similarities in the trends, STS Location Recommender significantly outperforms UBCF in terms of precision and recall in all cases. This indicates that incorporating temporal information improves both the quality and accuracy of location recommendation.

We record low values for both precision and recall using STS Location Recommender and UBCF. This is not surprising because GPS trajectory datasets are typically sparse in nature and thus contribute to the low values of precision and recall. However, it is much better than random prediction. In this work, we emphasize on the relative improvements achieved instead of the absolute values we obtained. Note that we do not compare precision and recall with UCLAF because we run UCLAF on our pre-processed dataset using the authors code which does not consider these metrics in their evaluations.

Comparison of RMSE Evaluation with Baseline Methods. We investigate the accuracy of our algorithm in comparison with UCLAF and UBCF using RMSE evaluation. We run each method 5 times for various percentages of test dataset and report the mean values and standard deviations in Table 2.

Method	RMSE@50%	RMSE@40%	RMSE@30%
UBCF	0.009962 ± 0.001	0.008146 ± 0.003	0.006443 ± 0.001
UCLAF	0.00824 ± 0.001	0.006889 ± 0.002	0.00423620 ± 0.001
STS Location Recommender	0.005511 ± 0.002	0.003567 ± 0.001	0.002252 ± 0.001

Table 2. Root Mean Square Error Evaluation

The results obtained show that STS Location Recommender outperforms both UCLAF and UBCF for all percentages of test data. Note that UCLAF also performs better than UBCF. This can be attributed to the fact that UBCF is purely based on conventional user-based collaborative filtering that relies only on user location features. UCLAF on the other hand employs additional information mentioned earlier to improve accuracy of recommendation. STS Location Recommender outperforms both approaches because incorporating temporal dimension of users' location histories significantly impacts positively on the accuracy our location recommendation. The results underscore the importance of temporal information as an additional factor for location recommendation.

5 Conclusions

In this work, we demonstrate how to model temporal and semantically meaningful user mobility behaviours using GPS trajectories. We also show how to leverage spatial, temporal and semantic information to estimate users' preference for location. Finally, using our proposed STS Location Recommender algorithm, we demonstrate how we achieve location recommendation based on our proposed mobility model. Through experimental evaluation, we show that our approach improves location recommendation compared with the baseline approaches.

References

1. Zheng, V.W., Cao, B., Zheng, Y., Xie, X., Yang, Q.: Collaborative filtering meets mobile recommendation: a user-centered approach. In: Proceedings of the 24rd AAAI Conference on Artificial Intelligence, pp. 236–241, (2010)
2. Zheng, Y., Zhang, L., Xie, X., Ma, W.-Y.: Mining interesting locations and travel sequences from GPS trajectories. In: Proceedings of the 18th International Conference on World Wide Web, pp. 791–800 (2009)
3. Yuan, Q., Cong, G., Ma, Z., Sun, A., Thalmann, N.M.: Time-aware point-of-interest recommendation. In: Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 363–372 (2013)
4. Ying, J.J.-C., Lee, W.-C., Tseng, V.S.: Mining geographic-temporal-semantic patterns in trajectories for location prediction. ACM Trans. Intell. Syst. Technol. **5**(1), 2:1–2:33 (2014)
5. Ye, M., Yin, P., Lee, W.-C., Lee, D.-L.: Exploiting geographical influence for collaborative point-of-interest recommendation. In: Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 325–334 (2011)
6. Zheng, V.W., Zheng, Y., Xie, X., Yang, Q.: Collaborative location and activity recommendations with GPS history data. In: Proceedings of the 19th International Conference on World Wide Web, pp. 1029–1038 (2010)
7. Takeuchi, Y., Sugimoto, M.: CityVoyager: an outdoor recommendation system based on user location history. In: Ma, J., Jin, H., Yang, L.T., Tsai, J.J.-P. (eds.) UIC 2006. LNCS, vol. 4159, pp. 625–636. Springer, Heidelberg (2006)
8. Ying, J.J.-C., Lu, E.H.-C., Lee, W.-C., Weng, T.-C., Tseng, V.S.: Mining user similarity from semantic trajectories. In: Proceedings of the 2Nd ACM SIGSPATIAL International Workshop on Location Based Social Networks, pp. 19–26 (2010)
9. Ankerst, M., Breunig, M.M., Kriegel, H.-P., Sander, J.: OPTICS: Ordering points to identify the clustering structure. In: Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data, vol. 28, pp. 49–60 (1999)
10. Ye, M., Shou, D., Lee, W.-C., Yin, P., Janowicz, K.: On the semantic annotation of places in location-based social networks. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 520–528 (2011)

11. Xiao, X., Zheng, Y., Luo, Q., Xie, X.: Finding similar users using category-based location history. In: Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 442–445 (2010)
12. Chon, Y., Lane, N.D., Li, F., Cha, H., Zhao, F.: Automatically characterizing places with opportunistic crowdsensing using smartphones. In: Proceedings of the 2012 ACM Conference on Ubiquitous Computing, UbiComp 2012, pp. 481–490 (2012) doi:10.1145/2370216.2370288
13. Alvares, L.O., Bogorny, V., Kuijpers, B., de Macelo, J., Moelans, B., Palma, A.T.: Towards semantic trajectory knowledge discovery, Data Mining and Knowledge Discovery
14. Spertus, E., Sahami, M., Buyukkokten, O.: Evaluating similarity measures: a large-scale study in the orkut social network. In: KDD 2005: Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining, pp. 678–684 (2005)
15. Herlocker, J.L., Konstan, J.A., Borchers, A., Riedl, J.: An algorithmic framework for performing collaborative filtering. In: Proceedings of the 22Nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 230–237 (1999)