# Location Prediction Through Activity Purpose: Integrating Temporal and Sequential Models

Dongliang Liao[1(✉)], Yuan Zhong[2], and Jing Li[1]

[1] University of Science and Technology of China, Hefei, China
liaodl@mail.ustc.edu.cn, lj@ustc.edu.cn
[2] Northeastern University, Boston, MA, USA
yzhong@ccs.neu.edu

**Abstract.** Based on the growing popularity of smart mobile devices, location-aware services become indispensable in human daily life. Location prediction makes these services more intelligent and attractive. However, due to the limited energy of mobile devices and privacy issues, the captured mobility data is typically sparse. This inherent challenge deteriorates significant principles in mobility modeling, i.e. temporal regularity and sequential dependency. To tackle these challenges, by utilizing temporal regularity and sequential dependency, we present a location prediction model with a two-stage fashion. Firstly, it extracts predictive features to effectively target the better performer from sequential and temporal models. Secondly, according to the inferred activity, it adopts non-parametric Kernel Density Estimation for posterior location prediction. Extensive experiments on two public check-in datasets demonstrate that the proposed model outperforms state-of-the-art baselines by 10.1% for activity prediction and 12.9% for location prediction.

**Keywords:** Location prediction · Activity prediction · Mobility modeling · Context-Aware Hybrid approach · Kernel Density Estimation

## 1 Introduction

With the ubiquity of smart mobile devices and the development of positioning technology, an overwhelming number of location-aware services have gained increasing popularity in recent years. These services have offered an unprecedented opportunity for both academia and industry to study human mobility behavior with access to various kinds of data, such as GPS trajectories, WiFi records, cellular phone logs, smart card transactions and social network check-ins, etc. They also shed light on a myriad of potential applications like user profiling, location understanding, urban planning and mobility modeling [11,19].

Among them, location prediction plays a key role. Generally, scholars handle this task with two-broad-category approaches, sequential modeling and temporal regularity modeling. Viewing that user activity serves as mobility motivation, activity prediction [5,8,9,16,17] is introduced as auxiliary to reduce vast


J. Kim et al. (Eds.): PAKDD 2017, Part I, LNAI 10234, pp. 711–723, 2017.
DOI: 10.1007/978-3-319-57454-7_55

location candidate space (million magnitude from population level and thousand magnitude from individual level). Unfortunately, many unresolved difficulties remain tough in location prediction: (1) Sensitive to sequential dependency, sequential models [1,3,7,16] deteriorate when the timespan of consecutive mobility records being far like days or even months [3,15]. This is often the case due to device energy limitation and user privacy concern. (2) Temporal model performs poorly at night or during weekends [13] owing to decayed regularity; (3) Even for the same activity purpose, people still conceive different preferences under different contexts. E.g., at midnight, Alice would buy snacks from 7-Eleven near home, instead of Stop&Shop where she usually visits in the daytime. One thing worth noting is that these three problems are non-trivial. Simply take the last example for illustration. Because of data sparsity, directly estimating the user location preference for the specific time is obsessed with under-fitting. In addition, "contexts" are highly diversified and even only for the time context, modeling "location open hour", "user rest period", etc. simultaneously can be overwhelming.

In this paper, we tackle above challenges by decomposing location prediction into two subtasks [5,8,16], user activity inference and location inference based on activity. For activity inference, sequential and temporal models can fit respectively. However, as previously indicated, both are ineffective in certain circumstances. Here we design a Context-Aware Hybrid (CAH) module to integrate temporal regularity and sequential dependency models dynamically. More specifically, a set of elaborate evaluation features (e.g. density of recent records, regularity strength of user historical activities) are extracted as context features and based on that, a supervised classifier is applied to select the better performer between sequential and temporal models. For location inference, we adopt a time-aware approach for posterior location distribution calculation. Technically, instead of employing parameterized models which usually fall into a training dilemma, Kernel Density Estimation is applied to capture the visit time distribution at specific locations. Last but not the least, we summarize these two phases to leverage final location prediction.

Our main contributions are summarized as follows:

1. With a set of features assessing the performance of sequential and temporal models, we develop a Context-Aware Hybrid approach to combine them for user activity prediction.
2. We introduce Kernel Density Estimation to model the time variation of location preference for a given user, and construct a two-stage model to predict future locations based on the inferred activity.
3. The experimental results on two public datasets validate that our model significantly outperforms state-of-the-art baselines in terms of both activity prediction accuracy and locations prediction accuracy.

The rest of paper is structured as follows: Sect. 2 reviews related mobility prediction works. Section 3 formulates the prediction problem and introduces the notations. Our proposed model is presented in Sect. 4. Experimental results based on two real world public datasets are presented in Sect. 5. Finally, the conclusion, limitation and future work outlook are offered in Sect. 6.

## 2 Related Work

### 2.1 Mobility Pattern Model

We categorize relevant mobility prediction models into sequential model, temporal model and hybrid model.

**Sequential Model.** Song et al. [12] found that Order-2 Markov with fallback had the best performance on the location prediction. Applied in mobility prediction by Cheng et al. [1], Factorizing Personalized Markov Chain extended the Markov Chain via factorization of transition matrix. Zhang et al. [18] extracted users' mobility sequential pattern from historical checkins as a Location-Location Transition Graph. The problem of sequential models lies in that when adjacent mobility records gap for a long time like several days or even months, the performance becomes undesirable [1,3].

**Temporal Model.** Cho et al. [2] proposed a time-aware Gaussian Mixture model combining periodic short-range movements and sporadic long-distance travels. Wang et al. [13] provided a Regularity Conformity Heterogeneous (RCH) model to predict user location at specific time, considering both the regularity and conformity. Yang et al. [15] employed a Tensor Factorization model to capture the user temporal activity preference. However, these methods depend heavily on temporal regularity and data with decayed mobility regularity (e.g. at night or during weekends) leads to low accuracy [13].

**Hybrid Model.** Lian et al. [6] incorporated Markov model and temporal regularity model into the hidden Markov framework to predict user regular locations. This method suffered the same drawback as sequential model. Feng et al. [3] developed Personalized Ranking Metric Embedding (PRME) method to balance sequential dependency and user preference, by a threshold of transition timespan. PRME ignored the temporal regularity and a fix threshold cannot satisfy all the scenarios. In contrast to these methods, the proposed CAH approach combine temporal and sequential models flexibly depending on mobility context.

### 2.2 Location Prediction with Activity Information

Some researchers exploited activity information to improve the location predictability [5,8,9,16,17]. Noulas et al. [9] captured factors driving user movements, including the activity preference and activity transition. Yuan et al. [17] came up with a unified model $W^4$ (who, when, where, what) to discover individual mobility behaviors from spatial, temporal and activity aspects. Ye et al. [16], Li et al. [5] and Liu et al. [8] modeled activity sequential pattern, and predicted locations based on above activity. However, none of them absorb temporal and sequential model simultaneously to infer user activity preference.

In addition, given the activity distribution, Yuan et al. [17] and Li et al. [5] assumed the user location preference followed multinomial distribution. Ye et al. [16] ranked locations based on check-in frequency. Liu et al. [8] applied Matrix Factorization to predict user preference of specific locations. However, these methods fail to capture the time variation of user location preference. Instead, we adopt a generative approach to model the time variation pattern.

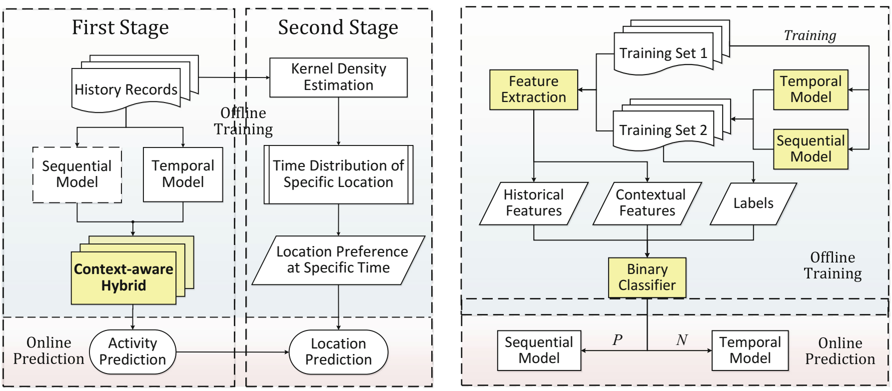
# 3 Problem Formulation

Let $\mathcal{V} = \{v_1, v_2, \ldots, v_{|\mathcal{V}|}\}$ and $\mathcal{C} = \{c_1, c_2, \ldots, c_{|\mathcal{C}|}\}$ represent locations and categories. Each location belongs to a certain category indicating the activity purpose of users. Given a set of users $\mathcal{U}$, each mobility record can be defined as a quadruple $r = (u, v, c, t)$, representing that user $u$ visits location $v$ at time $t$ for activity $c$. Here, for the ease of calculation, $t$ is discretized from continuity to discrete by 24 h. Our goal is to predict user $u$'s next location $\hat{v}$, given the next visit time $\hat{t}$ and the recent visit sequence before $\hat{t}$, $\tau^u_{\hat{t}}$.

# 4 Methodology

## 4.1 Overview

We construct a two-stage model to predict activities and locations. The overall framework of the proposed model is presented in Fig. 1(a). It consists of two stages for activity and location prediction respectively and each stage incorporates offline model training and online prediction. In the first stage, Context-Aware Hybrid (CAH) approach is adopted to dynamically select the better performer from sequential and temporal models for activity prediction, i.e. inferring




(a) Framework of proposed model (b) Context-Aware Hybrid module

**Fig. 1.** Overall framework and CAH module

$P_u(c|\tau^u_{\hat{t}}, \hat{t})$. In second stage, based on inferred user activity, Kernel Density Estimation is exploited to approximate $P_u(v|c, \tau^u_{\hat{t}}, \hat{t})$. Finally, location prediction is achieved by $\hat{v} = \arg\max_v P_u(v|\tau^u_{\hat{t}}, \hat{t})$, where

$$P_u(v|\tau^u_{\hat{t}}, \hat{t}) = \sum_{c_j} P_u(v|c_j, \tau^u_{\hat{t}}, \hat{t}) P_u(c_j|\tau^u_{\hat{t}}, \hat{t}) = P_u(v|c_v, \hat{t}) P_u(c_v|\tau^u_{\hat{t}}, \hat{t}) \tag{1}$$

Note that during the second phrase, we ignore the sequential pattern of location by simplifying $P_u(v|c_v, \tau^u_{\hat{t}}, \hat{t})$ to $P_u(v|c_v, \hat{t})$. The reason is that the sequential dependency of the user mobility has been captured in the first stage. Although the geo-distance may influence user location preference, introducing distance does not significantly improve the prediction performance [6], due to the highly uncertain timespans between adjacent records and the convenient transportation in the modern world.

Figure 1(b) shows specific details of Context-Aware Hybrid module. We partition the training data into training set 1 and training set 2. The former is utilized for learning sequential and temporal models, and the latter is employed to evaluate the performances of them. In this work, we assign Tensor Factorization as temporal model and smoothed Order-1 Markov Chain as sequential model. With features of user contextual and historical factors and labels of the better performer between sequential and temporal models, we build a binary classifier for online prediction.

### 4.2 User Activity Prediction

**Sequential Model.** Markov model has been proved effective in mobility prediction [12]. Due to the data sparsity, we filter the transitions with timespans larger than threshold $\varepsilon$, and merely consider Order-1 Markov Chain. The transition probability is estimated by Kneser-Ney smoothing technique [6]. In particular, let $n^u_\varepsilon(c_i, c_j)$ indicate the times of user $u$ transferring from activity $c_i$ to $c_j$ within $\varepsilon$. The transition probability is derived as:

$$P_u(c_j|c_i) = \frac{\max\{n^u_\varepsilon(c_i, c_j) - \delta, 0\}}{\sum_k n^u_\varepsilon(c_i, c_k)} + \frac{\delta \sum_k \mathbf{I}\{n^u_\varepsilon(c_i, c_k) > 0\} \cdot \sum_k \mathbf{I}\{n^u_\varepsilon(c_k, c_j) > 0\}}{\sum_k n^u_\varepsilon(c_i, c_k) \cdot \sum_k \sum_l \mathbf{I}\{n^u_\varepsilon(c_l, c_k) > 0\}}$$

where $\mathbf{I}\{\cdot\}$ is an indicator function and $\delta$ is the discount parameter. The basic intuition of this equation is to discount the observed times of transition from $c_i$ to $c_j$, and turn them over to low frequency transitions.

**Temporal Model.** We adopt the non-negative Tensor Factorization (TF) method for inferring the activity preference at specific time [15]. A user-time-activity tensor $\mathbf{T} \in \mathbb{R}^{|\mathcal{U}| \times 24 \times |\mathcal{C}|}$ is built, in which the element $\mathbf{T}_{u,h,c}$ equals to the frequency of activity $c$ at hour of day (HOD) $h$ by user $u$. Using Canonical decomposition model [4], $\mathbf{T}$ is decomposed into three matrices, user feature matrix $\hat{U} \in \mathbb{R}^{|\mathcal{U}| \times L}$, time feature matrix $\hat{T} \in \mathbb{R}^{24 \times L}$, and activity feature matrix

$\hat{A} \in \mathbb{R}^{|\mathcal{C}|\times L}$ ($L$ is the latent space dimension). User $u$'s preference of activity $c$ at $h$ could be described as: $Pref_{u,h,c} = \sum_{i=1}^{L}\sum_{j=1}^{L}\sum_{k=1}^{L}\hat{U}_{u,i}\cdot\hat{T}_{h,j}\cdot\hat{A}_{c,k}$. For user $u$, the probability of activity $c$ given time $t$ is formulated as follows, where $t_h$ is HOD of $t$.

$$P_u(c|t) = \frac{Pref_{u,t_h,c}}{\sum_{c'\in\mathcal{C}} Pref_{u,t_h,c'}} \tag{2}$$

**Context-Aware Features Extraction.** Given the Markov and TF models, we extract several kinds of features determining the accuracies of these two models, including temporal contextual features, sequential contextual features and historical features.

*Temporal Contextual Features:* This group of features refer to factors severely affecting temporal model at next visit HOD $h$, i.e. temporal regularity strength and data density at $h$. (1) Temporal regularity strength determines the limit of predictability, measured by *entropy* [11], defined as: $H(Z) = -\sum_i P(z_i)\log P(z_i)$ over random variable $Z$. We introduce random variable $A_h^u$, activity at $h$ of user $u$, whose entropy $H(A_h^u)$ can be calculated based on $u$'s history records $\Gamma_u$. Moreover, the number of distinct activities at $h$ in $\Gamma_u$, correlating with $H(A_h^u)$, is also considered here, signified by $N_a^u(h)$. (2) Data density is represented by $N_r^u(h)$, the number of history records at $h$ of user $u$. In summary, $H(A_h^u)$, $N_a^u(h)$ and $N_r^u(h)$ constitute temporal contextual features.

*Sequential Contextual Features:* The accuracy of sequential model depends on contexts of user recent records, i.e. the timespan and recent record frequency. (1) Timespan feature: As the sequence dependency decays over time, $D_1^u(\hat{t})$, the interval between $\hat{t}$ and nearest record time of user $u$, is introduced to model it. (2) Recent record frequency features: Data sparsity means missing latest activities and reduced performance of sequential model. Thus we propose two features: the length of $S_{\hat{t}}^u$, the longest mobility sequence ending by $\hat{t}$, satisfying that timespans between any adjacent records is less than $\varepsilon$; $D_2^u(\hat{t})$, the timespan between $\hat{t}$ and the earliest record time in $S_{\hat{t}}^u$. In summary, we use $D_1^u(\hat{t})$, $D_2^u(\hat{t})$, $|S_{\hat{t}}^u|$ as sequential contextual features.

*Historical Features:* From the whole mobility historical sequences, we consider user specific features (independent of context) of temporal regularity, sequential dependency and activity regularity strengths. (1) User specific temporal regularity strength is defined by $E_h(N_a^u(h))$ and $E_h(H(A_h^u))$, where $N_a^u(h)$ and $H(A_h^u)$ are defined above, and $E_h(Y) = \sum_{i=1}^{24} P_u(h_i)Y(h_i)$. (2) User specific sequential dependency strength is captured by $E_c(M_a^u(c))$ and $E_c(H(A_c^u))$, where $M_a^u(c)$ is the number of distinct activities of $u$ after activity $c$, $A_c^u$ is a random variable of the activity for user $u$ after activity $c$, and $E_c(Y) = \sum_i P_u(c_i)Y(c_i)$. (3) User specific activity regularity strength is measured by the number of distinct activities $N^u$ and the activity entropy $H(A^u)$ in history records $\Gamma_u$, where $A^u$ is a random variable of the activity for user $u$. In summary, the historical features include $E_h(N_a^u(h))$, $E_h(H(A_h^u))$, $E_c(H(A_c^u))$, $E_c(M_a^u(c))$, $N^u$ and $H(A^u)$.

**Context-Aware Hybrid.** Given the user $u$, the visit sequence $\tau_{\hat{t}}^u$ and the next activity time $\hat{t}$, feature vector $X$ is calculated as mentioned before. We build a binary classifier to target at the better performer between TF and Markov models, taking feature vector $X$ as input. Let positive class represent that Markov model is more effective, then $P_u(c_i|\tau_{\hat{t}}^u, \hat{t})$ is estimated as follows, where $c_n$ is the latest activity in $\tau_{\hat{t}}^u$ and $y$ is the output of classifier:

$$P_u(c_i|\tau_{\hat{t}}^u, \hat{t}) = \begin{cases} P_u(c_i|c_n), & \text{if } y = 1 \\ P_u(c_i|\hat{t}), & \text{if y} = -1 \end{cases} \tag{3}$$

We split user $u$'s history records $\Gamma_u$ into two parts, $\Gamma_u^{(1)}$ for training Markov and TF models, and $\Gamma_u^{(2)}$ for training the classifier. For the record $r : (u, t_r, c_r, v_r)$ in $\Gamma_u^{(2)}$, let $Rank_m(c_r)$ represent the probability rank of actual activity $c_r$ generated by Markov model, and $Rank_t(c_r)$ is the probability rank generated by TF model. Then the record can be labeled as positive or negative depending on the sign of $Rank_t(c_r) - Rank_m(c_r)$. However, apart from contextual and historical features, the capacity of these two models may also be slightly affected by some random factors, such as the stochastic error. When the Markov and TF models perform similarly on the activity prediction, these random errors lead to wrong labeling. Therefore, we only take the records satisfying $|Rank_t(c_r) - Rank_m(c_r)| > \xi$ as training examples of the classifier. $\xi$ is called filtering parameter. At last, considering that the numbers of positive and negative examples may be unbalanced, we set the negative-rate of training examples as the weight of positive class and the positive-rate as the weight of negative class.

### 4.3 User Location Prediction

As we have discussed in Sect. 1, the user preference of specific location changes over time. Without sufficient training data, directly estimating the probability $P_u(v|t, c_v)$ leads to the under fitting problem. The generative approach is more effective to address this missing data situation than the discriminative approach. If the time variation pattern of the user location preference could be modeled as probability distribution $P_u(t_h|v)$, we can approximate the probability of location $v$ to be visited at time $t$ as follows, where $t_h$ is the HOD of $t$:

$$P_u(v|t, c_v) = \frac{P_u(t_h|v)P_u(v)}{\sum_{v' \in c_v} P_u(t_h|v')P_u(v')} \tag{4}$$

However, the time variation pattern varies from location to location. For example, some restaurants have three peak periods in a day including breakfast time, lunch time and dinner time, while some other restaurants only focus on dinner time. Due to this case, we perform non-parametric Kernel Density Estimation to reckon $P_u(t_h|v)$, which is widely used to estimate the shape of unknown probability density. The density of location $v$ at $t_h$ is formulated by

$$P_u(t_h|v) = \frac{1}{n_h d} \sum_{i=1}^{n} K\left(\frac{\Delta(t_h, h_i)}{d}\right) \tag{5}$$

where $\Delta(t_h, h_i) = min(|t_h - h_i|, 24 - |t_h - h_i|)$ is the interval between HOD $t_h$ and $h_i$, $K(\cdot)$ is the kernel function, $d$ is the bandwidth, and $n_h$ is the number of distinct record hours on this location.

## 5 Experiments

### 5.1 Datasets

We evaluate our model on public check-in datasets in two big cities (New York and Tokyo), collected by Yang et al. [15]. In these two datasets, check-in records last from Apr 2012 to Feb 2013, and locations are classified into 251 categories. The statistics description is shown in Table 1. We do not study other public datasets due to the lack of activity information, such as the Gowalla dataset [2].

**Table 1.** Datasets statistic

| | #User | #Location | #Check-in | #Location per user | #Category per user |
|---|---|---|---|---|---|
| NYC | 1,083 | 38,333 | 227,420 | 84.04 | 40.22 |
| TKY | 2,293 | 61,858 | 573,703 | 92.43 | 32.40 |

### 5.2 Experiment Setting

**Evaluation Plan.** In the following experiments, we set the proportion of training set $\Gamma_u^{(1)}$, $\Gamma_u^{(2)}$ and test dataset as 7:2:1. For more convincing results, we repeat each experiment 10 times and take the average of metrics into comparison.

**Parameter Setting.** We set the timespan threshold $\varepsilon$ as 6 h following the empirical rule [1,3], and the discount parameter as empirical formula $\delta = \frac{n_1}{n_1+2n_2}$ ($n_1$ and $n_2$ are the number of one-time transitions and two-times transitions)[6]. The latent space dimension $L$ of TF model is recommended as 64 on these datasets by Yang et al. [15]. We select the standard normal kernel function and rule-of-thumb bandwidth $d = (4\hat{\sigma}/3n)^{\frac{1}{5}} \approx 1.06\hat{\sigma}^{-\frac{1}{5}}$ for KDE [10]. We study the effect of filtering parameter $\xi$ in Sect. 5.3 and set it as 60.

### 5.3 Activity Prediction Evaluation

**Effect of Features and Parameters.** Firstly, we study the performance of binary classifier with different features. After attempting several methods such as logistic regression, decision tree and SVM, we apply the one with high performance and low training cost: Classification and Regression Tree (CART). The classification performance is measured by accuracy $Acc$ and weighted average F-score $F$, following [14].

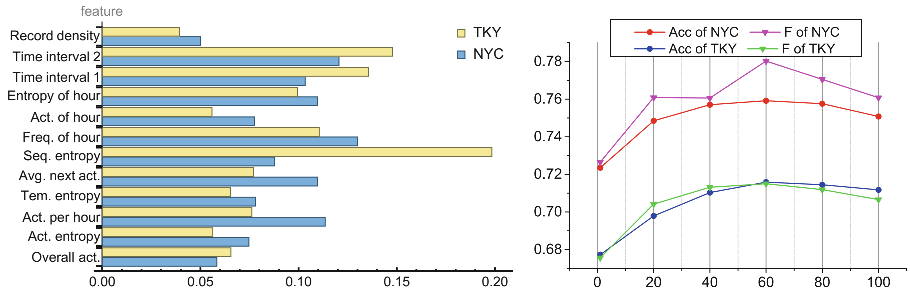
The classification performance evaluation based on different feature groups is shown in Table 2, where Seq, Tem and His are the abbreviation of sequential

**Table 2.** Features evaluation

| | Seq | Tem | His | His+Seq | Tem+His | Tem+Seq | All |
|---|---|---|---|---|---|---|---|
| *Acc* of NYC | 0.7154 | 0.7440 | 0.7511 | 0.7587 | 0.7534 | 0.7570 | **0.7593** |
| *F* of NYC | 0.7362 | 0.7677 | 0.7794 | 0.7701 | 0.7747 | 0.7666 | **0.7803** |
| *Acc* of TKY | 0.6822 | 0.7040 | 0.6998 | 0.7044 | 0.7054 | 0.7062 | **0.7158** |
| *F* of TKY | 0.6719 | 0.7022 | 0.6719 | 0.7074 | 0.7019 | 0.7029 | **0.7150** |

contextual features, temporal contextual features and historical features. We can observe that every paired feature groups combination outperforms the individual one, and combining all the features gets the best performance, implying that all three feature groups are effective and necessary.

Figure 2(a) describes the importance of features. Sequential contextual features and the sequential entropy take a larger proportion in TKY dataset. One possible reason is that the sequence regularities of users are stronger in TKY dataset, which makes sequential model more important in CAH approach.




(a) Feature Importance

(b) Effect of $\xi$

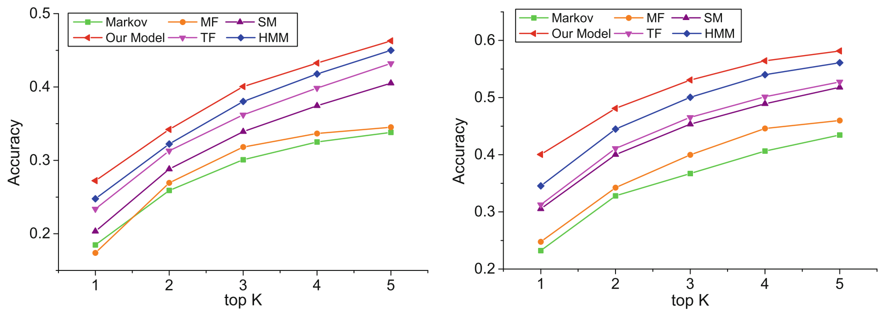
**Fig. 2.** Feature importance and parameter effect

Besides, Fig. 2(b) reports the effect of filtering parameter $\xi$. As $\xi$ increases, the labels of training examples become more credible. Thus the performance gets better when $\xi$ varies from 0 to 60. However, there is a negative correlation between $\xi$ and the number of classifier training examples. Owing to the insufficiency of training examples, the classification accuracy will fall back when $\xi$ is bigger than 60.

**Activity Prediction.** After training the classifier, we apply the most frequently used metric of mobility prediction performance, Acc@top$k$, to contrasting the performance of proposed CAH approach with following 5 baselines:

1. Most Frequent: This method assigns the most frequent activity of user $u$ at time $t$ as the result of prediction.
2. Fallback Markov: Order-2 Markov with fallback has been utilized widely in mobility prediction on GPS trajectories and WiFi network [12].

3. Smooth Markov and Tensor Factorization: The sequential and temporal models we used, which have been introduced in Sect. 4.2.
4. HMM of CEPR: This model integrates temporal regularity and Markov models into a hidden Markov framework [6].




(a) NYC dataset (b) TKY dataset

**Fig. 3.** Acc@top$k$ of activity prediction

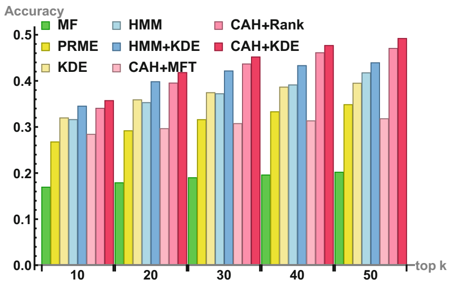
Figure 3 shows the top-$k$ ($k = 1, 2, 3, 4, 5$) accuracy of activity prediction. It can be observed that (1) the proposed Context-Aware Hybrid (CAH) approach achieves the highest accuracy for all $k$ values, and outperforms Smooth Markov and Tensor Factorization models by a large margin. In particular, when we choose the activity with the maximum probability as the prediction result, the CAH approach shows at least 10.1% and 15.7% improvement over any other method on NYC dataset and TKY dataset; (2) As the prediction list size $k$ increases, the performance gaps between CAH and some baselines become smaller, such as HMM and TF. This result is not surprising since a user usually prefer about 30–40 activities according to Table 1. In addition, it is clear that a large prediction list size $k$ is meaningless for practical applications, thus getting higher accuracy with a small $k$ is much more valuable.

### 5.4 Location Prediction Evaluation

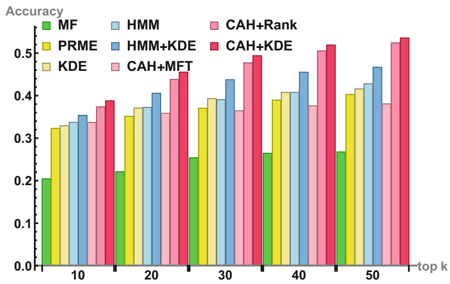
For location prediction evaluation, we use the same metric as activity prediction (i.e. Acc@top$k$) and study following methods for comparison:

1. Most Frequent: Returning the most frequent locations of user as result.
2. KDE: Predicting locations only with generative method of the second stage, without the first stage.
3. PRME: This method [3] constructs a metric embedding model to balance sequential information and individual preference.
4. HMM of CEPR: We provide two versions of this approach. HMM represents the original approach of [6], predicting locations without activity information. HMM&KDE uses the hidden Markov framework of [6] to predict activities and the proposed generative approach to predict locations.

5. CAH&Rank/CAH&MFT: We apply other two methods to predict user future location based on CAH's results. CAH&Rank ranks locations by overall frequency [16]. CAH&MFT(Most Frequent of Time) directly estimates $P_u(v|t, c_v)$ based on frequency of location $v$ at time $t$ by user $u$.
6. CAH&KDE: The integrity version of the proposed model in this article.




(a) NYC dataset




(b) TKY dataset

**Fig. 4.** Acc@top$k$ of Location Prediction

Figure 4 depicts the Acc@top$k$ ($k = 10, 20, 30, 40, 50$) of above methods. We can learn from that: (1) the integrity version of the proposed model (CAH&KDE) gets the best results for all the $k$ values. Specifically, it shows 12.9% and 14.4% improvement over HMM of CEPR and 28.7% and 20.1% improvement over PRME, when $k = 10$; (2) the proposed generative approach (CAH&KDE) outperforms any other location prediction method based on CAH's results(i.e. CAH&Rank/CAH&MFT). Note that CAH&MFT gets the worst result, which is in line with the discussion in Sect. 4.2; (3) the performances of CAH&KDE and HMM&KDE are obviously better than KDE and HMM, implying that exploiting activity information facilitates location prediction. In addition, it also proves, to some extent, our two-stage framework is suitable for other activity prediction approaches; (4) the comparison of CAH&KDE and HMM&KDE indicates that improving activity prediction accuracy is beneficial to location prediction.

## 6 Conclusion

In this article, we propose a two-stage method to predict locations. In the first stage, we study the contextual and historical features that impact the prediction accuracy of sequential and temporal models, then we adopt a binary classifier to switch between these two models depending on predicting context. In the second stage, Kernel Density Estimation is performed to capture the time variation of the user location preference. Based on the evaluation results, our model significantly outperforms existing approaches.

Several interesting future directions exist for further exploration. For example, the sequential dependency and temporal regularity of user activities may affect each other, which makes it possible to improve the predictability.

**Aknowledgements.** The research is funded by National Key Research and Development Program, investigation on global optimization of resource scheduling method based on the specific application under the contract number No. 2016YFB0201402.

## References

1. Cheng, C., Yang, H., Lyu, M.R., King, I.: Where you like to go next: successive point-of-interest recommendation. IJCAI **13**, 2605–2611 (2013)
2. Cho, E., Myers, S.A., Leskovec, J.: Friendship and mobility: user movement in location-based social networks. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1082–1090. ACM (2011)
3. Feng, S., Li, X., Zeng, Y., Cong, G., Chee, Y.M., Yuan, Q.: Personalized ranking metric embedding for next new POI recommendation. In: Proceedings of IJCAI (2015)
4. Kim, S., Park, H.: Fast nonnegative tensor factorization with an active-set-like method. High-Performance Scientific Computing. Springer, Heidalberg (2012). doi:10.1007/978-1-4471-2437-5_16
5. Li, X., Lian, D., Xie, X., Sun, G.: Lifting the predictability of human mobility on activity trajectories. In: 2015 IEEE International Conference on Data Mining Workshop (ICDMW), pp. 1063–1069. IEEE (2015)
6. Lian, D., Xie, X., Zheng, V.W., Yuan, N.J., Zhang, F., Chen, E.: CEPR: a collaborative exploration and periodically returning model for location prediction. ACM Trans. Intell. Syst. Technol. (TIST) **6**(1), 8 (2015)
7. Liu, Q., Wu, S., Wang, L., Tan, T.: Predicting the next location: a recurrent model with spatial and temporal contexts. In: Thirtieth AAAI Conference on Artificial Intelligence (2016)
8. Liu, X., Liu, Y., Aberer, K., Miao, C.: Personalized point-of-interest recommendation by mining users' preference transition. In: Proceedings of the 22nd ACM International Conference on Information & Knowledge Management, pp. 733–738. ACM (2013)
9. Noulas, A., Scellato, S., Lathia, N., Mascolo, C.: Mining user mobility features for next place prediction in location-based services. In: 2012 IEEE 12th International Conference on Data Mining, pp. 1038–1043. IEEE (2012)
10. Silverman, B.W.: Density Estimation for Statistics and Data Analysis, vol. 26. CRC Press, Boca Raton (1986)
11. Song, C., Qu, Z., Blumm, N., Barabási, A.L.: Limits of predictability in human mobility. Science **327**(5968), 1018–1021 (2010)
12. Song, L., Kotz, D., Jain, R., He, X.: Evaluating location predictors with extensive wi-fi mobility data. In: INFOCOM 2004, Twenty-Third Annual Joint Conference of the IEEE Computer and Communications Societies, vol. 2, pp. 1414–1424. IEEE (2004)
13. Wang, Y., Yuan, N.J., Lian, D., Xu, L., Xie, X., Chen, E., Rui, Y.: Regularity and conformity: location prediction using heterogeneous mobility data. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1275–1284. ACM (2015)

14. Witten, I.H., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, Burlington (2005)
15. Yang, D., Zhang, D., Zheng, V.W., Yu, Z.: Modeling user activity preference by leveraging user spatial temporal characteristics in LBSNs. IEEE Trans. Syst. Man Cybern.: Syst. **45**(1), 129–142 (2015)
16. Ye, J., Zhu, Z., Cheng, H.: Whats your next move: user activity prediction in location-based social networks. In: Proceedings of the SIAM International Conference on Data Mining. SIAM (2013)
17. Yuan, Q., Cong, G., Ma, Z., Sun, A., Thalmann, N.M.: Who, where, when and what: discover spatio-temporal topics for twitter users. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 605–613. ACM (2013)
18. Zhang, J.D., Chow, C.Y.: Spatiotemporal sequential influence modeling for location recommendations: a gravity-based approach. ACM Trans. Intell. Syst. Technol. (TIST) **7**(1), 11 (2015)
19. Zhong, Y., Yuan, N.J., Zhong, W., Zhang, F., Xie, X.: You are where you go: inferring demographic attributes from location check-ins. In: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, pp. 295–304. ACM (2015)