Modeling Temporal Effects of Human Mobile Behavior on Location-Based Social Networks

Huiji Gao, Jiliang Tang, Xia Hu, and Huan Liu
Computer Science and Engineering, Arizona State University, Tempe, AZ, USA
{huiji.gao, jiliang.tang, xia.hu, huan.liu}@asu.edu

ABSTRACT

The rapid growth of location-based social networks (LBSNs) invigorates an increasing number of LBSN users, providing an unprecedented opportunity to study human mobile behavior from spatial, temporal, and social aspects. Among these aspects, temporal effects offer an essential contextual cue for inferring a user's movement. Strong temporal cyclic patterns have been observed in user movement in LBSNs with their correlated spatial and social effects (i.e., temporal correlations). It is a propitious time to model these temporal effects (patterns and correlations) on a user's mobile behavior. In this paper, we present the first comprehensive study of temporal effects on LBSNs. We propose a general framework to exploit and model temporal cyclic patterns and their relationships with spatial and social data. The experimental results on two real-world LBSN datasets validate the power of temporal effects in capturing user mobile behavior, and demonstrate the ability of our framework to select the most effective location prediction algorithm under various combinations of prediction models.

Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social and Behavioral Sciences

Keywords

Location-Based Social Networks; Location Prediction; Temporal Effect; Human Mobile Behavior

1. INTRODUCTION

The wide use of mobile devices has greatly enriched users' urban experience and promoted the development of location-based social services in recent years. Typical location-based social networking sites (e.g., Foursquare[1] and Facebook Places[2]) have attracted billions of users around the world and generated massive location-based social network data, providing us with both opportunities and

----

[1] http://foursquare.com/

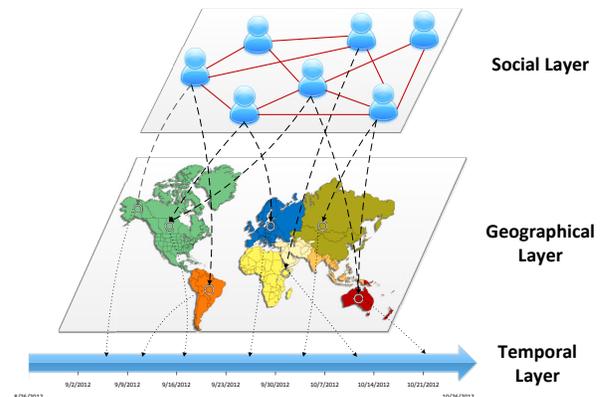[2] http://www.facebook.com/about/location/

Figure 1: The Information Layout of LBSNs

challenges for investigating a user's mobile behavior, with the purpose of designing more advanced location-based services such as location-based marketing [9] and disaster relief [3].

Location-based social network data contain three distinct information layers: a social layer, a geographical layer, and a temporal layer, as shown in Figure 1. The social layer consists of social friendships, the geographical layer displays historical check-ins of users, and the temporal layer indicates temporal stamps of each check-in action. The availability of multiple information sources presents various views to study a user's mobile behavior from spatio-temporal, social-temporal, socio-spatial, and spatial-temporal-social aspects. Previous research studied the social and spatial layers on LBSNs in terms of social-historical ties [4], social-spatial properties [10], geographical influence [14], and "geo-social" correlations [5], etc., while the temporal layer in terms of temporal effects has been rarely studied to model user mobile behavior on LBSNs.

In this paper, we aim to present a comprehensive study of temporal effects on LBSNs to model user mobile behavior. The temporal layer on LBSNs is usually leveraged as an order indicator to connect check-ins chronologically for generating location trajectories [15], which has not been fully exploited. As observed in [13, 8], human movement exhibits strong temporal cyclic patterns in terms of hour of the day and day of the week. For example, a user regularly goes to a restaurant for lunch around 12:00 pm, watches a movie on Friday night, and shops during weekends. Previous work has leveraged these patterns as features under supervised learning to solve location prediction problem [1]. Since human movement is observed as a stochastic process [7], the temporal features generated from a user's movement become very sparse in the large temporal feature space, while the unobserved features severely affect the prediction performance. Due to the temporal continuity of human movement, it is possible to infer an unobserved temporal

feature with observed temporal patterns. For example, a temporal pattern of visiting a restaurant at 10:00 am and 12:00 pm could imply the user's potential presence at that restaurant at 11:00 a m, although such presence at 11:00 am is unobserved in the user's previous movement. This inspires us to exploit these temporal cyclic patterns for modeling a user's temporal preferences.

In addition, temporal cyclic patterns strongly correlate to the spatial and social layers. Firstly, they imply the chronological information for generating location trajectories, which form a user's spatial context [6]. Secondly, social correlation suggests that human movement is usually affected by their social context, such as having lunch with friends at noon. Therefore, to model the temporal effects on LBSNs, we inevitably need to consider two aspects: (1) **temporal preferences** in terms of cyclic patterns; and (2) **temporal correlations** in terms of temporal-spatial correlations and temporal-social correlations.

In this paper, we model temporal effects of user mobile behavior on location-based social networks in terms of the above two aspects. To the best of our knowledge, this work presents the first comprehensive study of temporal effects on LBSNs. The contributions of our work are the following:

- We propose a general framework to study temporal effects in terms of temporal preferences and temporal correlations on location-based social networks.

- We model the temporal cyclic patterns to capture a user's mobile behavior through check-in data, and investigate their correlations to the spatial context and social context.

- We use location prediction to evaluate the temporal effects on two real-world location-based social network datasets. The results validate the power of temporal effects on a user's mobile behavior.

2. A FRAMEWORK FOR MODELING USER MOBILE BEHAVIOR

To investigate the temporal effects, we first propose a general framework to model a user's mobile behavior w.r.t. the temporal cyclic patterns and temporal correlations. Let $\mathcal{U} = \{u_1, u_2, ..., u_n\}$ be the set of users and $\mathcal{L} = \{l_1, l_2, ..., l_m\}$ be the set of locations where n and m are the numbers of users and locations, respectively. Each check-in action is represented as a tuple $\langle u_i, l_j, t_k \rangle \in C$, indicating user $u_i \in \mathcal{U}$ checks in at location $l_j \in \mathcal{L}$ at time t_k, where C is the observed check-in set. Let $\mathcal{F}(u)$ denote u's social friends, $H_{u,t} = \{\langle u_i, l_j, t_k \rangle | \langle u_i, l_j, t_k \rangle \in C, u_i = u, t_k < t\}$ be the observed historical check-in actions of u before time t, and $S_{u,t} = \{\langle u_i, l_j, t_k \rangle | \langle u_i, l_j, t_k \rangle \in C, u_i \in \mathcal{F}(u), t_k < t\}$ be the observed check-in actions of u's friends before time t. t_k and t are all represented by standard time format "YYYY-MM-DD hh:mm:ss am/pm", e.g., "2013-02-22 11:09:59 pm".

Figure 2 shows a user u's mobile behavior (represented by "?") at time t w.r.t. his/her personal check-in history $H_{u,t}$ and friends' check-in history $S_{u,t}$. Given the corresponding observations of $H_{u,t}$ and $S_{u,t}$, the probability distribution over the check-in locations of u at time t is governed by the following formula:

$$P(c_u = l | t, H_{u,t}, S_{u,t}), \quad (1)$$

where c_u denotes the check-in location of user u. Various temporal information related to cyclic patterns can be implied by t (e.g., "2013-02-22 11:09:59 pm") to indicate a user's check-in state, such as a specific hour of the day (11:00 pm), a day of the week (Friday), or a month of the year (February), etc. We use **temporal**
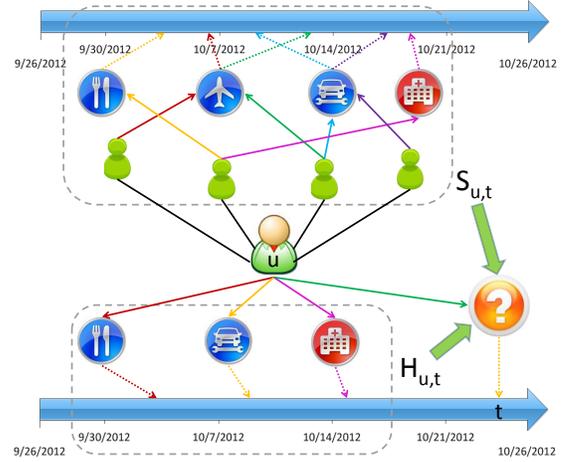

Figure 2: Mobile Behavior of User u at Time t

state to represent such information and introduce $r(t)$ to denote the temporal state extracted from the time t. Depending on the type of temporal state, $r(t)$ can be a different function. For example, if $r(t)$ denotes temporal state in terms of hour of the day, then $r(t) \in \{0, 1, ..., 23\}$, with $r(t) = 2$ indicating that the temporal state is 2 am. If $r(t)$ denotes temporal state in terms of day of the week, then $r(t) \in \{0, 1, ..., 6\}$, with $r(t) = 2$ indicating that the temporal state is Wednesday. Without loss of generality, we use $r(t)$ to denote a type of temporal state in the following description. Eq.(1) is then reformulated as

$$P(c_u = l | r(t), H_{u,t}, S_{u,t}). \quad (2)$$

Applying Bayes' rule, we reformulate Eq. (2) and obtain

$$P(c_u = l | r(t), H_{u,t}, S_{u,t})$$
$$\propto P(r(t) | c_u = l, H_{u,t}, S_{u,t}) P(c_u = l | H_{u,t}, S_{u,t}). \quad (3)$$

The above equation decomposes a user's check-in probability into two components w.r.t. spatial and temporal context, with each context associated with the observed $H_{u,t}$ and $S_{u,t}$. $P(c_u = l | H_{u,t}, S_{u,t})$ is the spatial context indicating the **location distribution** of u's check-in given his/her personal check-in history and friends' check-in history. $P(r(t) | c_u = l, H_{u,t}, S_{u,t})$ is the temporal context representing the **temporal state distribution** of u's check-in, given the observed check-in location at l with the corresponding $H_{u,t}$ and $S_{u,t}$.

Various approaches have been proposed to model the spatial context $P(c_u = l | H_{u,t}, S_{u,t})$. Since the check-in temporal state $r(t)$ is not observed in spatial context, the chronological information of $H_{u,t}$ and $S_{u,t}$ is commonly utilized to generate location trajectories. For example, the Order-k Markov Model [11] considers only $H_{u,t}$ and generates the probability based on the most frequent location sequential patterns. The Social-Historical Model [4] considers both $H_{u,t}$ and $S_{u,t}$ and generates the probability based on the combination of their weighted location sequential patterns.

Compared to the spatial context, the temporal context has been little exploited in previous research. It builds the nexus between a user's visited location and the corresponding visiting time, and provides context of how likely a user would visit a location at a specific temporal state according to his/her personal preferences and social networks. In this paper, we model it as a probability function w.r.t. the combinational effect of personal preferences from $H_{u,t}$ and social influence from $S_{u,t}$. Following the common assumption of modeling user behavior in online social media [12], we consider the personal preferences and social influence of a user as two inde-

pendent parts and adopt a combinational approach similar to [4]:

$$P(r(t)|c_u = l, H_{u,t}, S_{u,t})$$
$$= \alpha P(r(t)|c_u = l, H_{u,t}) + (1-\alpha)P(r(t)|c_u = l, S_{u,t})$$
$$= \alpha P(r(t)|c_u = l, H_{u,t})$$
$$+ (1-\alpha)\frac{\sum_{u_i \in \mathcal{F}(u)} sim(u, u_i)P(r(t)|c_u = l, H_{u_i,t})}{\sum_{u_i \in \mathcal{F}(u)} sim(u, u_i)}, \qquad (4)$$

where α is a parameter that controls the contribution of personal preferences and social influence. $sim(u, u_i)$ is the similarity between u and u_i. In this paper, we compute it as cosine similarity based on the visited locations of u and u_i.

According to Eq.(3) and Eq.(4), we obtain the framework of modeling temporal effects of user mobile behavior:

$$P(c_u = l|r(t), H_{u,t}, S_{u,t})$$
$$\propto P(c_u = l|H_{u,t}, S_{u,t})(\alpha P(r(t)|c_u = l, H_{u,t})$$
$$+ (1-\alpha)\frac{\sum_{u_i \in \mathcal{F}(u)} sim(u, u_i)P(r(t)|c_u = l, H_{u_i,t})}{\sum_{u_i \in \mathcal{F}(u)} sim(u, u_i)}) \qquad (5)$$

In the above framework, $P(r(t)|c_u = l, H_{u,t})$ plays an important role as an elemental distribution. It captures a user u's temporal preferences, and generates the social context $P(r(t)|c_u = l, S_{u,t})$ by incorporating social friend u_i's temporal preferences $P(r(t)|c_u = l, H_{u_i,t})$. The temporal preferences and social context are correlated through Eq.(4) forming a user's temporal context, which is correlated to the spatial context through Eq.(5). These temporal preferences and correlations form the temporal effects on LBSNs, resulting in three major aspects that we study in this paper:

- **Temporal Preferences**
  $P(r(t)|c_u = l, H_{u,t})$ represents the effects of temporal preferences. It captures the temporal state distribution of a user's check-in at a location based on the observation of his/her historical check-ins.

- **Temporal-Social Correlations**
  Eq.(4) presents the temporal-social correlations between a user and his/her friends. It fuses the personal check-in preferences from a user and his/her social friends, providing the perspective of investigating how likely a user's temporal check-in preferences would be affected by his/her social friends.

- **Temporal-Spatial Correlations**
  Eq.(5) combines a user's temporal context $P(r(t)|c_u = l, H_{u,t}, S_{u,t})$ and spatial context $P(c_u = l|H_{u,t}, S_{u,t})$ together, corresponding to the correlations between temporal cyclic information and chronological information. It provides a mathematical analysis on how likely a user would visit a location w.r.t. his/her recently visited locations and visiting time.

3. MODELING TEMPORAL EFFECTS ON LOCATION-BASED SOCIAL NETWORKS

In the above section, we have modeled a user's mobile behavior and presented the importance of the temporal distributions $P(r(t)|c_u(t) = l, H_{u,t})$ in capturing a user's temporal preferences and connecting spatial and social contexts. In this section, we discuss how to model this distribution $P(r(t)|c_u(t) = l, H_{u,t})$. We first analyze two different types of temporal cyclic patterns from a user's check-in history, followed by the modeling of these patterns and then discuss how to integrate them together.

3.1 Analyzing Temporal Cyclic Patterns

Previous work has discovered that human mobility exhibits strong temporal cyclic patterns, and suggested that daily patterns (hour of the day) and weekly patterns (day of the week) are the two most fundamental patterns in reflecting a user's mobile behavior [13, 8]. It is also reported as a mobile property that a user mostly visits a location during one or more specific periods of time, while rarely visiting it during other time periods [6]. Figure 3 plots a user's daily and weekly check-in distribution at a location l from our dataset. Each point in Figure 3(a) and Figure 3(b) represents the total number of check-ins that occurred at a specific hour of the day (day of the week) at location l by that user, respectively. We can observe the phenomenon of "check-in probability centering on certain time periods and decreasing during other time periods" from this figure, which is consistent with the above property.
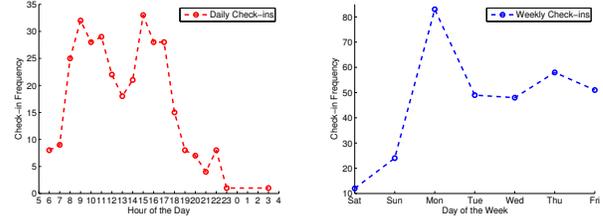
(a) Daily Check-in Frequency (b) Weekly Check-in Frequency

Figure 3: Daily and Weekly Check-in Distribution of a User at Location l

To model the temporal preferences $P(r(t)|c_u = l, H_{u,t})$, e.g., how likely the user would visit location l at 16:00 pm, one can simply count the number of a user's previous check-ins at location l at 16:00 pm, and compute its ratio to the total number of his/her check-ins at location l through Figure 3. However, users usually do not have check-ins at each hour of the day (day of the week) on a location. For example, there are no check-ins from hour 0 to hour 2 in Figure 3(a). It would be arbitrary to determine that the probability of his/her check-in during that time is zero, as the visit time of a user at the same location tends to be within a user-biased time period. Therefore, a smoothing technique is inevitable to evaluate the probability of a check-in at unvisited time.

3.2 Modeling Temporal Cyclic Patterns

Mathematically, we need a probability distribution to model a user's temporal preferences at a location $P(r(t) = q|c_u = l, H_{u,t})$. Such distribution should satisfy the following requirements: (1) probability distribution centers on one or more temporal states; (2) probability decreases as the distance to the center point increases; and (3) each user has a biased probability decreasing speed around a center. Among various distributions, Gaussian mixture distribution is such a distribution capturing these properties. Thus, in this paper, we utilize the Gaussian mixture model (**GMM**) to capture a user's temporal preferences. The hypothesis is that a user u's visiting time at a location l is a stochastic process centered around several time points, as shown below:

$$P(r(t)|c_u = l, H_{u,t}) \sim \sum_{i=1}^{k} A_i \mathcal{N}(r(t)|\mu_{u,l}^i, \sigma_{u,l}^i) \qquad (6)$$

where k is the number of centers considered to model $P(r(t)|c_u = l, H_{u,t})$. A_i controls the maximum power of Gaussian distribution centered on the i-th center. $\mu_{u,l}^i$ and $\sigma_{u,l}^i$ are the corresponding mean and variance.

Each user may have different temporal preferences on different locations, resulting in different numbers of centers in the above

distribution. To determine the value of k, we design a center detection strategy to compute the number of possible centers in the temporal state distribution on each observed user-location pair. A temporal state is considered as a center as long as the check-in frequency on this temporal state is higher than that of the previously observed temporal state and the next observed temporal state. For a user-location pair with less than three observed temporal states, we consider it a single center case whose center corresponds to the temporal state with the highest check-in frequency.

We applied the center detection strategy to our dataset, and observed that *more than 90% of daily and weekly patterns have less than two centers*, indicating that $k = 2$ represents the majority. In addition, due to the data sparseness, the daily pattern has a maximum of 24 input points for training, while the weekly pattern has 7 points at most, which is insufficient to accurately model a complicated process with too many center points. Therefore, in this paper, we select the number of centers k as two for modeling temporal preferences and take the gradient descent method to solve Eq. (6).

3.3 Integrating Temporal Cyclic Patterns

In previous sections, we modeled daily and weekly patterns of a user's mobile behavior with Gaussian mixture distribution. In this section, we discuss how to integrate the two temporal cyclic patterns together. We first formally define $r(t)$ in terms of daily and weekly temporal states. We define $r(t) = \{r_d(t), r_w(t)\}$ as two indicator functions mapping the time stamp for each type of time state, where $r_d(t) \in \{0, 1, ...23\}$ is the hour of the day (e.g., 10:00 am), indicating the daily temporal state. $r_w(t) \in \{0, 1, ..., 6\}$ is the day of the week (e.g., Monday), indicating the weekly temporal state. Assume the time stamp t = "2012-11-24 12:30:00pm", then $r_d(t)$=12 (12:00 pm) and $r_w(t)$=5 (Saturday). We follow the independence assumption [6] of daily and weekly patterns and obtain

$$
\begin{aligned}
&P(r(t)|c_u = l, H_{u,t}) \\
&= P(r_d(t), r_w(t)|c_u = l, H_{u,t}) \\
&= P(r_d(t)|c_u = l, H_{u,t})P(r_w(t)|c_u = l, H_{u,t}), \quad (7)
\end{aligned}
$$

where $P(r_d(t)|c_u = l, H_{u,t})$ indicates the probability of the check-in happening at hour $r_d(t)$ given the observation that the check-in actually occurred at location l. Similarly, $P(r_w(t)|c_u = l, H_{u,t})$ indicates the probability of the check-in happening at day $r_w(t)$ given the observation that the check-in actually occurred at location l.

4. EXPERIMENTS

In this section, we evaluate the significance of temporal effects with our proposed models under the general framework on two real-world location-based social network datasets. We use location prediction as an application for evaluation. Since we focus on investigating the temporal effects of user mobile behavior, we perform the evaluation in two stages: (1) we evaluate the models based on temporal preferences, temporal-social correlations, and temporal-spatial correlations, respectively, to see their ability in improving location prediction performance; and (2) we compare the selected location prediction models from our framework with a state-of-the-art location prediction method, which utilizes spatial, temporal and social information in order to give a general idea of the effectiveness and reasonableness of our framework.

4.1 Datasets and Experiment Setup

We use two publicly available datasets from two location-based social networking sites, i.e., Foursquare[3] and Brightkite[4], to evalu-

Table 1: Statistical Information of Two Datasets

	Brightkite	Foursquare
Duration	03/2008-10/2010	03/2010-01/2011
# of Users	26,915	18,107
# of Check-ins	4,666,732	2,073,740
# of Unique Locations	751,176	43,063
# of Links	261,982	231,148
# of Test Check-ins	134,575	90, 935

ate the temporal effects of user mobile behavior. Both sites allow a user to check-in at a physical location through his/her cellphone and then let his/her online friends know where he/she is. In both datasets, the friendships are undirected.

We select users who have at least 10 check-ins, and obtain 26,915 and 18,107 users on each dataset. For each user, we randomly sample five of his/her total check-ins as test check-ins. The statistic information of these two datasets is listed in Table 1. For each test check-in from a user u, we consider its check-in time t, the user's historical check-ins before t ($H_{u,t}$), and the check-ins from u's friends before t ($S_{u,t}$) as observed data. We then predict the test check-in location based on models of temporal preferences, temporal social-correlations, and temporal-spatial correlations, respectively. We use **prediction accuracy**, i.e., the ratio of the number of accurately predicted check-ins to the total number of test check-ins, to evaluate the prediction performance.

For the **temporal preference** model, we use **GMM** to compute the temporal state probability $P(r(t)|c_u = l, H_{u,t})$, as discussed in Section 4. We rank all the locations observed in $H_{u,t}$ based on their $P(r(t)|c_u = l, H_{u,t})$ with respect to the temporal state of test check-in time t and report the top ranked location for prediction. For the **temporal-social correlation** model, we first apply **GMM** to compute $P(r(t)|c_u = l, H_{u_i,t})$ for each user u's friend u_i and then obtain the $P(r(t)|c_u = l, S_{u,t})$ and $P(r(t)|c_u = l, H_{u,t}, S_{u,t})$ through Eq.(4) for location prediction. Finally, for the **temporal-spatial correlation** model, we compute the check-in probability $P(c_u = l|r(t), H_{u,t}, S_{u,t})$ through Eq.(5) with the **temporal-social correlation** model and existing spatial models.

4.2 Temporal Preferences

In this section, we evaluate the effectiveness of temporal preferences in terms of (1) whether applying smooth technology can capture a user's temporal preferences more accurately; and (2) how do different temporal cyclic patterns affect a user's temporal preferences. Table 2 shows the detailed prediction performance, where "Daily" and "Weekly" indicate the type of patterns used in Eq. (7). "Daily_Weekly" represents the combination of the two patterns through Eq.(7). The non-smooth approaches simply consider the ratio of a user's previous check-ins on l at $r(t)$:

$$
\begin{aligned}
&P(r(t)|c_u = l, H_{u,t}) \\
&= \frac{|\{\langle u_i, l_j, t_k\rangle|\langle u_i, l_j, t_k\rangle \in H_{u,t}, l_j = l, r(t_k) = r(t)\}|}{|\{\langle u_i, l_j, t_k\rangle|\langle u_i, l_j, t_k\rangle \in H_{u,t}, l_j = l\}|} \quad (8)
\end{aligned}
$$

We summarize the essential observations from Table 2:

- Approaches with **GMM** based on Gaussian smooth all perform better than non-smooth approaches, demonstrating the property of "**check-in probability centering on certain time periods and decreasing during other time periods**", which explains a user's mobile behavior to a certain extent.

- The non-smooth approaches obtain much better performance compared to the **random guess** approach (approximately **3.84%** and **4.49%** accuracy) on two datasets. This proves that users

Table 2: Location Prediction with Temporal Preferences

	Patterns	Brightkite	Foursquare
Non-Smooth	Daily	0.2221	0.1526
	Weekly	0.2156	0.1345
	Daily_Weekly	0.2485	0.1875
GMM	Daily	0.2949	0.2084
	Weekly	0.2721	0.1729
	Daily_Weekly	**0.3360**	**0.2614**

on LBSNs do exhibit temporal cyclic patterns, and such patterns are helpful in capturing a user's temporal preferences.

- The daily pattern and weekly pattern both capture a user's temporal preferences. Their integration outperforms each individual one, indicating that these two patterns contain complementary information to explain a user's mobile behavior.

In sum, both daily and weekly cyclic patterns observed from a user's check-in history can be utilized for improving location prediction performance. By exploiting their relationship to the unobserved cyclic patterns through smoothing, we are able to capture a user's mobile behavior more accurately. In addition, the combination of daily and weekly patterns can better explain a user's temporal preferences than each pattern.

4.3 Temporal-Social Correlations

In this section, we investigate the effectiveness of temporal-social correlations. We are interested in how a user's mobile behavior on a location is affected by his/her temporal preferences and his/her friends' temporal preferences. According to the temporal-social correlation model in Eq.(4), α is a parameter to control the contribution of personal preferences and social influence. Therefore, to investigate the temporal-social correlations, we consider three variants of the model:

- **Soc**: Setting $\alpha = 0$, **Soc** considers friends' temporal preferences only, corresponding to $P(r(t)|c_u = l, S_{u,t})$.

- **Tmp**: Setting $\alpha = 1$, **Tmp** considers personal temporal preferences only, corresponding to $P(r(t)|c_u = l, H_{u,t})$.

- **Tmp-Soc**: Increasing α from 0 to 1 with an increment step of 0.01, observing the location prediction performance at each α. **Tmp-Soc** reports the best performance, corresponding to considering both personal preferences and social influence.

The results are reported in Table 3. Due to the space limit, we only present the results on Brightkite, as similar results are observed on Foursquare. We summarize the essential observations below:

- **Soc** performs better than the random guess approach. It achieves, on average, **94.97% relative improvement over the random guess approach**, suggesting that a user's temporal preferences do have correlations to his/her friends.

- **Tmp-Soc** performs slightly better than **Tmp**. According to social correlation, similar users tend to become friends and social friends tend to behave similarly, suggesting that two friends may share more common preferences. This explains the small improvement which may be due to the overlapping of temporal preferences between a user and his/her friends.

- The best performance in **Tmp-Soc** is obtained with α around 0.8, indicating the different contributions of personal preferences and social networks on a user's mobile behavior, which is also consistent with the observations in [4].

Table 3: Location Prediction with Temp-Social Correlations

	Pattern	Soc	Tmp	Tmp-Soc
Brightkite	Daily	0.0761	0.2949	0.2982
	Weekly	0.0751	0.2721	0.2724
	Daily_Weekly	0.0734	0.3360	0.3389

4.4 Temporal-Spatial Correlations

In this section, we investigate the effectiveness of temporal-spatial correlations in Eq. (5), which also corresponds to our framework. We first present a set of state-of-the-art approaches for modeling spatial context $P(c_u = l|H_{u,t}, S_{u,t})$ and then evaluate the temporal-spatial correlations model.

4.4.1 Spatial Models

We introduce three state-of-the-art spatial models, which utilize location trajectory patterns for location prediction.

- Most Frequent Check-in Model (**MFC**)
  The Most Frequent Check-in Model assigns the probability of a user u checking in at a location l as the probability of l appearing in u's check-in history; social information is not considered in this approach.

- Order-1 Markov Model (**OMM**)
  The Order-1 Markov Model [11] considers the latest check-in location as context and searches for frequent patterns to predict the next location. Social information is usually not considered in this model.

- Social Historical Model (**SHM**)
  The Social Historical Model [4] utilizes the Hierarchical Pitman-Yor language model to capture the n-gram location sequential patterns for location prediction w.r.t. a user's social and historical ties.

4.4.2 Performance of Temporal-Spatial Correlations

Table 4 lists the location prediction performance with various spatial-temporal correlations. Each cell represents the prediction accuracy with the correlated models from the corresponding row and column, where "**Temporal Only**" indicates that no spatial context is used in Eq. (5), corresponding to the temporal-social correlation model **Tmp-Soc** in Table 3. "**Spatial Only**" indicates no temporal context is used, corresponding to the above spatial models. The percentage listed next to the accuracy represents the relative improvement over the corresponding spatial model. We summarize several observations below:

- The temporal-spatial correlation models consistently outperform the corresponding **spatial only** models. For example, the proposed temporal-spatial correlation models have, on average, 8.26%, 3.70%, and 4.25% relative improvement over MFC, OMM, and SHM, respectively, on Brightkite. Similar improvements can also be observed on Foursquare data. Considering the low accuracy of random guess approach on two datasets, this improvement is actually significant. The results also indicate that temporal cyclic patterns provide complementary information to spatial context and are helpful in improving the prediction accuracy.

- The **spatial only** models perform better than the proposed **temporal only** models, indicating the importance of chronological information in explaining a user's mobile behavior. The results suggest that cyclic information may not be as valuable as chronological information to model a user's mobile behavior, while the latter contains more helpful information to improve the location prediction accuracy.

Table 4: Location Prediction with Temporal-Spatial Correlations

		Temporal Only	MFC	OMM	SHM
Brightkite	Spatial Only	–	0.3377	0.3244	0.3950
	Daily	0.2982	0.3651 (+8.11%)	0.3374 (+4.01%)	0.4127 (+4.48%)
	Weekly	0.2724	0.3411 (+1.01%)	0.3309 (+2.00%)	0.3962 (+0.30%)
	Daily_Weekly	0.3389	0.3906 (+15.66%)	0.3409 (+5.09%)	0.4265 (+7.97%)
Foursquare	Spatial Only	–	0.2706	0.2498	0.3120
	Daily	0.2123	0.2888 (+6.73%)	0.2591 (+3.72%)	0.3276 (+5.00%)
	Weekly	0.1781	0.2810 (+3.84%)	0.2535 (+1.48%)	0.3213 (+2.98%)
	Daily_Weekly	0.2646	0.3140 (+16.04%)	0.2601 (+4.12%)	0.3423 (+9.71%)

Table 5: Comparison of SHM+T and PSMM

	SHM+D	SHM+W	SHM+DW	PSMM
Brightkite	0.4127	0.3962	0.4265	0.3874
Foursquare	0.3276	0.3213	0.3423	0.3127

- The performance of the combined model is highly related to the prediction ability of each individual model. For example, a strong-strong combination (e.g., Daily_Weekly and SHM) performs better than a strong-weak combination (e.g., Daily_Weekly and OMM), while the latter is better than a weak-weak combination (e.g., Weekly and OMM). This provides an opportunity to **select the most effective prediction algorithm from various model combinations** by evaluating their performance through our framework.

4.4.3 Comparison to State-of-the-Art Prediction Model

To summarize, our framework provides a way to evaluate temporal effects on various location prediction models. The empirical comparison indicates that SHM, considering temporal information (denoted as SHM+T), performs the best. Since SHM+T takes into account temporal, spatial and social information, we would like to further investigate *how it fares in comparison with the state-of-the-art location prediction approach* which also considers temporal, spatial and social information.

PSMM [2] is a state-of-the-art location prediction method in LBSNs that leverages spatial, temporal and social information for location prediction. It also utilizes Gaussian distribution; however, different from our work, it models spatial patterns instead of temporal patterns as Gaussian distribution. We compare SHM+T with **PSMM** on our two datasets and the results are shown in Table 5, where D (daily) and W (weekly) denote the type of temporal information considered. The experiment results show that SHM+T performs consistently better that **PSMM**. According to the different types of cyclic patterns used in SHM+T, it achieves 6.30% and 5.66% relative improvement, on average, over **PSMM** on Brightkite and Foursquare, respectively. This indicates that the prediction models selected from our framework could be very effective in capturing a user's mobile behavior.

5. CONCLUSION AND FUTURE WORK

In this paper, we model the temporal effects of user mobile behavior on LBSNs with respect to two aspects: temporal preferences and temporal correlations. We study each aspect under a proposed framework which not only models these aspects together but also provides the ability to select the most effective location prediction algorithm. Our experimental results show that a user's mobile behavior is affected by various temporal cyclic patterns whose distribution can be modeled as a Gaussian mixture distribution and therefore captures a user's temporal preferences more accurately. We also observe that a user's temporal preferences are correlated with his social friends with a lot of preference overlapping, and conclude that temporal context is complementary to spatial context in improving location prediction performance. Among various

directions for future work, it would be interesting to consider continuous temporal states for modeling cyclic patterns and explore other types of temporal cyclic patterns.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] J. Chang and E. Sun. Location 3: How users share and respond to location-based data on social networking sites. *AAAI*, 2011.

[2] E. Cho, S. Myers, and J. Leskovec. Friendship and mobility: user movement in location-based social networks. *SIGKDD*, 2011.

[3] H. Gao, G. Barbier, and R. Goolsby. Harnessing the crowdsourcing power of social media for disaster relief. *Intelligent Systems, IEEE*, 26(3):10–14, 2011.

[4] H. Gao, J. Tang, and H. Liu. Exploring social-historical ties on location-based social networks. *ICWSM*, 2012.

[5] H. Gao, J. Tang, and H. Liu. gSCorr: Modeling geo-social correlations for new check-ins on location-based social networks. *CIKM*, 2012.

[6] H. Gao, J. Tang, and H. Liu. Mobile location prediction in spatio-temporal context. *Nokia Mobile Data Challenge Workshop*, 2012.

[7] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.

[8] E. Malmi, T. Do, and D. Gatica-Perez. Checking in or checked in: Comparing large-scale manual and automatic location disclosure patterns. *MUM*, 2012.

[9] A. Research. Mobile advertisers forecast to spend $1.8 billion on location-based campaigns in 2015. 2010.

[10] S. Scellato, A. Noulas, R. Lambiotte, and C. Mascolo. Socio-spatial properties of online location-based social networks. *ICWSM*, 11, 2011.

[11] L. Song, D. Kotz, R. Jain, and X. He. Evaluating location predictors with extensive wi-fi mobility data. In *INFOCOM 2004*, volume 2, pages 1414–1424. IEEE, 2004.

[12] J. Tang, H. Gao, H. Liu, and A. D. Sarma. eTrust: Understanding trust evolution in an online world. *KDD*, 2012.

[13] M. Ye, K. Janowicz, C. Mülligann, and W. Lee. What you are is when you are: the temporal dimension of feature types in location-based social networks. *SIGSPATIAL*, 2011.

[14] M. Ye, P. Yin, W. Lee, and D. Lee. Exploiting geographical influence for collaborative point-of-interest recommendation. *SIGIR*, 2011.

[15] Y. Zheng, X. Xie, and W.-Y. Ma. Geolife: A collaborative social networking service among user, location and trajectory. *IEEE Data Engineering Bulletin*, 2010.