Inferring and Exploiting Categories for Next Location Prediction

Ankita Likhyani[1] Deepak Padmanabhan[2] Srikanta Bedathur[2] Sameep Mehta[2]
[1]IIIT-Delhi, [2] IBM Research - India
ankital@iiitd.ac.in, {deepak.s.p,sbedathur,sameepmehta}@in.ibm.com

ABSTRACT

Predicting the next location of a user based on their previous visiting pattern is one of the primary tasks over data from location based social networks (LBSNs) such as Foursquare. Many different aspects of these so-called "check-in" profiles of a user have been made use of in this task, including spatial and temporal information of check-ins as well as the social network information of the user. Building more sophisticated prediction models by enriching these check-in data by combining them with information from other sources is challenging due to the limited data that these LBSNs expose due to privacy concerns.

In this paper, we propose a framework to use the location data from LBSNs, combine it with the data from maps for associating a set of venue categories with these locations. For example, if the user is found to be checking in at a mall that has cafes, cinemas and restaurants according to the map, all these information is associated. This category information is then leveraged to predict the next checkin location by the user. Our experiments with publicly available check-in dataset show that this approach improves on the state-of-the-art methods for location prediction.

Categories and Subject Descriptors

H.2.8 [**DATABASE MANAGEMENT**]: Database Applications; J.4 [**COMPUTER APPLICATIONS**]: Social And Behavioral Sciences

Keywords

Category Information; Human Mobility; Location based Social Networks

1. INTRODUCTION

The location prediction problem in LBSNs has been widely studied in recent years. This has wide applications in areas such as targeted marketing since knowing that a specific user's next check-in is likely to be at a cinema could be used to prioritise movie ticket offers to be sent to her. LBSNs typically do not expose the check-in information of users; most popular LBSNs such as FourSquare[1] and BrightKite[2] have their datasets publicly available that are annonymised, wherein only the latitude and longitude of each check-in is available because of privacy concerns. The unavailability of the category (viz., restaurant, cinema etc.) information with historical check-ins in LBSN data has led to development of location prediction methods that do not look to leverage category information. However, it may be observed that map information from sources such as Google Maps (or FourSquare itself) have information about popular venues such as restaurants and cinemas along with their lat-long information. Thus, approximate spatial joins across check-in and map data can be intuitively used to infer category information for each check-in. Such joins are restrained to provide category information only at a coarse granularity since there could be multiple types of venues (e.g., restaurants, coffee-shops) at the same location (e.g., a mall or a food court). In this paper, we consider the problem of inferring category information with each check-in and leveraging such coarse-grained category information to improve location prediction in LBSNs. The location prediction model in general does not perform good because of sparisity in the check-in data. Therefore, analysis on sparse and annonymised data is useful to achieve atleast some improvement. To the best of our knowledge this is the first work that exploits such coarse-grained category information for next check-in location prediction.

features \ techniques	PMM and PSMM [1]	SHM [3]	SHM+T [2]	M5 Trees [5]	gSCorr [4]
Social Correlation	✓	✓	✓	✓	✓
Geographic Distance	✓		✓		✓
User Mobility	✓	✓	✓	✓	✓
Category Information				✓	
Periodic Patterns	✓		✓	✓	

Table 1: List of features used in different techniques

Related Work: In Table 1 we have broadly divided and summarised the feature space that has been used for location prediction problem by state-of-the-art methods. It can be seen from table 1 that only Noulas et al in [5] have explored the category information in predicting where the user would go next. However, [5] assume that the check-in information contains not only fine-grained category information (e.g., restaurant, cinema etc.) but also contains information about the precise venue the user has checked-in, i.e., the name of the restaurant or cinema. Thus, this technique is impractical in common cases where we have check-ins specified only at the lat-long level. The other methods listed in Table 1 are able to work with check-in data at the level of lat-long information.

[1]http://www.public.asu.edu/∼hgao16/dataset.html
[2]http://snap.stanford.edu/data/loc-brightkite.html

WWW 2015 Companion, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05.
http://dx.doi.org/10.1145/2740908.2742770.

Techniques / Datasets	SHM	CLM	SHM +T			SHM+CLM	(SHM+T)+CLM		
			SHM+D	SHM+W	SHM+DW		(SHM+D)+CLM	(SHM+W)+CLM	(SHM+DW)+CLM
FourSquare (Mar'10-Jan'11)	22.45	23.50	23.70	23.52	23.81	24.96	24.98	24.87	24.88
FourSquare (Jan'11-Dec'11)	30.45	24.87	31.59	31.31	32.03	31.32	32.36	32.21	32.74
BrightKite (Mar'08-Oct'10)	23.25	21.22	24.21	24.51	24.61	24.38	25.74	25.54	25.57

Table 2: Comparison in terms of Accuracy (in %) for different models

2. METHODOLOGY

We first discuss inferring coarse-grained category information by correlating lat-long check-in data with map information. Next, we briefly describe the SHM+T model proposed in [3] and describe our method that extends SHM+T by exploiting category information to improve location prediction accuracy.

2.1 Inferring Categories

We use a publicly available annonymised check-in dataset [3, 1] where each check-in has only a lat-long representation. We associate each check-in $l = [lat, long]$ with a set of categories of venues (using FourSquare API) that are within 50 meters radius of the location. i.e.,
$cat(l) = \{v.category | distance(v, l) \leq 50m\}$, where, v represents venue. Empirically, we found that each location gets associated with approximately k categories. The average value of k for the three datasets i.e. FourSquare (Mar'10-Jan'11), FourSquare (Jan'11-Dec'11) and BrightKite (Mar'08-Oct'10) is 22, 13 and 9, respectively.

2.2 Data Modelling

We denote the set of categories associated with a location l as C_l. If the last location is l', we find a likely category c' for l'. Then we determine a category c that is likely to be followed after c' (for example, restaurant after cinema). Lastly, we find a location l that is likely for the category c. This is aggregated over multiple values for c and c' to associate a probability with each value of l, as follows:

$$P_C(l) = \sum_{c,c'} P_u(l|c).(\alpha * P_u(c|c') + (1-\alpha) * P_g(c|c')).P(c'|l'), \quad (1)$$

where $P_u(c|c')$ and $P_g(c|c')$ denote the probability of the user visiting a location of category c right after visiting one of category c', as estimated using the user's own check-in history, and the global check-in history across all users respectively. α is an interpolation parameter that determines the relative weighting of P_u and P_g. We estimate $P(c'|l')$ and $P(l|c)$ as follows:

$$P(c'|l') = \frac{\text{number of venues of category c close to } l}{\text{total number of venues close to } l}$$

$$P(l|c) = \frac{\text{number of venues of category c close to } l}{\sum_{loc} \text{number of venues of category c close to } loc}$$

We will call $P_C(.)$ as the category language model location prediction method; as in the case of other models, once this distribution is estimated, the most probable location could be recommended.

2.3 Leveraging Category Language Model (CLM)

Using just the CLM for location prediction is not likely to be effective since it does not model other aspects of user movement. Thus, we devise a method to leverage CLM with SHM+T [2], a state-of-the-art model for location prediction that uses social, historical and temporal information. The SHM+T model may be simplistically represented as a function $P(l|t, H_{u,t}, S_{u,t})$ that estimates a probability distribution of next location using: (1) t, the time of the day and week, (2) $H_{u,t}$, the user's check-in history and (3), $S_{u,t}$, the user's social circle.

We combine SHM+T and CLM by estimating the combined distribution as a weighted sum of the distributions by the individual models, with the relative weighting determined by λ as follows:

$$P_{SHM+T+CLM}(l) = \lambda * P(l|t, H_{u,t}, S_{u,t}) + (1-\lambda) * P_C(l) \quad (2)$$

3. EXPERIMENTAL EVALUATION

Implementation Details: We have tested the proposed model over all the users who have made at least 10 check-ins. For each test user, we divide her checkin history into 4:1, where 80% of checkins are used for training and rest 20% are used for testing. Note that, checkins are sorted chronologically. *Accuracy* is used as the evaluation metric.

Results are reported in table 2. T in table 2 denotes the periodic patterns i.e. Daily(D), Weekly(W) and Daily-Weekly(DW). For CLM and SHM+T+CLM model, $\alpha = 0.6$ and $\lambda = 0.7$ (in equations 1 and 2 respectively) are used. At these values we obtain the best performance. It can be observed that SHM+T+CLM achieves 1-2 % of improvement over SHM+T, that is similar to quantum of improvement that SHM+T has achieved over SHM in [2].

4. CONCLUSION AND FUTURE WORK

In this paper, we have inferred coarse-grained category information and have leveraged this information for enhancing the performance of the existing state of the art methods for next-checkin location prediction. In continuation of this work, we would like to explore other auxiliary information such as text information (reviews) that can be used to improve the current state of the art.

5. REFERENCES

[1] E. Cho, S. A. Myers, and J. Leskovec. Friendship and mobility: User movement in location-based social networks. In *KDD*, 2011.

[2] H. Gao, J. Tang, X. Hu, and H. Liu. Modeling temporal effects of human mobile behavior on location-based social networks. In *CIKM*, 2013.

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

[4] H. Gao, J. Tang, and H. Liu. gSCorr: Modeling Geo-social Correlations for New Check-ins on Location-based Social Networks. In *CIKM*, 2012.

[5] A. Noulas, S. Scellato, N. Lathia, and C. Mascolo. Mining user mobility features for next place prediction in location-based services. In *ICDM*, 2012.