This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2017.2705083, IEEE Transactions on Knowledge and Data Engineering

GALLOP: GlobAL feature fused LOcation Prediction for Different Check-in Scenarios

Yuxing Han, Junjie Yao, Liping Wang, Xuemin Lin

Abstract—Location prediction is widely used to forecast users' next place to visit based on his/her mobility logs. It is an essential problem in location data processing, invaluable for surveillance, business, and personal applications. It is very challenging due to the sparsity issues of check-in data. An often ignored problem in recent studies is the variety across different check-in scenarios, which is becoming more urgent due to the increasing availability of more location check-in applications.
In this paper, we propose a new feature fusion based prediction approach, GALLOP, *i.e.*, GlobAL feature fused LOcation Prediction for different check-in scenarios. Based on the carefully designed feature extraction methods, we utilize a novel combined prediction framework. Specifically, we set out to utilize the density estimation model to profile geographical features, *i.e.*, context information, the factorization method to extract collaborative information, and a graph structure to extract location transition patterns of users' temporal check-in sequence, *i.e.*, content information. An empirical study on three different check-in datasets demonstrates impressive robustness and improvement of the proposed approach.

Index Terms—Location Prediction, Geographical Closeness, Trajectory Data, Check-in Behavior Analysis,

1 Introduction

We have witnessed a stunning availability of smart devices in recent years. With these devices' easy carrying and always online capabilities, nowadays lots of users are getting used to 'check-in' their mobility activities on many popular services and applications, *i.e.*, report their location and reviews from their smart phones or other digital accessories. Representative check-in applications include Foursquare, Yelp, and general social platforms like Facebook, Twitter and Weibo.

A check-in heat map of one thousand users in Beijing is shown in Figure 1. We choose a social media check-in dataset and visualize these users' check-in locations. It is apparent that there exhibits an impressive check-in pattern across prominent places like CBD, airport and tech parks. Users' check-in activities present a unique angle into their life, and the distribution patterns reflect their interest and preference. In recent years, the value of check-in data have been demonstrated in many applications, including but not limited in mobile advertising, promotion recommendation, traffic management and social surveillance [1], [2], [3], [4].

Behind the check-in data processing, *location prediction* [5], [6], [7] is a fundamental task. However, it is very challenging due to the check-in data's inherent characteristics. First, *Sparsity*: There is a large possible space which users can visit, but in fact they only cover a small set of the places [8], [9]. Second, *Heterogeneousness*: Location data consists of different kinds of features, i.e., the location, text and temporal information.

- *Yuxing Han, Junjie Yao, Liping Wang are affiliated with East China Normal University, Shanghai, 200062, P.R.China. E-mail: yxhan@student.ecnu.edu.cn, {junjie.yao, lipingwang}@sei.ecnu.edu.cn.*
- *Xuemin Lin is affiliated with East China Normal University, Shanghai, 200062, P.R.China, and CSE, University of New South Whales, Sydney, N.S.W., Australia. E-mail: lxue@cse.unsw.edu.au.*

Fig. 1: Users' Check-in Heat map in Beijing

The recent emerging trend of more and more check-in applications makes the prediction problem complex. It leads to the third and usually ignored but becoming more difficult challenge, which we will focus in this article, *i.e, Variety*. The check-in activities in different applications are not following the same patterns. A growing dataset of different kinds of users' trajectories and mobility records become available [10], [11], [12]. For example, local area check-ins in specific App are intense in users' home and office routines. In contrast, travel logs are able to cover a much larger scope and usually very sparse.

Figure 2 shows the difference of time/distance gap between consecutive check-ins within each check-in log. We choose three kinds of check-in scenarios, *i.e.*, Foursquare as local check-in, micro-blog as log sharing and Yelp as place reviewing. The time/distance gap of consecutive check-ins in a check-in log is aggregated and compared. The check-ins in micro-blog and Foursquare follow similar patterns. They exhibit varying but smaller distance span, and the time gap is shorter, compared with Yelp reviews. Under the similar

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2017.2705083, IEEE Transactions on Knowledge and Data Engineering

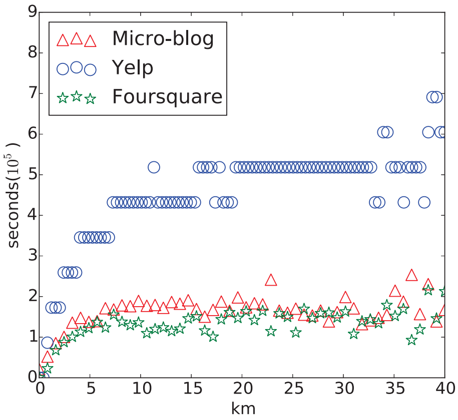


Fig. 2: Time v.s. Distance Gap of Different Consecutive Check-ins

check-in distance gap, the micro-blog check-in has slightly but impressively longer time gap. Different time gaps reveal varying mobility patterns of interest or preference.

Though some recent work [8], [9], [12], [13] pay attention to these check-in behavior differences, the inherent variety between these check-ins and their affect to location prediction is usually missing. State of the art location prediction methods can be categorized into three lines, *i.e.*, collaborative feature based [14], spatial feature based [15], and fusion feature based [16], [17]. Collaborative feature based approaches utilize the Collaborative Filtering or factorization methods to cope with the sparsity challenge but they are sensitive to spatial features. In contrast, spatial feature based methods use the gravity and locality closeness measurements to fit into the location setting, but are usually difficult to generalize the temporal and other related features. The feature fusion approach shows the advantage of feature combination to deliver improved accuracy. In the later parts of this article, we will show that existing prediction methods can not be directly applied to all check-in scenarios, where their performance vary greatly. Empirical details can be referred to Section 5.

In this paper, we propose a new feature fusion approach, *i.e.*, GlobAL feature fusion for LOcation Prediction (GALLOP), to cope with the variety problem in location prediction. To improve the applicability of location prediction approach, We utilize several kinds of features and discuss their different characteristics in the variety of check-in scenarios. Three classes of features are used in GALLOP: context feature(geographical aspects), collaboration feature(users' latent interest space) and content feature(places' description attributes). We introduce intuitive ways to model these check-in features and then formalize a combination framework to deliver the predicted target places to end users.

The proposed GALLOP prediction approach is not only general over different check-in scenarios but also comprehensive of different features. In the context feature, we design a multiple granularity model to profile the geographical closeness. We select the predicted candidates based on the combination of local district, local city and state scales. The weights of each scale are learned from training data. This approach is similar to recent introduced density model [15], but we improve it with new indexing and granularity organization methods, guaranteeing its flexibility. In the collaborative feature, we resort to the de facto factorization techniques to model the other users' check-in logs and supplement it with a weighted version. In the content feature, we transform users' consecutive check-in sequence, *i.e.*, transition patterns of the check-in sequence into a graph representation, and extract these attributes' closeness with a Random Walk with restart model.

To conclude the contribution of this paper, we summarize the following ones:

- First, we investigate the difference of several representative check-in scenarios from the spatial and temporal aspects. We argue that these varying check-in scenarios ask for more general location prediction methods.
- Second, we demonstrate the combination power of location features in a novel angle. We not only utilize the different classes of context, collaboration and content information, but also factorize them in a new way to improve the prediction robustness and generality.
- At last, the extensive study over several real datasets reveals the improvement and advantage of our approach. We provide an empirical study with several competition methods. Detailed experiments show the different behaviors of the prediction methods, and prove that the general location prediction approach is a better choice to tackle the location prediction challenges.

The remaining of this paper is organized as follows. We first discuss the check-in characteristics and the approach preliminaries in Section 2. We then illustrate the check-in features in Section 3 and provide the prediction method details in Section 4. The empirical study is presented in Section 5. We proceed to review the related work in Section 6. At last we conclude this paper and outline the future directions.

2 PRELIMINARIES

This section first presents the data characteristics of different check-in datasets and then lists the notions used in this paper.

2.1 Check-in Activities Analysis

Three large real check-in datasets are chosen in this paper, as we introduced in Figure 2.

- **Foursquare**[1]: the online check-in records collected within the areas of New York. Foursquare is a unique check-in service and used to report users' position with comments [10].
- **Yelp**[2]: the online business review records, supporting user-contributed reviews about local businesses.

1. http://foursquare.com
2. http://yelp.com

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2017.2705083, IEEE Transactions on Knowledge and Data Engineering

TABLE 1: Check-in Data Statistics

	#Users	#POIs	#Categories	#Check-ins	User-POI density	Time Span
Foursquare	32,421	41,663	354	399,034	1.14×10^{-4}	May. 2008 to Jul. 2011
Yelp	80,789	33,986	754	495107	4.62×10^{-4}	Apr. 2009 to Jan. 2015
Microblog	11,191	48,318	253	158,628	2.21×10^{-4}	Dec. 2011 to Nov. 2013

- **Micro-blog**[3]: Here we extract check-in logs from *Sina Weibo*, which is one of the most popular micro-blog platforms. Of its hundreds of millions of users' life records, lots are check-in logs.

These three check-in datasets span from the check-in service, reviewing website to micro-blog platforms, covering a diverse set of different check-in behaviors. We carefully select them based on the respective scenarios of check-in services. Table 1 shows the statistics of these datasets. Each dataset has tens of thousands of users and POIs. The user-poi density in the user-poi matrix is very low, in a range of 10^{-4}.

In addition to the time gap v.s. distance gap distribution shown in Figure 2, we also provide other check-in difference studies. We sort each user's check-in record based on the chronological date order, and generate the trajectories. We analyze the spatial-temporal patterns of users' check-in activities and show the following CCDF(Complementary Cumulative Distribution Function[4]) plots of time and distance gaps in Figure 3,4.

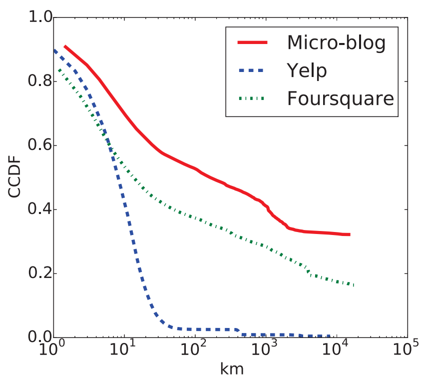


Fig. 3: Distance Gap Distribution in Consecutive Check-ins

It is shown that, micro-blog data has the slowest decreasing shape, both in distance and time dimensions, meaning that micro-blog users have a much larger activity range in spatial and temporal scopes. In contract, the Yelp check-in activities enjoy a sharp decline in spatial interval but a slow decline in temporal interval, which reveals that Yelp users are focusing on specific restaurant areas. Besides, Foursquare has a moderate pace in both distance and time. It shows general check-in activities like Foursquare are more local than micro-blog records but more active than those specific category favors, *i.e.*, Yelp line.

The above data analysis reveals that, different kinds of check-in activities bring up varying characteristics. User intention and preference behind these are different. Thus

3. http://weibo.com
4. https://en.wikipedia.org/wiki/Cumulative_distribution_function

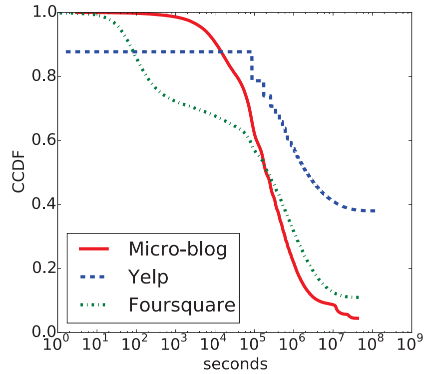


Fig. 4: Time Gap Distribution in Consecutive Check-ins

the prediction methods should be tailored. However, most of current location prediction methods are not general for these different kinds of check-in scenarios [11].

This paper proposes a novel and general solution to cope with varying scenarios. The basic motivation of the proposed approach is that, by carefully design the feature extraction and combination approach, we can include varying requirements of different scenarios and provide a general solution for the location prediction problems, finally improve the prediction performance under different scenarios.

2.2 Definitions

The notations used in this paper are listed in Table 2.

TABLE 2: List of Notations

Notations	Explanation
$\mathcal{U}$	Users
$\mathcal{L}$	Check-in Places
$\mathcal{C}$	Attributes/categories of check-in places
Tr_u	Check-in trajectories of user u
C_l	Categories of location l
$\mathcal{R}$	Users' check-in matrix, $\lvert\mathcal{U}\rvert \times \lvert\mathcal{L}\rvert$

In the setting of location prediction, we focus on the users, POI places and the users' check-in log, places' description information to fulfill the location prediction tasks.

We continue to provide the definitions used in this paper.

Definition 1 (Check-in).
A user u's check-in record is represented as a multiple fields tuple $r = (u, t, l)$, where t is the time stamp, l is the location coordinate comprised of (*latitude*, *longitude*) elements.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2017.2705083, IEEE Transactions on Knowledge and Data Engineering

Here we assume that $\phi = \{c_1, c_2, \ldots, c_i, \ldots, c_{|r|}\}$ are the description attributes of check-in places. ϕ_l is the set of corresponding attributes of the location l.

Definition 2 (Check-in Trajectory).
One of a user's check-in trajectory is his/her ordered check-in sequence, represented as $\{r_1, r_2, \ldots r_n\}$.

The check-in records profile users' mobility activities and have several different classes of features. Though this has been investigated in recent studies, here we treat and divide them in a different angle, *i.e.*, global feature combination, for better clarification and combination. We list the features used in this work.

- **Context Feature**: refers to the spatial dimension. Users' check-in activities are distributed in a spatial scope. Nearby places can contribute to the representation of users' check-in records, especially when the users visit a focused set of places [9].
- **Collaboration Feature**: refers to the collaborative dimension. It is contributed from a large group of users' records. Similar to the collaborative filtering formalization in recommendation systems, here the users' check-in activities can be reduced into a latent dimension, representing their collaboration preferences.
- **Content Feature**: refers to the place dimension. Places are not merely visited by users. These places have inherent attributes, *i.e.*, categories, text descriptions and other kinds of annotations. We discuss how the transition between places reveals users' interest/preference over time. In further, the transition patterns benefit the closeness extraction of places for better prediction. In this paper, we use the categories of the location as its attribute description, and the details would be discussed in Section 3.3.

On top of the collection of a set of users' check-in trajectories, the objective of location prediction is to predict users' future check-in activities, *i.e.*, which places they would like to visit. For Foursquare/Micro-blog users, predict future places they will check-in, and for Yelp users, predict future restaurants they will visit.

Problem Statement[Location Prediction]: The location prediction problem is to predict the preference probability $p(l|u)$, *i.e.*, a user u to an unvisited place l.

The prediction problem studied in this paper is more challenging than successive POI recommendation. While general POI recommendation focuses on estimating users' preference over POI, successive POI recommendation pays more attention on most recent checked-in locations to deliver satisfied recommendations. The location prediction problem is related to sequential patterns of check-in behaviors and aims to identify the next place to be visited, given the target users [11].

3 Check-in Features Extraction

This section investigates the chosen features used in the location prediction. For each feature, we first discuss their extraction method and the then corresponding characteristics used for prediction.

3.1 Context Feature

Density closeness of users' check-in logs from the spatial perspective has received a lot of attention [16]. It has shown advantage over traditional spatial modeling, both in flexibility and accuracy. We follow similar motivation and provide our own design to extract users' preference from the spatial aspects.

Concretely, we choose to use kernel density estimation (KDE)[5] to model the individual distribution of position of location visited by the user, since this non-parametric can be used with arbitrary distribution and without prior knowledge of the form for the distance distribution. The density model is very flexible and useful in the spatial feature extraction.

We focus on the individual historical check-in sequence in the following way. Denote $\mathcal{L} = \{l_1, l_2, ..., l_n\}$ as the historical check-in record of a user u where $l_j = (x_j, y_j)$ records the longitude/latitude of spatial location.

To take advantage of two-dimensional spatial records, we employ a two-dimensional normal KDE model. The basic procedure of KDE model in location prediction consists of two steps:

Step 1. For each user u, we estimate his/her kernel location density and derive his/her probability over places based on the trajectories.

$$\hat{f}(u|\mathcal{L}, h) = \frac{1}{n}\sum_{j=1}^{n} K_h(l - l_j), \tag{1}$$

together with

$$K_h(l - l_j) = \frac{1}{2\pi h} exp\Big(-\frac{(x - x_j)^2 + (y - y_j)^2}{2h}\Big), \tag{2}$$

where bandwidth h is estimated as a fixed optimal parameter, based on the density estimation algorithm [18].
Step 2. For a new location l' that has not been visited by user, we derive its to be visited probability as:

$$P(l|\mathcal{L}) = \frac{1}{n}\sum_{j=1}^{n} K_h(l' - l_j). \tag{3}$$

It is the combination of this place's density based closeness to the users' visited places.

A case study of the context feature selection is shown in Figure 5. One user's check-in sequence at Phoenix, Arizona is visualized. With the color degree as the check-in density, the figure shows this user's activity zone and the corresponding hubs, *i.e.*, home or office. It is vivid that the context features can profile this user's check-in pattern, thus spatial preference.

Though the above method seems feasible, due to the inherent varying check-in behaviors among users, we argue that a fixed model setting is not applicable in the practical situation. Besides our discussion of different check-in applications in Section 2, even in the same check-in application, human-location data where dense urban areas will tend to have high event density and sparsely-populated rural areas will have low event density. And, users who are enthusiastic about social network may tend to have more

5. https://en.wikipedia.org/wiki/Kernel_density_estimation

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2017.2705083, IEEE Transactions on Knowledge and Data Engineering

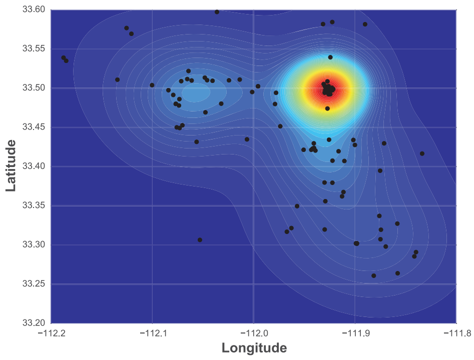


Fig. 5: One User's Check-in Heatmap at Phoenix, Arizona

diverse check-in activities than others. Be aware that, the work in this paper follows in the similar way of [15] and we go beyond further with new adaptive strategies and indexing structures.

Here we continue to introduce an adaptive bandwidth method on top of the basic model. Instead of fixed density bandwidth, we relax the kernel bandwidth h_i for each location l_i in check-in dataset. That is to say, h_i is chosen as the distance to the $\boldsymbol{k}$-th nearest neighbor to l_i in the training data in $K_h(*)$ in Eq 3.

In the actual implementation of density based context feature extraction, we utilize the quad-tree [19] instead of tradition R-tree to speed up the $\boldsymbol{k}$NN processing. A quad-tree is usually used to partition a two-dimensional space by recursively subdividing it into four quadrants or regions. Different from R-tree, the nodes of a quad-tree will be adaptively split only when enough points are located within the corresponding region. As the check-in distribution might be highly skewed, quad-tree can achieve a better performance.

3.2 Collaborative Feature

We first list the definition used tin present the collaborative features with user and check-in places.

Definition 3 (Check-in matrix).

A check-in data matrix $\mathcal{R}_{|\mathcal{U}|\times|\mathcal{L}|}$ is the historical check-in data of a user set, in which each entry represents the check-in frequency (e.g. Micro-blog and Foursquare) or rating(e.g. Yelp) of user $u \in \mathcal{U}$ on a location $l \in \mathcal{L}$.

The user check-in matrix $\mathcal{R}_{|\mathcal{U}|\times|\mathcal{L}|}$, as shown in Table 1, is fairly sparse. Compared with 10^{-2} density in content rating or web recommendation problems [20], here the 10^{-4} is very challenging. One necessary step to extract users' preference interest is to resort to the latent representation, for the efficiency and robustness issues.

Similar to the conventional recommendation problems, a feasible method choice is low rank matrix factorization to get users' latent location interests and then predict future candidates [11], [20].

Here we formalize the factorization process as a mapping from the raw space with users and places into a joint latent space with dimension $\mathcal{K} \ll \min(|\mathcal{U}|, |\mathcal{L}|)$. Users' check-in records can thus be modeled as the product between them in the latent space. That is to say,

$$\mathcal{R}_{|\mathcal{U}|\times|\mathcal{L}|} = \boldsymbol{U}_{|\mathcal{U}|\times\mathcal{K}} \times \boldsymbol{L}^T_{|\mathcal{L}|\times\mathcal{K}}, \tag{4}$$

$U_{|\mathcal{U}|\times\mathcal{K}}$ measures users' location interests over latent space and $L_{|\mathcal{L}|\times\mathcal{K}}$ represents the locations' latent space.

The factorization can be usually solved via several candidate methods, such as singular value decomposition. However, due to the large scale and sparsity of $\mathcal{R}$ in this work, it is very prohibitive to directly solve this. With the luck that the weights in the raw matrix are non-negative, we resort to the Non-Negative Matrix Factorization methods to tackle this problem. It is achieved by approximating the raw matrix by solving the following optimization problem:

$$\min_{\boldsymbol{U},\boldsymbol{L}} \sum_{i,j} (R_{ij} - \boldsymbol{U}\boldsymbol{L}^T_{ij})^2 + \lambda(||\boldsymbol{U}||^2 + ||\boldsymbol{L}||^2) \tag{5}$$

here we use the squared norm of the loss function. The factor λ controls the regularization power. We choose the projected gradient descent algorithm [21] to derive the factorization parameters and then the latent matrices for location prediction.

The probability of a user u to visit the location l, is thus determined by:

$$p(l|u)_{coll} \propto \boldsymbol{U}_u \boldsymbol{L}^T_l \tag{6}$$

The next section will illustrate the details of collaborative feature selection for the combined location prediction approach.

3.3 Content Feature

We continue to introduce the content features of check-in records. Lots of current location prediction methods prefer spatial collaborative features [14], [15]. And other methods choose temporal or social network information as the supplement sources [11], [16]. However, the rich representation of check-in places is still missing. Here we discuss a general way to obtain the content description features.

POI, *i.e.*, places in check-in records are usually annotated with categories or attributes. Users' transition pattern from one place to another one shows the interest flow between these places. The extracted closeness from these transition patterns can be used to predict the potential locations that user will take a visit next.

Here we design a graph structure to organize the place closeness from the category dimension. and then conduct the category closeness mining algorithm. We construct a directed graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, where each category is represented as a node in $\mathcal{V}$. For each pair of consecutive records from the check-in dataset, there exist several edges in $\mathcal{E}$ with each connecting corresponding categories in the check-in sequences.

Specifically, let $r_i = (u, t_i, l_i)$ and $r_j = (u, t_j, l_j)$ representing two consecutive check-ins of user u, where $\phi_i = \{c_{i1}, c_{i2}, \ldots, c_{i|r_i|}\}$ and $\phi_j = \{c_{j1}, c_{j2}, \ldots, c_{j|r_j|}\}$ represent l_i and l_j's corresponding categories. Then we connect links between categories of l_i and l_j. That is to say, we set up connections between a category pair with one picked from ϕ_i and the other picked from ϕ_j.

After the processing, this step brings up a graph representation of the categorical transition of all the users'

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2017.2705083, IEEE Transactions on Knowledge and Data Engineering

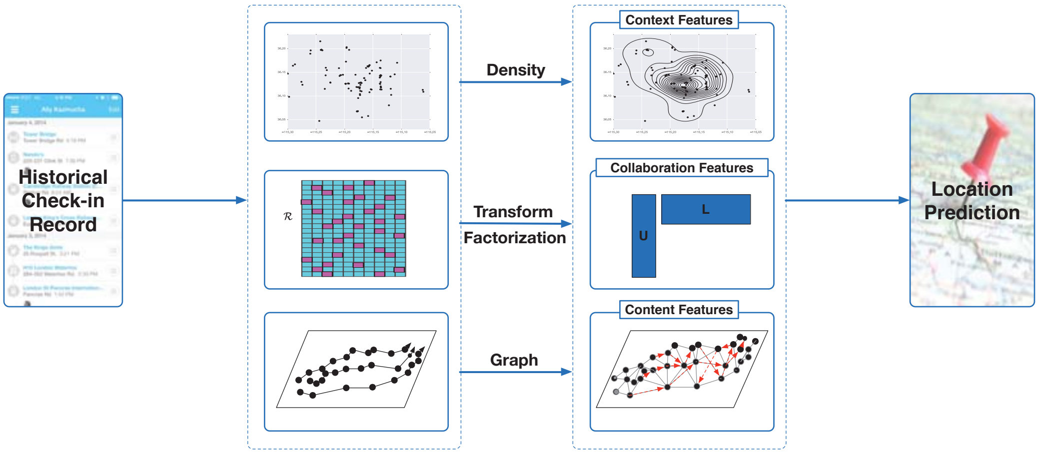


Fig. 6: Global Feature Fusion for Location Prediction

check-in activities. The weight of links can be measured by the corresponding frequency. Be aware that this step may introduce the self-loop links, of which the weights are measured by the self transition.

TABLE 3: Categorical Transitions in New York

Laundry Service	$\Longrightarrow$	Gym/Fitness Center
Sculpture Garden	$\Longrightarrow$	BBQ Joint
Hobby Shop	$\Longrightarrow$	Gaming Cafe
School	$\Longrightarrow$	Gift Shop
Academic Building	$\Longrightarrow$	Hobby Shop
Game Cafe	$\Longrightarrow$	Gift Shop
Stadium	$\Longrightarrow$	Nightlife Spot
Vietnamese Restaurant	$\Longrightarrow$	Sculpture Garden
	......	

Table 3 shows the category transition examples extracted from the Foursquare dataset. We construct the graph structure of place categories from users' check-in sequences in New York city and extract the highly frequent category transition pairs. We can find that, users exhibit reasonable transition patterns. There exist remarkable activities, i.e., from the stadium/gym to nightlife spot, the clusters of hobby/game shop and cafe. With the help of these content feature patterns, we can improve the location prediction performance.

After the above check-in feature investigation and the extraction method discussion, we continue to present the feature combination methods and location prediction implementation in the next section.

4 Location Prediction Approach

4.1 Approach Framework

Figure 6 shows the overall design of the proposed location prediction approach. As shown in the previous sections, We use three kinds of check-in features and combine them in an intuitive way, in which we merge them for a final location prediction score. We employ the multiplication way to combine the features, guaranteeing the extensibility and easy tuning.

$$p(l|u) = p_{con}(l|u)p_{coll}(l|u)p_{cat}(l|u) \tag{7}$$

We continue to discuss the prediction features in this section. Specifically, we utilize the locality patterns of context features and propose the improved density closeness estimation model. We discuss the weighted factorization method to extract the collaboration feature to alleviate the sparsity of check-in records. And, we continue to enrich the content feature in the graph structure and extract the transition patterns with Random Walk with restart model.

4.2 Context Feature

We continue to present a multi-scale smoothing method, improving the context feature's generality performance over different datasets [15]. An concerning issue of individual context feature extraction method discussed in Section 3.1 is the sparsity problem, *i.e.*, for many users, there is relatively small data, making it difficult to formalize accurate models of location influence at individual level. To alleviate these, we propose to include some smoothing factors into the feature extraction method. That is to say, combine several kernel density models of different population levels.

In this work, we design a new model for multi-granularities and improve it with new indexing structure since users who have more activities in the same area are more like to share more similar check-in behavior than others.

In the mixture kernel density model for an individual user u, we include him/her in several different coarse spatial levels, such as city and state. We utilize all the historical records within high-level areas to generate corresponding kernel estimators, which can be augmented to the kernel estimate only from individual data. The formal estimation for mixture-KDE with k-level is:

$$p_{con}(l|u) = P_{mix}(l|\mathcal{L}^u) \quad = \sum_{j=1}^{k} \gamma_j P(l|\mathcal{L}_j^u), \tag{8}$$

where $\gamma_1, \gamma_2, ..., \gamma_k$ are non-negative mixture weights with $\sum_{j=1}^{k} \gamma_j = 1$. $P(l|\mathcal{L}_1^u), P(l|\mathcal{L}_2^u), ..., P(l|\mathcal{L}_k^u)$ represent estimator contributions from different levels respectively. In the model training stage, we usually adopt a three-level strategy, *i.e.*, individual, city and state level. It is also possible that more levels can be constructed, which is a trade-off between

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2017.2705083, IEEE Transactions on Knowledge and Data Engineering

accuracy and training cost. Inevitably, more levels lead to more training costs, and better accuracy, but bring up over-fitting concern.

Under the above assumption, each sample (location) of historical check-in dataset is independent and identically distributed, and drawn from the mixture distribution based on Equation 8. In the inference stage, to fit the parameters of a mixture of m densities, we pick a random i from $\{1, 2, ..., m\}$ where probability of picking i is γ_i, then pick a sample according to the density $P(l|\mathcal{L}^u_j)$.

We choose the commonly used Expectation-Maximization (EM) algorithm to learn the model parameters, which is described in Algorithm 1. The algorithm runs as follows. Based on Eq 1-Eq 3, we first create several KDE models of check-in records on different granularity levels (Line 1) and perform the initialization steps for the mixture model (Line 2). In each iteration, to update the weight γ_i for each KDE model $\mathcal{M}^i$ in the M-step (Line 4), the E-step calculates the statistical intermediate variables $\theta^{(m)}_{ij}$ and $\varphi^{(m)}_j$(Line 3). Then we determine whether the convergence status is achieved and decide whether to continue EM steps or finish the algorithm (Line 5-6).

Algorithm 1: Context Feature Extraction

1 Initialize KDE models $\mathcal{M}^1, \mathcal{M}^2, ..., \mathcal{M}^s$ of s different levels in training set $\mathcal{L}^u_1, \mathcal{L}^u_2, ..., \mathcal{L}^u_s$. Annotate $\mathcal{M}^i_j$ as the visiting probability for location j according to model $\mathcal{M}^i$;

2 Set the initial estimator $\gamma^{(0)}_j = \frac{1}{s}, j = 1, ..., s$ for the mixture model, and get the initial log-likelihood:

$$\mathbb{W}^{(0)} = \frac{1}{|\mathcal{L}^u|} \sum_{i=1}^{|\mathcal{L}^u|} log \left(\sum_{j=1}^{s} \gamma^{(0)}_j \mathcal{M}^i_j \right)$$

3 **E-Step**: Get

$$\theta^{(m)}_{ij} = \frac{\gamma^{(m)}_j \mathcal{M}^i_j}{\sum_{l=1}^{k} \gamma^{(m)}_l \mathcal{M}^l_j}, i = 1, 2, ..., |\mathcal{L}^u|, j = 1, 2, ..., s$$

$$\varphi^{(m)}_j = \sum_{i=1}^{|\mathcal{L}^u|} \theta^{(m)}_{ij}, j = 1, 2, ..., s$$

4 **M-step**: Update the new estimates:

$$\gamma^{m+1}_j = \frac{\varphi^{(m)}_j}{|\mathcal{L}^u|}, j = 1, 2, ..., s$$

5 Derive the new log-likelihood score function:

$$\mathbb{W}^{(m+1)} = \frac{1}{|\mathcal{L}^u|} \sum_{i=1}^{|\mathcal{L}|} log \left(\sum_{j=1}^{s} \gamma^{(m+1)}_j \mathcal{M}^i_j \right)$$

6 Return to **E-Step** if $|\mathbb{W}^{(m+1)} - \mathbb{W}^{(m)}| > \delta$; otherwise finish the training process ;

After the inference algorithm, we get the context features' weighting scores and thus the feature representation for users' check-in records. It should be noted that this density based context feature extraction method does not take the check-in sequence information into consideration. We would continue discuss the collaborative and content features used for location prediction. In the next experiment section, we will report the context feature extraction improvement and its contribution to the final fusion approach.

4.3 Collaborative Feature

While the commonly used collaborative feature extraction method usually delivers a reasonable performance in many scenarios, it does not perform well in the check-in usage, which is revealed in our initial study. The causes for the unsatisfying performance are two-folds. First, the check-in matrix $\mathcal{R}$ is very sparse, *i.e.*, it has many zero items. Zero check-in record means that the user never visit the corresponding place. This is probably because he/she is not a fan of that place or his/her location scope is rather focused. Second, for some rated places, the frequency information is not enough. In the counterpart representation of movie or product recommendation, users not only reveal their favorite items with high ratings, but also their least favorite items with low rating. However, this situation does not hold in the case of location prediction, frequency of check-ins merely imply the confident level of users' preference for the corresponding location, which makes them barely serve as explicit ratings for locations given from users.

To solve these concerning issues, we propose to design a re-weighting scheme of check-in matrix $\mathcal{R}$ before utilizing the collaborative feature extraction approach.

Specifically, for user u and location l, we update the weights in $\mathcal{R}$ as follows:

$$\mathcal{R}_{u,l} = \begin{cases} 1 + log(\mathcal{R}_{u,l} + 1), & \text{if } \mathcal{R}_{u,l} > 0 \\ 1, & \text{otherwise} \end{cases} \tag{9}$$

,we employ logarithm normalization to smooth the weighted scores.

Then we map the weights in $\mathcal{R}$ into discrete ratings as explicit preferences of users. First, we denote the list of intervals for a user u as $I_u = \{1, 2, ..., M\}$. Then we divide the open scale of check-ins for each user to M intervals with equal width, and get the width of intervals w_u for user u as follows:

$$w_u = \frac{max(\mathcal{R}_{u*}) - min(\mathcal{R}_{u*})}{M} \tag{10}$$

,where $max(\mathcal{R}_{u*})$ and $min(\mathcal{R}_{u*})$ represent the maximal and minimal weight in the vector $\mathcal{R}_{u*}$, respectively.

The lower and upper bound of an interval $i_{uk} \in I_u$ for user u is derived as follows:

$$LOW(i_{uk}) = min(\mathcal{R}_{u*}) + (k-1) * w_u \tag{11}$$

and

$$UP(i_{uk}) = min(\mathcal{R}_{u*}) + k * w_u \tag{12}$$

Finally we set the rates for user u and location l as the index of the interval which includes the corresponding weights.

To summarize the above re-weighting process, we first manage to assign zero check-ins with small weights and non-zero check-ins with larger weights in the check-in matrix $\mathcal{R}$. Then we derive a discrete rating for each user-location item on top of the new weights. The weighted

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2017.2705083, IEEE Transactions on Knowledge and Data Engineering

schema reduces the sparsity and alleviate the later fusion process with other features.

After updating the weight with the rating in $\mathcal{R}$, we then conduct the factorization process to get the latent collaborative features for location prediction. We will report its performance in the experiment section.

4.4 Content Feature

We proceed to present suitable methods to derive the content feature for check-in records. In the new content feature graph, we utilize the Random Walk with restart algorithm [22] to get the required category closeness scores.

Random walk is an iteration method to update node scores across the graph. A random walk on a directed graph consists of a visiting sequence of vertices, from a starting vertex, for each visit, selects an out edge, traverses this edge to the new connected vertex, and repeating the visit process. The commonly used random walk with restart variant adds a restart option to go back to starting vertex in each step.

After several iterations, the random walk with restart process will usually converge to a stable state. Then relative importance of each node can be measured by its stable score.

Considering the directed graph $\mathcal{G}$ we built in Section 3.3, A as its adjacency matrix, $\vec{r}$ as a *preference vector*, inducing a probability distribution over V and λ as the damping factor. 1- λ is the restarting probability. Random Walk vector $\vec{p}$ over all vertexes V is defined as the following equation:

$$\vec{p} = \lambda \cdot A \cdot \vec{p} + (1-\lambda) \cdot \vec{r} \tag{13}$$

In the iteration process, a random walk starts from a node and chooses out edge randomly among the available out-going edges, according to the edge weights. With a restart probability $(1-\lambda)$, it goes back to a node based on the weighted distribution in $\vec{r}$. After several iterations, node scores will converge, which is guaranteed in most cases.

If preference vector $\vec{r}$ is uniform over V, $\vec{p}$ is referred as a global random walk vector. For non-uniform $\vec{r}$, the solution $\vec{p}$ will be treated as a *personalized random walk* vector. By appropriately selecting $\vec{r}$, the rank vector $\vec{p}$ can be made to bias certain nodes [23].

Here in the category transition graph $\mathcal{G}$, to measure one user's potential to visit next category, we conduct a personalized random walk. We model each user with a set of random surfers over the categorical transition graph.

The initial non-zero starting surfers are assigned by user's previous visited category nodes. From these nodes, the random walk process follows the out-edges based on the weighed distribution of these defined transition frequency. After the converge of several iterations, we select top ranked candidate categories C_u. Be aware that, shown in Table 1, the size of category nodes is usually not large, thus the random walk process would not take much time to derive.

The prediction probability of the candidate position l, given the user u, can be measured by the aggregation of the corresponding categories nodes' closeness, derived from the random walk process.

$$p_{cat}(l|u) = \sum_{c \in Tr_u} \sum_{c' \in C_{l'} \cap C_u} sim(c, c') \tag{14}$$

It is noticed that, the proposed transition pattern extraction method is different from the recent temporal sequence mining or Markov chain based pattern mining work [24], [25]. First, we represent the transition patterns in a more global way, with the help of a graph structure. It overcomes the limitation of current chain based check-in pattern mining work. Second, we provide a new extended closeness measurement in a more general form with the help of Random Walk based closeness extraction algorithm. Besides, it opens the combination space of future different features to be included. The actual improvement will be shown in the later experimental study.

5 Experiments

In this section, we report the empirical study on three large real datasets to demonstrate the advantages of our proposed approach. We first introduce the datasets, metrics and baseline methods used in the experiment, then we present the experiments results.

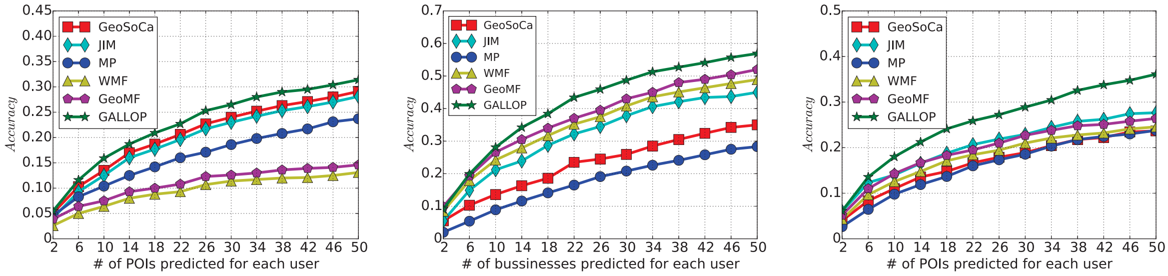
5.1 Experimental Settings

Dataset Setup. Section 2.1 has already introduced the chosen datasets. Here in the experimental study of model training and testing, we pre-process the check-in datasets as follows: Each user's check-in trajectory is split into two parts in effectiveness evaluation. The first parts are used as training set and the remaining are testing set and in our experiments with the default percentage 50%.

Baseline Methods. Several state-of-the-art prediction models are chosen to compare the characteristics and improvements of our proposed method. We select first two methods as the feature oriented baseline and the other three as the fusion oriented competitors, thus we can test the contributions of features and the fusion power of the proposed GALLOP approach.

- **MP**: It models human location data at an individual level using density estimation method [15]. Mixtures of density estimates at different spatial scales are employed to alleviate the problem of data sparsity. This method is used to test the performance of basic context information extraction.
- **WMF**: Weighted Matrix Factorization [20] is an attractive collaborative filtering model applied in traditional recommendation system. It can be used in POI recommendation which takes check-in records as implicit feedbacks and fit both nonzero and zero records in the same way. This method is used to test the performance of collaboration information.

- **JIM**: This is a joint probabilistic generative model [26] which integrates the effect of temporal, content and geographical(context) to model users' check-in behaviors in LBSNs. This model is used to measure the influence of temporal factors in prediction performance. In our experiments, the query time is set to the 48 hours after the most checked-in timestamp.
- **GeoMF**: This model [14] follows the line of factorization of the user-POI check-in matrix and augments



(a) Foursquare (b) Yelp (c) Micro-blog

Fig. 7: Predication Performance of Different Methods

latent space with geographical information. Technically, users' activity areas and POIs' influence areas are incorporated into their latent factors, respectively. This method is used to test the performance of collaboration feature based fusion approach.

- **GeoSoCa** This method [16] exploits geographical, social and categorical correlations for POI recommendation. GeoSoCa models the geographical correlations between different locations by utilizing kernel estimation with an adaptive bandwidth, and models both social check-in frequency/rating and the popularity of POI categories as power-law distribution. To deliver a final result, GeoSoCa integrates three sources of features by product rule. This method is used to test fusion of different features.

Default Parameters. In the proposed mixture model of context feature extraction, we set the default value of $\boldsymbol{k}$ to 3 which means we choose the distance of third nearest neighbour and we utilize a 3-level which represents individual, city and state respectively. To utilize collaborative feature, we set the default value of latent dimension $\mathcal{K}$ to 20. In the personalized random model for extracting content feature, the default value of λ is set to 0.85. The default value of the number of POI predicted for a user is 30.

Evaluation Metric. To compare different prediction implementations, we resort to $Accuracy@\Bbbk$, which is a standard metric in prediction evaluations [27]. Accuracy is crucial, especially for predication scenarios, where most users browse or consider only the first k items.

To compute the metric $Accuracy@\Bbbk$, we denote $hit@\Bbbk$ for a single test case as either the value 1, if the ground truth POI v appears in the top-$\Bbbk$ predicted locations, or the value 0, if otherwise. The overall $Accuracy@\Bbbk$ is defined by averaging over all test cases:

$$Accuracy@\Bbbk = \frac{\#hit@k}{|D_{test}|}.$$

$Accuracy@\Bbbk$ measures what percentage of testing check-in locations can be returned with the highest probability.

5.2 Approach Performance

We first report the $Accuracy$ evaluations on three different datasets individually on Figure 7. As the number of places to be predicted for the end user grows, $Accuracy$ get higher. This is intuitive because more predictions we provide, more places we can cover that users will check-in next.

Apparently, the proposed GALLOP method shows advantages over all three datasets, and the competitor baselines can not win on all. Different baselines get ahead in different datasets. It shows the generality and robustness of our proposed approach.

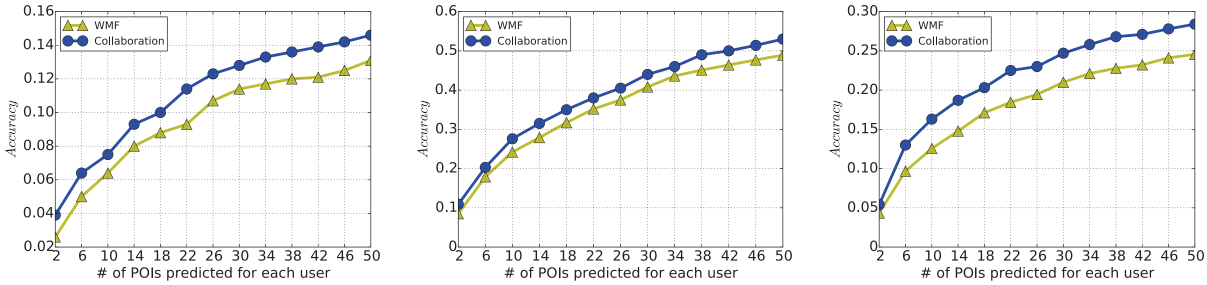
Specifically, Figure 7a presents the results on Foursquare. The WMF and GeoMF has the lowest performance in all three metrics. The MP method can get reasonable improvement respectively. The GeoSoCa and JIM methods are better. The proposed method GALLOP is on the top. Foursquare is a specific check-in service and users report their locations periodically. The locality factors are more impressive and geographical influence plays an important role in the users' mobility patterns. With those methods favoring the spatial features, *i.e.*, GALLOP, GeoSocaa and JIM have advantages in common check-in scenarios. Besides, GALLOP is able to favor the context features under its flexible feature fusion framework.

Figure 7b shows the results on the Yelp dataset. The MP and GeoSoCa methods are falling behind, both of which are context information lines. Matrix Factorization methods, such as GeoMF and WMF and temporal method, JIM get small improvement. The proposed method GALLOP ranks on the top. Yelp is usually used as a restaurant and related rating hub. The check-in activities fall into the food category. The locality factors are relatively small. Instead, the collaborative preference information plays an important role and GALLOP can better tune different features and deliver the satisfying results.

Figure 7c displays the results on the Micro-blog dataset. We can find that the feature combination methods have advantages, *i.e.*, GALLOP method gets top position in $Accuracy$ metric. Micro-blog is a life recording service. As a result, the check-in activities vary greatly, since the check-in records include local behaviors and remote travels. Though the locality factors and interest preference make the problem more complex, this experiment demonstrates the versatility of our proposed approach and shows a significant advantage over the other baselines.

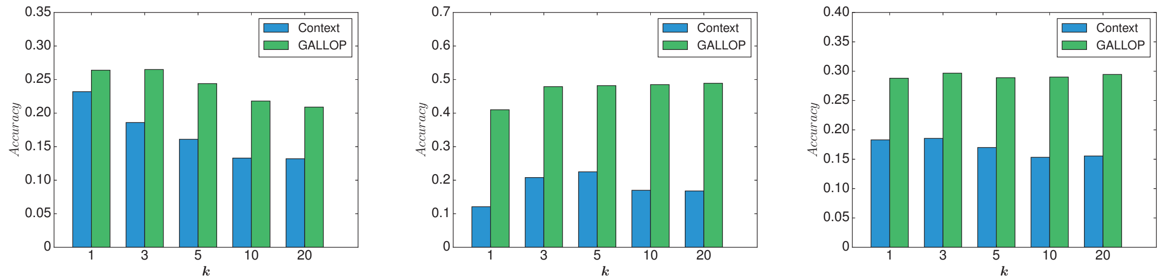
Besides the overall fused prediction performance, we also test GALLOP's individual feature extraction performance. Due to the space limits, we report the collaborative feature extraction comparison. Figure 8 shows the difference between WMF and our proposed collaboration extraction

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2017.2705083, IEEE Transactions on Knowledge and Data Engineering



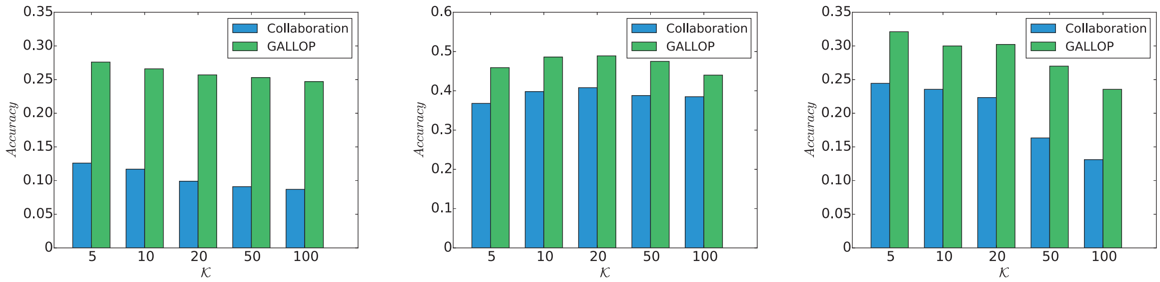
(a) Foursquare (b) Yelp (c) Micro-blog

Fig. 8: Performance Comparison of Two Collaboration Extraction Methods



(a) Foursquare (b) Yelp (c) Micro-blog

Fig. 9: Content Parameter Sensitivity



(a) Foursquare (b) Yelp (c) Micro-blog

Fig. 10: Collaborative Parameter Sensitivity

methods on three different datasets. We can find that, the proposed approach has remarkable improvement.

To conclude the effectiveness evaluation studies, the proposed GALLOP approach shows significant advantage over all the other competitors in different datasets. GALLOP has impressive and stable performance over different datasets, which are vastly different in check-in characteristics.

The global feature selection mechanism and the adaptive feature fusion framework is useful to cope with the variance challenges in the current check-in location prediction tasks.

5.3 Training Ratio

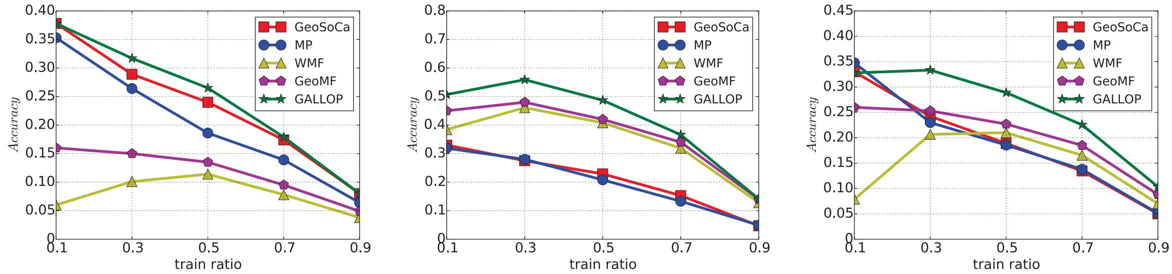
5.4 Parameter Sensitivity

We test the sensitivity of model parameters used in our approach. Besides the other default values, two parameters are manually tuned, *i.e.*, the number of nearest neighbors count $\boldsymbol{k}$ in the context feature extraction and the latent dimensions $\mathcal{K}$ in collaborative feature extraction.

Figure 9 shows the effects of varying $\boldsymbol{k}$ in three datasets. The context feature component and the proposed approach GALLOP are included. We can find that, different datasets have their own optimal $\boldsymbol{k}$ to achieve the highest $Accuracy$. Larger $\boldsymbol{k}$ value does not promise a better performance due to the limited activity scope of the individual user. Note that the performance of Yelp also does not change significantly because context information contributes less in this dataset.

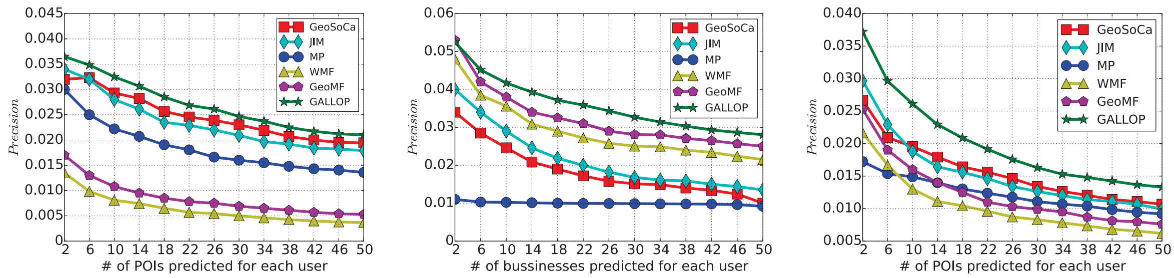
Nevertheless, we also include the experiments of latent dimension $\mathcal{K}$ affects. Figure 10 illustrates the similar phenomenon, though the proposed method looks a bit decline with more dimensions. GALLOP shows better performance

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2017.2705083, IEEE Transactions on Knowledge and Data Engineering



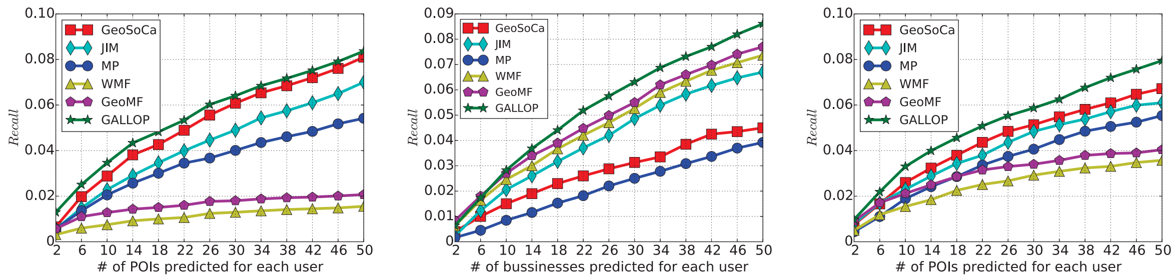
(a) Foursquare (b) Yelp (c) Micro-blog

Fig. 11: Accuracy Sensitivity under different Training Ratio



(a) Foursqaure (b) Yelp (c) Micro-blog

Fig. 12: Predication Effectiveness, $Precision$



(a) Foursquare (b) Yelp (c) Micro-blog

Fig. 13: Predication Effectiveness, $Recall$

and general stableness.

These parameter studies demonstrate that the designed model structure also contributes the proposed method's advantage, besides the integration of more features.

We continue to test the models' training percentage choice. In the commonly used training/testing setting, we have ordered users' check-in trajectory based on the time order and split a user's first part as training and the remaining as testing dataset. Here we vary the ratio of training and testing subsets.

Figure 11 shows $Accuracy$ changes over different percentages of training datasets. In general, GALLOP always exhibits best performance. GeoSoCa and MP methods are good at low training ratio, while GeoMF and WMF are preferable at large training ratio. In each dataset, the $Accuracy$ of GeoSoCa and MP methods would fall with increasing training ratio. On the other side, the matrix factorization methods and proposed methods first exhibit an increase and then gradually decline with the increasing training ratio. Though it may seem wrong at the first glance, the empirical results still make sense. The score changes are caused by the data drifting phenomenon on check-in data. That is to say, users don't provide much repeated check-in reports in their trajectory and the old data would not benefit the future prediction.

In the difference of these methods, the spatial influence lines methods are not capable for the long trend prediction and the collaborative factorization approaches are better at the latent long trend mining. The combination of spatial like context features, transition based content information and the collaborative factorization information can benefit the check-in activity prediction across different datasets. These results also prove the advantage of the proposed GALLOP combination method in another perspective.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2017.2705083, IEEE Transactions on Knowledge and Data Engineering

5.5 Effectiveness Study

Beside the $Accuracy$ metric, here we include other ones to systematically reveal the effectiveness and characteristics of the proposed approach and state of the art ones. We choose $Precision@\mathbb{k}$ and $Recall@\mathbb{k}$ to perform more effectiveness study.

Figure 12-13 present the $Precision$ and $Recall$ experiments individually. As the number of candidate places grows, $Precision$ becomes lower while instead $Recall$ gets higher. It is intuitive because more predictions we provide, more places we can cover that users will check-in next, but some predicted locations are less possible to be visited by users and increase the information overload for the end users.

In accord with the $Accuracy$ study, the proposed GAL-LOP method shows advantages over all three datasets, but the competitor baselines can not win on all. Different baselines get ahead in different datasets. These experiments show the generality and robustness of our proposed approach. Admittedly, the $Precision$ and $Recall$ scores are still not very high, due to the scope of check-in scenarios.

To conclude the experimental study, the proposed GAL-LOP shows advantage in both the model structure and model input. It has better performance in different dataset settings and guarantees the usefulness in the parameter tuning and training process. These make it rather versatile for many check-in prediction applications.

6 Related Work

The location prediction work in this paper is related to several recent research directions. We briefly review them as follows:

Check-in Data Analysis: With the availability of more and more users' mobility and check-in records in recent years, the value of check-in data analysis becomes attractive both in the user behavior understanding and spatial data processing.

Empirical analysis of user check-in and mobility activities was investigated in [8], [12], [13]. The authors identified the scaling shrinkage patterns in users' mobility activities and also reported the correlations of these activities with users' home and office locations. The fact that the users' check-in activities usually have a high centralized distribution around several hubs is revealed. The influence of social network to the check-in activities was discussed in [9]. The authors reported the locality behaviors between home and office like nearby areas are more periodical, and social network friendship contributes some proportion to long distance travels.

To cope with the rich check-in data, many spatial enriched techniques become possible. The co-located patterns and temporal changes were investigated in [28], [29]. Nearest neighbor queries are always attractive in both the spatial data processing and the recent check-in scenarios. [30], [31] discussed new scalable methods for the recent trajectory data tornado.

Check-in data can ignite many new types of applications. The spatial object tagging problem was illustrated in [32]. Users' check-in sequence, *i.e.*, trajectory logs can be divided and extracted. The efficient ways include subset selection and semantic enrichment [33]. The extraction of local events was studied in [34]. Location Tagging with social information was proposed in [35] to improve location estimation. To infer users' interest and the retail location, human mobility data from Foursquare were carefully analyzed to understand the popularity of retail store chains in a metropolis in terms of number of check-ins [36]. [37] discussed the real-time query processing on the micro-blog data.

In this paper, we discuss the mobility differences between several check-in scenarios and reveal their trends and characters. Then we proceed to design the general and global feature fusion method of our work.

Location Prediction: Driven by users' more active check-in behaviors and desire for assistant in their ordinary life, place recommendation and location prediction methods become an increasing hot topic in the recent years [10], [11]. As we discussed in Introduction, there are several lines of prediction methods in the location prediction problem.

First category of work is derived from the check-in data's inherent distance closeness and spatial distribution to generate prediction results. For example, in [24], a location feature based method was introduced. [15] extended a kernel density estimation model to profile the spatial distribution. These methods follow the hub distribution of user's mobility behaviors and always deliver stable results but the temporal dynamics and various scenarios hinder its performance to a higher level.

Approaches based on *geographical closeness* perform the prediction task by taking advantage of geographical feature mining. For example, In [38], Jeung et al. combined the motion function and the extracted objects movement patterns to estimate a moving objects future locations. In [24], Monreale et al. utilized the trajectory patterns equipped with temporal information as predictive rules. [28] proposed a unified framework to include various temporal tightness patterns. In these methods, the authors usually defined a spatial trajectory as an ordered-sequence of locations, and extract associate rules from historical trajectory dataset using two different methods. Given a query sub-trajectory, he choose a best-match historical trajectory based on the associate rules for prediction.

Besides these data driven trajectory pattern mining, some generative models were also utilized to infer the trajectory patterns for better location prediction [25], [39], [40]. Since here a challenging problem is the data sparsity in the check-in records. Direct data driven mining usually fails to provide a robust prediction solution.

In the model training, [41] chose the whole data space as a set of disjoint square cells. All the locations that locate within the same grid are treated as the same position. Then the authors defined the state transition model, and infer the transition probabilities for later prediction evaluation. Another way [25] was to take the predictive task as inferring the cell that the end point of the query trajectory locates. A formal formula is provided to get the probability of a particular cell being the destination given a query trajectory.

Another line of prediction work is the extension of collaborative filtering and feature factorization methods commonly used in other recommendation tasks [20]. Methods include Probabilistic Latent Semantic Analysis used in [42], Matrix Factorization with spatial features proposed

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2017.2705083, IEEE Transactions on Knowledge and Data Engineering

in [14], and the re-rank metrics guided factorization in [43]. Inevitably, temporal affects of check-in activities were also investigated in [44].

Feature Fusion based Prediction: The bottleneck of traditional prediction methods are lacking of flexibility to cope with the noise and sparsity in the location prediction. Some recent work resorted to feature fusion model for stable location prediction solutions. In [14], a factorization method was introduced to include the check-in matrix and augmented the raw latent space with geographical information. Technically, users' activity areas and places' influence areas are incorporated into their latent factors, respectively. Another work exploited spatial, social and place correlations for location recommendation [16]. The authors modeled the spatial correlations between different location by utilizing spatial features, and included both social information and place attributes for the combined prediction.

By the fusion of multiple features, the recommendation performance can usually be improved [11]. Besides the above methods, other general solutions also include the new model structures. For example, regularity and conformity were discussed in [17], in which the periodical and abnormal patterns are modeled and combined in a unified framework. Recently, a novel embedding method to transfer the representation of different check-in features into a new dimensional space was proposed [45].

The approaches proposed in these paper followed recent feature combination line. However, we focus on the global solution, in which more scales of features are utilized and generalized to tackle the variance of different scenarios, guaranteeing the prediction performance.

7 Conclusions

This paper presents a new feature fusion method for location prediction problem. We systematically analyze the check-in characteristics of different scenarios and propose to model three categories of features and combine them in a global way. The geographical, collaborative and categorical information are all utilized. We propose new models to include more global features to improve the generality and robustness of the prediction method. Besides, the approach is versatile and easy to extend. It shows impressive advantage on different datasets and significantly improve the prediction accuracy.

This research has several interesting future directions. For example, better ways to improve the feature preprocessing stage and design the compact structure to maintain the extracted features. It is also valuable to exploit the evolving factors in the location prediction. Additionally, the feature extraction methods we proposed in this work can be extended to enable incremental updating. And new comprehensive location prediction and update setting can be utilized.

References

[1] J. Sang, T. Mei, J.-T. Sun, C. Xu, and S. Li, "Probabilistic sequential pois recommendation via check-in data," in *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*, 2012, pp. 402–405.

[2] X. Liu, Y. Liu, K. Aberer, and C. Miao, "Personalized point-of-interest recommendation by mining users' preference transition," in *Proc. of CIKM*, 2013, pp. 733–738.

[3] J.-D. Zhang, C.-Y. Chow, and Y. Li, "Lore: exploiting sequential influence for location recommendations," in *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2014, pp. 103–112.

[4] W. Wang, H. Yin, S. Sadiq, L. Chen, M. Xie, and X. Zhou, "SPORE: A sequential personalized spatial item recommender system," in *Proc. of ICDE*, 2016, pp. 954–965.

[5] D. Lian, X. Xie, V. W. Zheng, N. J. Yuan, F. Zhang, and E. Chen, "CEPR: A collaborative exploration and periodically returning model for location prediction," *TIST*, vol. 6, no. 1, p. 8, 2015.

[6] A. Noulas, S. Scellato, N. Lathia, and C. Mascolo, "Mining user mobility features for next place prediction in location-based services," in *Proc. of ICDM*. IEEE, 2012, pp. 1038–1043.

[7] J. Ye, Z. Zhu, and H. Cheng, "Whats your next move: User activity prediction in location-based social networks," in *Proc. of SDM*, 2013.

[8] A. Noulas, S. Scellato, C. Mascolo, and M. Pontil, "An empirical study of geographic user activity patterns in Foursquare," in *Proc. of ICWSM*, 2011, pp. 70–573.

[9] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: user movement in location-based social networks," in *Proc. of SIGKDD*, 2011, pp. 1082–1090.

[10] J. Bao, Y. Zheng, D. Wilkie, and M. Mokbel, "Recommendations in location-based social networks: A survey," *Geoinformatica*, vol. 19, no. 3, pp. 525–565, Jul. 2015.

[11] D. Lian, X. Xie, F. Zhang, N. J. Yuan, T. Zhou, and Y. Rui, "Mining location-based social networks: A predictive perspective," *IEEE Data Engineering bulletin*, vol. 38, no. 2, pp. 35–46, 2015.

[12] D. Brockmann, L. Hufnagel, and T. Geisel, "The scaling laws of human travel," *Nature*, vol. 439, no. 7075, pp. 462–465, 2006.

[13] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, pp. 779–782, 2008.

[14] D. Lian, C. Zhao, X. Xie, G. Sun, E. Chen, and Y. Rui, "GeoMF: Joint geographical modeling and matrix factorization for point-of-interest recommendation," in *Proc. of SIGKDD*, 2014, pp. 831–840.

[15] M. Lichman and P. Smyth, "Modeling human location data with mixtures of kernel densities," in *Proc. of SIGKDD*, 2014, pp. 35–44.

[16] J.-D. Zhang and C.-Y. Chow, "GeoSoCa: Exploiting geographical, social and categorical correlations for point-of-interest recommendations," in *Proc. of SIGIR*, 2015, pp. 443–452.

[17] Y. Wang, N. J. Yuan, D. Lian, L. Xu, X. Xie, E. Chen, and Y. Rui, "Regularity and conformity: Location prediction using heterogeneous mobility data," in *Proc. of SIGKDD*, 2015, pp. 1275–1284.

[18] B. W. Silverman, *Density estimation for statistics and data analysis*. CRC press, 1986.

[19] R. A. Finkel and J. L. Bentley, "Quad trees a data structure for retrieval on composite keys," *Acta informatica*, vol. 4, no. 1, pp. 1–9, 1974.

[20] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30–37, Aug. 2009.

[21] D. Saad, "Online algorithms and stochastic approximations," *Online Learning*, 1998.

[22] R. Motwani and P. Raghavan, *Randomized algorithms*. Chapman & Hall/CRC, 2010.

[23] T. H. Haveliwala, "Topic-sensitive PageRank," in *Proc. of WWW*, 2002, pp. 517–526.

[24] A. Monreale, F. Pinelli, R. Trasarti, and F. Giannotti, "Wherenext: a location predictor on trajectory pattern mining," in *Proc. of SIGKDD*, 2009, pp. 637–646.

[25] M. Li, A. Ahmed, and A. J. Smola, "Inferring movement trajectories from GPS snippets," in *Proc. of WSDM*, 2015, pp. 325–334.

[26] H. Yin, X. Zhou, Y. Shao, H. Wang, and S. Sadiq, "Joint modeling of user check-in behaviors for point-of-interest recommendation," in *Proc. of CIKM*, 2015, pp. 1631–1640.

[27] C. D. Manning, P. Raghavan, H. Schütze *et al.*, *Introduction to information retrieval*. Cambridge university press Cambridge, 2008.

[28] J. Lee, J. Han, and X. Li, "A unifying framework of mining trajectory patterns of various temporal tightness," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 6, pp. 1478–1490, 2015.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2017.2705083, IEEE Transactions on Knowledge and Data Engineering

[29] S. Barua and J. Sander, "Mining statistically significant co-location and segregation patterns," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 5, pp. 1185–1199, 2014.

[30] J. Niedermayer, A. Züfle, T. Emrich, M. Renz, N. Mamoulis, L. Chen, and H. Kriegel, "Probabilistic nearest neighbor queries on uncertain moving object trajectories," *PVLDB*, vol. 7, no. 3, pp. 205–216, 2013.

[31] Y. Fang, R. Cheng, W. Tang, S. Maniu, and X. S. Yang, "Scalable algorithms for nearest-neighbor joins on big trajectory data," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 3, pp. 785–800, 2016.

[32] D. Flatow, M. Naaman, K. E. Xie, Y. Volkovich, and Y. Kanza, "On the accuracy of hyper-local geotagging of social media content," in *Proc. of WSDM*, 2015, pp. 127–136.

[33] C. Zhang, J. Han, L. Shou, J. Lu, and T. La Porta, "Splitter: Mining fine-grained sequential patterns in semantic trajectories," *PVLDB*, vol. 7, no. 9, 2014.

[34] J. Foley, M. Bendersky, and V. Josifovski, "Learning to extract local events from the web," in *Proc. of SIGIR*, 2015, pp. 423–432.

[35] J. McGee, J. Caverlee, and Z. Cheng, "Location prediction in social media based on tie strength," in *Proc. of CIKM*, 2013, pp. 459–468.

[36] D. Karamshuk, A. Noulas, S. Scellato, V. Nicosia, and C. Mascolo, "Geo-Spotting: Mining online location-based services for optimal retail store placement," in *Proc. of SIGKDD*, 2013, pp. 793–801.

[37] A. Magdy, M. F. Mokbel, S. Elnikety, S. Nath, and Y. He, "Venus: Scalable real-time spatial queries on microblogs with adaptive load shedding," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 2, pp. 356–370, 2016.

[38] H. Jeung, Q. Liu, H. T. Shen, and X. Zhou, "A hybrid prediction model for moving objects," in *Proc. of ICDE*, 2008, pp. 70–79.

[39] E. Horvitz and J. Krumm, "Some help on the way: Opportunistic routing under uncertainty," in *Proc. of UbiComp*, 2012, pp. 371–380.

[40] B. D. Ziebart, A. L. Maas, A. K. Dey, and J. A. Bagnell, "Navigate like a cabbie: Probabilistic reasoning from observed context-aware behavior," in *Proc. of UbiComp*, 2008, pp. 322–331.

[41] J. Krumm and E. Horvitz, "Predestination: Inferring destinations from partial trajectories," in *Proc. of UbiComp*, 2006, pp. 243–260.

[42] J. McInerney, A. Rogers, and N. R. Jennings, "Improving location prediction services for new users with probabilistic latent semantic analysis," in *Proc. of UbiComp*, 2012, pp. 906–910.

[43] W. Wang, H. Yin, L. Chen, Y. Sun, S. Sadiq, and X. Zhou, "Geo-SAGE: A geographical sparse additive generative model for spatial item recommendation," in *Proc. of SIGKDD*, 2015, pp. 1255–1264.

[44] Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. M. Thalmann, "Time-aware point-of-interest recommendation," in *Proc. of SIGIR*, 2013, pp. 363–372.

[45] F. Figueiredo, B. Ribeiro, J. M. Almeida, and C. Faloutsos, "Tribeflow: Mining & predicting user trajectories," in *Proc. of WWW*, 2016, pp. 695–706.

Yuxing Han received BS degree in School of Computer Science and Software Engineering from East China Normal University in 2013. He is currently working toward the PhD degree in East China Normal University. His research interest includes spatial-temporal data mining and location-based services analysis.

Junjie Yao received BS degree from Department of Computer Science, Nankai University in 2007 and PhD degree from Institute of Network, Peking University in 2013. He is currently an Associate Professor at School of Computer Science and Software Engineering, East China Normal University. Before joining ECNU, he was a Post Doc visitor at UC Santa Barbara. His current research interest includes Data Mining, Information System and Social Media Analysis. He is a member of IEEE.

Liping Wang received the PhD degree in Computer Application and Technology from East China Normal University in 2013. She is currently an Associate Professor in the School of Computer Science and Software Engineering, East China Normal University. Her research interests include temporal spatial data management and Pattern Recognition.

Xuemin Lin received the BSc degree in Applied Math from Fudan University in 1984. During 1984-1988, he studied for the PhD degree in Applied Math at Fudan University. He received the PhD degree in Computer Science from the University of Queensland in 1992. He is a Scientia Professor in the School of Computer Science and Engineering, University of New South Wales (UNSW). He has been the head in the Database Research Group, UNSW since 2002. He is a concurrent professor in the School of Computer Science and Software Engineering, East China Normal University. Before joining UNSW, he held various academic positions at the University of Queensland and the University of Western Australia. He is currently the Editor-in-Chief of IEEE Transactions on Knowledge and Data Engineering. His current research interests lie in data streams, approximate query processing, spatial data analysis, and graph visualization. He is an IEEE Fellow.