On Interesting Place Finding in Social Sensing: An Emerging Smart City Application Paradigm

Chao Huang, Dong Wang
Department of Computer Science and Engineering
University of Notre Dame
Notre Dame, IN 46556, USA
chuang7@nd.edu, dwang5@nd.edu

***Abstract*—Social sensing has emerged as a new application paradigm for smart cities where a crowd of social sources (humans or devices on their behalf) collectively contribute a large amount of observations about the physical world. This paper focuses on an *interesting place finding* problem in social sensing where the goal is to accurately identify the interesting places in a city where people may have strong interests to visit (e.g., parks, museums, historic sites, scenic trails, etc.). Solving this problem is not trivial because (i) many interesting places are not necessarily frequently visited by the average people and hence less likely to be found by the traditional recommendation systems; (ii) the user's social connections could directly affect their visiting behavior and the interestingness judgment of a given place. In this paper, we develop a new *Social-aware Interesting Place Finding (SIPF)* approach that solves the above problem by explicitly incorporating both the user's *travel experience* and *social relationship* into a rigorous analytical framework. The evaluation results showed that the new approach significantly outperforms the state-of-the-arts using two real-world datasets collected from location-based social network service.**

***Keywords*-Interesting Place Finding, Social Sensing, Smart City, Maximum Likelihood Estimation**

I. Introduction

This paper develops a new principled approach to accurately identify interesting places in a city through social sensing applications. This work is motivated by the emergence of social sensing as a new smart city application paradigm of collecting observations about the physical world from social sources (humans or devices on their behalf) [16]. This paradigm is enabled by a few recent technical trends: (i) the proliferation of smart devices (e.g., smartphones) owned by average individuals; (ii) the ubiquitous coverage of wireless communication (e.g., 4G, WiFi, WiMax); (iii) the advent of online social media (e.g., Twitter, Foursquare, Facebook). For example, common citizens can now easily use a Location-Based Social Network (LBSN) service (e.g., Foursquare) on their mobile phones to upload the "check-in" points at the places they visit in a city. A key challenge in social sensing lies in correctly identifying the interesting and useful information from the massive, noisy and unvetted data contributed by the crowd. In this paper, we focus on an *interesting place finding* problem where the goal is to correctly identify the interesting locations in a city where people may have strong interests to visit (e.g., parks, museums, historic sites, scenic trails, etc.). The results of this work can be used to develop various smart city applications (e.g., smart city navigation systems, intelligent travel recommendation systems, mobile guide systems, etc) [8], [9].

Previous work in information retrieval [15], [29] and social sensing [17], [20] have made significant efforts to address the interesting place finding problem using the *crowdsourcing* methods. The main idea behind those approaches is to automatically infer the locations of interesting places in a city from the check-in points or GPS traces that users share using location-aware applications [6]. The advantages of using crowdsourcing methods compared to the traditional methods (e.g., search engine, travel websites) are threefold. *First*, the cost of data collection using crowdsorucing is low [12]. *Second*, the interestingness of a place may change over time and the crowdsourcing methods can track such changes by analyzing the most recent trajectory data uploaded by the crowd [26]. *Third*, the crowdsourcing traces normally have a better spatial-temporal coverage of the interesting places as the crowd are naturally distributed across the region [15].

However, two important limitations have not been fully addressed in the crowdsourcing methods. *First*, the current techniques are mostly heuristic-based and make strong assumptions when they handle users in the problem. For example, they either assume all users have exactly the same travel experience [1] or the correlation between a user's travel experience and the number of places he/she visited is simply linear [29]. This problem becomes more challenging when neither the user's travel experience nor the interestingness of a place is known *a priori* in social sensing [28]. Hence, we need to develop a new framework that can accurately model both the user's travel experience and the interestingness of places based on the social sensing data observed. *Second*, the social connections between users could easily affect their visiting behavior and the judgment on the interestingness of

[1]The travel experience of a user is highly correlated the user's ability to find interesting places in a city [11]

places they visited. For example, a group of colleagues who work in the same company are more likely to visit the same building every day. However, the building of their company may not necessarily be interesting to the general public. Unfortunately, current interesting place finding techniques totally ignored the impact of user's social dependency in their solutions.

In this paper, we develop a Social-aware Interesting Place Finding (SIPF) scheme that addresses the above limitations by explicitly incorporating both the user's *travel experience* and *social dependency* into a maximum likelihood estimation framework. We evaluate the SIPF using two real world datasets collected from online social network services. The evaluation results show that our approach significantly outperforms the state-of-the-art baselines by correctly identifying more interesting places in a city while keeping the false positives the least. The results of this paper are important because they allow social sensing applications to accurately identify interesting places by taking into account the user's travel experience and social dependency under a principle framework. The contributions of this work are as follows:

- To the best of our knowledge, we are the first to explicitly consider both the user's travel experience and social dependency in the interesting place finding problem in social sensing.
- We develop a principled framework that allows us to derive an optimal solution (in the sense of maximum likelihood estimation) for the social-aware interesting place finding problem.
- We show non-trivial performance gains achieved by our SIPF scheme (i.e., the SIPF scheme increased the interesting place identification precision by up to 36% and the recall by up to 20% compared to the state-of-the-art baselines on real world datasets.).

The rest of this paper is organized as follows: we discuss the related work in Section II. In Section III, we present the new social-aware interesting place finding problem in social sensing applications. The proposed maximum likelihood estimation framework and the expectation maximization solution is presented in Section IV. Evaluation results are presented in Section V. Finally, we conclude the paper in Section VI.

II. Related Work

Social sensing has emerged as a new smart city application paradigm of crowdsourcing sensory measurements about the physical world from humans or devices on their behalf [2], [16]. Current recent research in social sensing addresses various challenges such as privacy perseverance [4], information dissemination [19], efficient communication [23], data provenance [18], and data sparsity problem in multi-dimensional social sensing spaces [3]. An emerging problem of finding interesting places through social sensing applications arises recently due to the proliferation of location-based social network services (e.g., Foursquare, Google Places, Gowalla) [1]. They empower common citizens to easily share their location and visiting information through mobile devices. To address this emerging problem, this paper develops a social-aware interesting place finding scheme that explicitly exploits both the user's travel experience and social dependency under a rigorous analytical framework.

In information retrieval and data mining, there exists a good amount of work on the topic of mining geo-spatial data traces to identify the points of interests (POI). For example, a *Hub and Authorities* based method has been proposed to recommend interesting places for visitors [29]. Zhang et al. [27] developed a novel method to predict links across partially aligned location-based social networks and address the data sparsity problem in POI recommendations. Furthermore, Tiwari et al. used the semantic features of geo-spatial regions to mine places of potential interests [15]. However, the above works either assumed *linear* correlations between the user's travel experience and the interestingness of places or ignored the *social dependency* between users. In contrast, this paper explicitly models both the *nonlinear* relationship between the user's travel experience and the interestingness of places and the user's social dependency, which is shown to greatly improve the accuracy of the interesting place finding results.

Our work is also related with a category of information filtering system called recommendation systems. For example, Purushotham et al. [14] designed a generalized hierarchical Bayesian model to make item recommendation by exploiting user's social network information. Yin et al. [25] developed a LCA-LDA probabilistic model to infer both the item content information and the local preference in the recommendation process. Furthermore, Chen et al. [5] proposed a greedy based algorithm that leverages the information coverage to encode the location categories in its recommendations. Our work differs from the above works by developing a new rigorous analytical framework that jointly estimates both the user's travel experience and the interestingness of places without knowing either of them a priori. The framework also considers both the physical and social correlations extracted from the social sensing data.

III. Problem Formulation

In this section, we formulate the social-aware interesting place finding problem as a maximum likelihood estimation problem. Consider a scenario where a group of M users, denoted by $U_1, U_2, ..., U_M$, who visit a set of N places, denoted by $P_1, P_2, .., P_N$. For simplicity, we focus on the binary case on the interestingness of a place (i.e., a place is either interesting or not) [2]. In particular, we let U_i denote

[2]It turns out our solution presented in the next section could also provide a probabilistic metric to evaluate how interesting a place would be.

the i^{th} user and P_k denote the k^{th} place. Futhermore, we let $P_k = I$ denote that place P_k is interesting and $P_k = \bar{I}$ denote that place P_k is not interesting. Additionally, we define a *User-Place Matrix* UP to reflect the visiting behavior of users. In particular, the element $U_iP_k = 1$ when user U_i visits place P_k and $U_iP_k = 0$ otherwise.

In this paper, we explicitly consider the *social dependency* between users in our model. This is motivated by the observation that the visiting behavior of users is highly correlated with their social connections. Simply counting the visits from nonindependent users in the same way as independent users could easily lead to many false positives in the interesting place identification results. Therefore, we define a *User-Dependency Matrix* UD to represent the social dependency between users. In particular, the elements $UD_{ij} = 1$ if user U_i and user U_j are friends and $UD_{ij} = 0$ otherwise. Note the UD is a symmetric matrix as we only consider bi-directional friendship (e.g., friendship on Facebook) in this paper. It is trivial to extend our model and solution to handle directional friendship as well. Using the UD matrix, we can divide the whole set of users into C independent groups where users in the same independent groups have non zero components in UD and users in different independent groups have zero components in UD.

We formulate the social-aware interesting place finding problem as follows. First, we define a few important items that will be used in the problem formulation. If user U_i is an independent user (i.e., U_i has no social connections with other users), we denote the *travel experience* of user U_i by t_i, which is the probability that a place is interesting given that user U_i visits it. If user U_i is a non-independent user (i.e., U_i has social connections with other users), for a friend user U_j of U_i, we denote the *dependent travel experience* of U_i by $t_{i,j}$ where $t_{i,j}$ is the probability that a place is interesting and the friend U_j visits this place given that U_i visits it. Formally, t_i and $t_{i,j}$ are defined as:

$$\begin{aligned} t_i &= \Pr(P_k = I | U_iP_k = 1) \\ t_{i,j} &= \Pr(P_k = I, U_jP_k = 1 | U_iP_k = 1) \end{aligned} \tag{1}$$

For independent users, let us further define T_i to be the probability that user U_i visits the place P_k given that the place is interesting, and let F_i be the probability that user U_i visits the place P_k given that the place is not interesting. For non-independent users, we define $T_{i,j}$ as the probability that user U_i visits the place P_k given that the place is interesting and his/her friend U_j also visits the place. Similarly, we also define $F_{i,j}$ as the probability that user U_i visits the place P_k given that the place is not interesting and his/her friend U_j also visits the place. Formally, T_i, $T_{i,j}$, F_i and $F_{i,j}$ are defined as follows:

$$\begin{aligned} T_i &= \Pr(U_iP_k = 1 | P_k = I) \\ T_{i,j} &= \Pr(U_iP_k = 1 | U_jP_k = 1, P_k = I) \\ F_i &= \Pr(U_iP_k = 1 | P_k = \bar{I}) \\ F_{i,j} &= \Pr(U_iP_k = 1 | U_jP_k = 1, P_k = \bar{I}) \end{aligned} \tag{2}$$

Additionally, we denote the prior probability that user U_i visits a place by s_i (i.e., $s_i = \Pr(U_iP_k = 1)$) and denote d as the prior probability that a randomly chosen place is interesting (i.e., $d = \Pr(P_k = I)$). Based on the Bayes' theorem, we have:

$$\begin{aligned} T_i &= \frac{t_i \times s_i}{d}, \quad F_i = \frac{(1-t_i) \times s_i}{(1-d)} \\ T_{i,j} &= \frac{t_{i,j} \times s_i}{t_j \times s_j}, \quad F_{i,j} = \frac{(1-t_{i,j}) \times s_i}{(1-t_j) \times s_j} \end{aligned} \tag{3}$$

Therefore, the social-aware interesting place finding problem can be formulated as a maximum likelihood estimation (MLE) problem: given the User-Place Matrix UP, the User-Dependency Matrix UD, our goal is to estimate both the *interestingness of each place* and the *travel experience of each user*. Formally, we compute:

$$\begin{aligned} &\forall k, 1 \leq k \leq N : \Pr(P_k = I | UP, UR) \\ &\forall i, 1 \leq i \leq M : \Pr(P_k = I | U_iP_k = 1) \end{aligned} \tag{4}$$

IV. Social-aware Interesting Place Finding

In this section, we solve the interesting place finding problem formulated in Section III by developing a Social-aware Interesting Place Finding (SIPF) scheme.

A. Likelihood Function Development

Expectation Maximization (EM) algorithm is a commonly used optimization technique to find the maximum-likelihood estimates of parameters in a statistical model [7]. To apply EM algorithm to solve a MLE problem, we first need to define a likelihood function $L(\theta; X, Z) = p(X, Z|\theta)$, where θ denotes the parameter vector, X is the observed data, and Z represents the latent variables.

The iterative computation of EM algorithm mainly contains two steps: the expectation step (E-step) and the maximization step (M-step). In particular, E-step estimates the conditional expectation of the latent variables Z and M-step finds the parameters θ that maximize the expectation function in E-step. Formally, they are given as:

$$\text{E-step: } Q(\theta|\theta^{(n)}) = E_{Z|x,\theta^{(n)}}[logL(\theta; x, Z)] \tag{5}$$

$$\text{M-step: } \theta^{(n+1)} = \arg\max_{\theta} Q(\theta|\theta^{(n)}) \tag{6}$$

In the interesting place finding problem, the observed data is the *User-Place* Matrix UP and the *User-Dependency* Matrix UD. The estimation parameter $\theta = (T_1, ..., T_M; F_1, ..., F_M; T_{1,j}, ..., T_{M,j}; F_{1,j}, ..., F_{M,j}; d)$, where T_i, F_i, $T_{i,j}$, $F_{i,j}$ and d are defined in Equation (2). Moreover, we need to define a vector of latent variables Z

to indicate the interestingness of places. Specifically, we have a corresponding variable z_k for each place P_k. Hence, the likelihood function of social-aware interesting place finding problem can be written as:

$$
\begin{aligned}
&L(\theta; X, Z) = \Pr(X, Z|\theta) \\
&= \prod_{k=1}^{N} \Bigg\{ \prod_{g\in C} \Bigg[\prod_{i\in g} (T_i^{U_iP_k}(1-T_i)^{(1-U_iP_k)})^{(|g|==1)} \\
&\prod_{j\in g} ((T_{i,j}^{U_iP_k \ \&\& \ U_jP_k}(1-T_{i,j})^{(1-U_iP_k) \ \&\& \ U_jP_k})^{UD_{i,j}})^{(|g|>1)} \Bigg] \\
&\times d \times z_k + \Bigg[\prod_{i\in g} (F_i^{U_iP_k}(1-F_i)^{(1-U_iP_k)})^{(|g|==1)} \\
&\prod_{j\in g} ((F_{i,j}^{U_iP_k \ \&\& \ U_jP_k}(1-F_{i,j})^{(1-U_iP_k) \ \&\& \ U_jP_k})^{UD_{i,j}})^{|g|>1} \Bigg] \\
&\times (1-d) \times (1-z_k) \Bigg\} \qquad (7)
\end{aligned}
$$

where $U_iP_k = 1$ when user U_i visits the place P_k and 0 otherwise. $UD_{i,j} = 1$ when user U_i is the friend of U_j and 0 otherwise. $|g|$ denotes the size of the independent group g. The "&&" represents the "AND" logic for binary variables. The likelihood function represents the likelihood of the observed data (i.e., UP and UD) and the values of hidden variables (i.e., Z) given the estimation parameters (i.e., θ).

B. Social-aware Interesting Place Finding Scheme

Given the above mathematical formulation, we derive E and M steps of the proposed SIPF scheme. First, we derive the E-step using the likelihood function derived in Equation (7). The E-step is given as follows:

$$
\begin{aligned}
&Q(\theta|\theta^{(n)}) = E_{Z|X,\theta^{(n)}}[logL(\theta; X, Z)] \\
&\sum_{k=1}^{N}\sum_{g\in C}\Bigg\{ Z(n,k) \\
&\times \Bigg[(|g|==1)\cdot \sum_{i\in g}((U_iP_k logT_i + (1-U_iP_k)log(1-T_i)) \\
&+ (|g|>1)\cdot (\sum_{j\in g} UD_{i,j}\cdot (U_iP_k \ \&\& \ U_jP_k) logT_{i,j} \\
&+ (UD_{i,j}\cdot (1-U_iP_k) \ \&\& \ U_jP_k) log(1-T_{i,j})) + logd) \Bigg] \\
&+ (1-Z(n,k)) \\
&\times \Bigg[(|g|==1)\cdot \sum_{i\in g}((U_iP_k logF_i + (1-U_iP_k)log(1-F_i)) \\
&+ (|g|>1)\cdot (\sum_{j\in g} UD_{i,j}\cdot (U_iP_k \ \&\& \ U_jP_k) logF_{i,j} \\
&+ (UD_{i,j}\cdot (1-U_iP_k) \ \&\& \ U_jP_k) log(1-F_{i,j})) \\
&+ log(1-d)) \Bigg] \Bigg\} \qquad (8)
\end{aligned}
$$

where $Z(n,k) = \Pr(z_k = 1|X_k, \theta^{(n)})$. It is the conditional probability of the place P_k to be interesting given the observed data X_k and current estimate of θ, where X_k represents the k^{th} column of the User-Place Matrix UP.

$Z(n,k)$ can be further expressed as:

$$
\begin{aligned}
Z(n,k) &= \frac{\Pr(z_k = 1; X_k, \theta^{(n)})}{\Pr(X_k, \theta^{(n)})} \\
&= \frac{A(n,k)\times d^{(n)}}{A(n,k)\times d^{(n)} + B(n,k)\times (1-d^{(n)})} \qquad (9)
\end{aligned}
$$

where $A(n,k)$ and $B(n,k)$ are defined as follows:

$$
\begin{aligned}
&A(n,k) = \Pr(X_k, \theta^{(t)}|z_k = 1) \\
&\prod_{g\in C}\prod_{i\in g}(T_i^{U_iP_k}(1-T_i)^{(1-U_iP_k)})^{(|g|==1)} \\
&\prod_{j\in g}((T_{i,j}^{U_iP_k \ \&\& \ U_jP_k}(1-T_{i,j})^{(1-U_iP_k) \ \&\& \ U_jP_k)})^{UD_{i,j}})^{(|g|>1)} \qquad (10)
\end{aligned}
$$

$$
\begin{aligned}
&B(n,k) = \Pr(X_k, \theta^{(t)}|z_k = 0) \\
&\prod_{g\in C}\prod_{i\in g}(F_i^{U_iP_k}(1-F_i)^{(1-U_iP_k)})^{(|g|==1)} \\
&\prod_{j\in g}((F_{i,j}^{U_iP_k \ \&\& \ U_jP_k}(1-F_{i,j})^{(1-U_iP_k) \ \&\& \ U_jP_k)})^{UD_{i,j}})^{(|g|>1)} \qquad (11)
\end{aligned}
$$

For the M-step, in order to get the optimal θ^* that maximizes Q function, we set partial derivatives of $Q(\theta|\theta^{(n)})$ given by Equation (8) with respect to θ to 0. In particular, we get the solutions of $\frac{\partial Q}{\partial T_i} = 0$, $\frac{\partial Q}{\partial F_i} = 0$, $\frac{\partial Q}{\partial T_{i,j}} = 0$, $\frac{\partial Q}{\partial F_{i,j}} = 0$ and $\frac{\partial Q}{\partial d} = 0$ in each iteration, we can get expressions of the optimal T_i^*, F_i^*, $T_{i,j}^*$, $F_{i,j}^*$ and d^*:

$$
\begin{aligned}
T_i^{(n+1)} &= T_i^* = \frac{\sum_{k\in UP_i} Z(n,k)}{\sum_{k=1}^{N} Z(n,k)} \\
F_i^{(n+1)} &= F_i^* = \frac{\sum_{k\in UP_i}(1-Z(n,k))}{\sum_{k=1}^{N}(1-Z(n,k))} \\
T_{i,j}^{(n+1)} &= T_{i,j}^* = \frac{\sum_{k\in UP_{i,j}} Z(n,k)}{\sum_{k\in UP_j} Z(n,k)} \\
F_{i,j}^{(n+1)} &= F_{i,j}^* = \frac{\sum_{k\in UP_{i,j}}(1-Z(n,k))}{\sum_{k\in UP_j}(1-Z(n,k)})) \\
d^{(n+1)} &= d^* = \frac{\sum_{k=1}^{N} Z(n,k)}{N} \qquad (12)
\end{aligned}
$$

where UP_i is the set of places that user U_i visits and $UP_{i,j}$ is the set of places both user U_i and U_j visit.

C. Summary of SIPF Scheme

In summary, the input of the social-aware SIPF scheme is the User-Place Matrix UP and User-Dependency Matrix UD obtained from the social sensing data. The output is the maximum likelihood estimation of estimation parameters and latent variables. The estimation results can be used to

Algorithm 1 Social-aware Interesting Place Finding (SIPF) Scheme

Initialize θ ($T_i = s_i, F_i = 0.5 \times s_i$, $T_{i,j} = 0.5$, $F_{i,j} = 0.25$, d =Random number in $(0, 1)$)
$n \leftarrow 0$
repeat
  for Each $k \in P$ **do**
    compute $\Pr(z_k = 1|X_k, \theta^{(n)})$ based on Equation (9)
  end for
  for Each $i \in U$ **do**
    compute $T_i^*, F_i^*, T_{i,j}^*, F_{i,j}^*, d^*$ based on Equation (12)
  end for
  $n = n + 1$
until $\theta^{(n)}$ and $\theta^{(n-1)}$ converge
Let $(Z_k)^c$ = converged value of $\Pr(z_k = 1|X_k, \theta^{(n)})$
for Each $k \in P$ **do**
  if $(Z_k)^c \geq 0.5$ **then**
    consider P_k as Interesting
  else
    consider P_k as Not Interesting
  end if
end for
for Each $i \in U$ **do**
  calculate t_i^*, $t_{i,j}^*$ from converge values of T_i, F_i, $T_{i,j}$, $F_{i,j}$ and d based on Equation (3)
end for
Return the MLE on travel experience t_i^* for U_i and the interestingness judgment for place P_k.

compute both user's travel experience and the interestingness of a place.

V. Evaluation

In this section, we evaluate the performance of the SIPF scheme using two real world datasets collected from location-based social network services. The evaluation results show that SIPF scheme significantly outperforms the state-of-the-art baselines in solving the interesting place finding problem in social sensing.

A. Experiment Settings

1) Dataset Statistics: In this evaluation, we use two different datasets which are collected from location-based social network services, namely, Brightkite[3] and Gowalla[4] [6]. In the location-based social network services, users check in and share their location information using the format as: (user ID, latitude, longitude, timestamp). The basic statistics of the two datasets are shown in Table I.

Table I
DATASET STATISTICS

Description	Brightkite	Gowalla
Number of Users	58,228	107,092
Number of Friendships	214,078	950,327
Number of Check-ins	4,491,143	6,442,890

[3]http://snap.stanford.edu/data/loc-brightkite.html
[4]http://snap.stanford.edu/data/loc-gowalla.html

2) Data Pre-Processing: To evaluate our method in real-world settings, we conducted data pre-processing in two steps: (i) clustering all raw geographical check-in points into meaningful clusters that represent places in the physical world; (ii) identifying independent groups from all users based on their social connections. Using the meta-data generated by the above steps, we can create the User-Place Matrix UP and User-Relationship Matrix UR we discussed in Section III. In our evaluation, we select San Francisco as our target city from the two real-world datasets.

For the clustering step, we used the K-means clutering algorithm to first cluster the raw check-in records into intermediate clusters without any geospatial-semantic meanings allocated to those clusters. Then we re-organized the raw clusters into meaningful places by referring to the Point-of-Interest information from *Google Map*. For Brightkite dataset, we found 83 places in total, of which 36 places are interesting and 47 places are not interesting. For Gowalla trace dataset, we found 92 places in total, of which 39 places are interesting and 53 places are not interesting. As a result, we created the User-Place Matrix UP by associating each user with the places the user visited.

For the independent group identification step, we used a community detection algorithm called SPLA [24] to find independent groups of users. We first obtain the social connections between users from the friendship information in the dataset. In particular, we generated the user dependency graph as a undirected graph $G = (V, E)$ where V and E represents the set of users and their friendship respectively: if the user u is a friend of user v in the dataset, we have a link between u and v. We then applied the SPLA algorithm on the graph G to partition the whole set of users into different independent groups. The users in the same independent group form a clique in graph G. Using the output of this step, we generated the User-Relationship Matrix UR.

B. Baselines and Evaluation Metric

1) Baselines: In the evaluation, we compare the performance of the *SIPF* scheme with the following state-of-the-art baselines from current literature. The first baseline is *Voting*, which computes the interestingness of a place simply by counting the number of times the place is visited. The second baseline is the *Sums and Hubs* [29], which explicitly considers the difference in user's travel experience when it computes the interestingness of a place. The third baseline is *Regular-EM* which is shown to outperform four state-of-the-art techniques in identifying interesting entities from noisy social sensing data [21], [22]. Note that the above baselines all assume that each user is an independent individual and ignore the social dependency between users. In contrast, the SIPF scheme explicitly considers the user dependency under a rigorous analytical framework.

In the evaluation, we compare the performance of the *SIPF* scheme with the following state-of-the-art schemes from current literature. The first baseline is *Voting*, which computes the interestingness of a place simply by counting the number of times the place is visited. The second baseline is the *Sums and Hubs* [29], which explicitly considers the difference in user's travel experience when it computes the interestingness of a place. The third baseline is *Regular-EM* which is proposed to find the truth from unreliable social sensing data. The principle of *Regular-EM* is that source weight is proportional to the probability of the source reporting trustworthy physical observations. We can consider the sources and physical observations in its application as users and clustered places respectively. Then we can apply Regular-EM to identify interesting places in our problem. Note that the above baselines differs from ours is that they assume that all users are independent observers and do not consider the social dependency among them. Such ignorance of source dependency will lead to sub-optimal solutions, which we will demonstrate in the remaining of this section.

2) Evaluation Metric: In the experiments, we use two sets of evaluation metrics. The first set of metrics are used to evaluate the accuracy of different techniques in terms of identifying interesting places. These metrics include *precision*, *recall* and *F1-measure* [13]. The second set of metrics are used to evaluate the ranking performance of different schemes. [5] These metrics are *normalized discounted cumulative gains (NDCG)* [10]. NDCG is an indicator of the average ranking performance of all compared schemes. Given a query, NDCG at position n is calculated as:

$$NDCG(n) = Nr(n) \times \sum_{l=1}^{n} \frac{2^{r(l)} - 1}{log(1+l)} \qquad (13)$$

where $r(l)$ indicates the score for rank l. In our case, $r(l)$ is equal to 1 if the l-th place is interesting and $r(l) = 0$ otherwise. $Nr(n)$ is the normalization factor that guarantees the NDCG of the perfect ranking scheme is equal to 1. Note that NDCG is also averaged over queries at all positions.

C. Evaluation Results

We conducted experiments on two real-world datasets to compare *SIPF* scheme with three state-of-the-art baselines (i.e., *Regular-EM*, *Sums-Hubs* and *Voting*) in terms of *estimation accuracy* and *ranking performance*. Independently from two datasets we used in evaluation, we collected ground truth values (i.e. whether a place is interesting) from three widely used travel recommendation websites: TripAdvisor, Planet Aware and San Francisco Travel. We then decide whether a place is interesting using the following rubric:

- *Interesting places:* Places that have been recommended by at least two of these travel recommendation websites or manually verified by human evaluators.
- *Unconfirmed places:* Places that do not satisfy the requirement of interesting places.

[5]To evaluate the ranking performance, we ranked all places using the estimated interestingness scores of places returned by different schemes.

Note that "unconfirmed places" may include both places that are not interesting or potentially interesting places that cannot be independently verified by using the above rubric. Hence, our evaluation results present *pessimistic* bounds on the performance.

1) Estimation Performance: We first evaluate the estimation performance of all schemes in terms of *precision*, *recall* and *F1-measure*. The results on Brightkite dataset are shown in Figure 1. We observe that the *SIPF* outperforms all the compared baselines in terms of precision, recall and F1-measure. The largest performance gain achieved by SPIF on precision over the best performed baseline (Regular-EM) is 36%. This performance improvement is achieved by explicitly differentiating social dependent and independent users when SIPF computes the estimates on the interestingness of places. The results on Gowalla dataset are shown in Figure 2. We observe similar results: SIPF continues to outperform other baselines and the largest performance gain achieved by SIPF on recall compared to the best performed baseline is 20%.

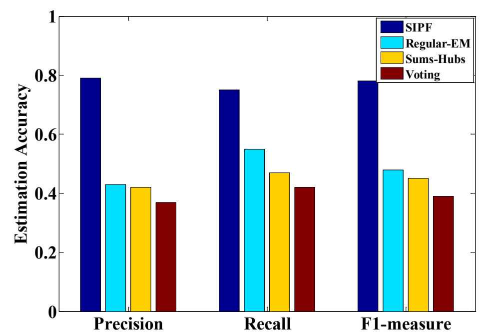

Figure 1. Estimation Accuracy on Brightkite Dataset

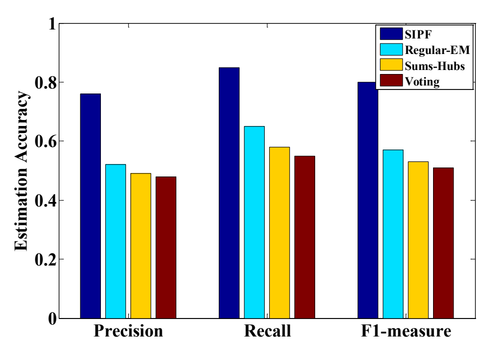

Figure 2. Estimation Accuracy on Gowalla Dataset

2) Ranking Performance: We also evaluate the ranking performance of all schemes and use *NDCG@10,*

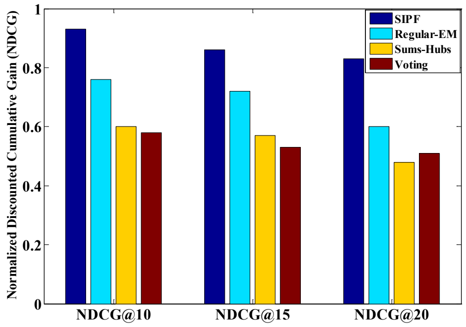

Figure 3. NDCG@n Evaluation on Brightkite Dataset

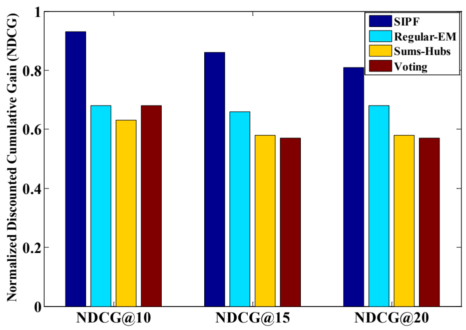

Figure 4. NDCG@n Evaluation on Gowalla Dataset

NDCG@15, NDCG@20 [10] as the evaluation metrics. In Figure 3 and Figure 4, we compare the performance of SIPF to all baselines in terms of NDCG@n on two datasets respectively. We observe that *SIPF* continues to outperform all baselines at different values of n. These results demonstrate that SIFP achieves the best ranking performance among all compared schemes.

VI. Conclusion

This paper develops a new social-aware maximum likelihood estimation framework to accurately identify interesting places in a *smart city application*. The proposed SIPF scheme explicitly incorporates both the user's *travel experience* and *social relationship* into a rigorous analytical framework. The proposed approach jointly estimates both the user's travel experience and the interestingness of a place using an expectation maximization algorithm. We evaluated the SIPF scheme on two real world datasets collected from location-based social network services. The results showed that the SIPF scheme achieved non-trivial performance gains in identifying more interesting places while keeping the falsely reported ones the least compared to the state-of-the-art baselines. The result of the paper is important because it lays out an analytical foundation to improve the interesting place finding accuracy using a principled approach.

Acknowledgment

This material is based upon work supported by the National Science Foundation under Grant No. IIS-1447795.

References

[1] T. Abdelzaher and D. Wang. Analytic challenges in social sensing. In *The Art of Wireless Sensor Networks*, pages 609–638. Springer, 2014.

[2] C. C. Aggarwal and T. Abdelzaher. Social sensing. In *Managing and Mining Sensor Data*, pages 237–297. Springer, 2013.

[3] H. Ahmadi, T. Abdelzaher, J. Han, N. Pham, and R. Ganti. The sparse regression cube: A reliable modeling technique for open cyber-physical systems. In *Proc. 2nd International Conference on Cyber-Physical Systems (ICCPS'11)*, 2011.

[4] I. Boutsis and V. Kalogeraki. Privacy preservation for participatory sensing data. In *IEEE International Conference on Pervasive Computing and Communications (PerCom)*, volume 18, page 22, 2013.

[5] X. Chen, Y. Zeng, G. Cong, S. Qin, Y. Xiang, and Y. Dai. On information coverage for location category based point-of-interest recommendation. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

[6] E. Cho, S. A. Myers, and J. Leskovec. 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*, pages 1082–1090. ACM, 2011.

[7] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, 39(1):1–38, 1977.

[8] Y. Ge, Q. Liu, H. Xiong, A. Tuzhilin, and J. Chen. Cost-aware travel tour recommendation. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 983–991. ACM, 2011.

[9] M. Harding, J. Finney, N. Davies, M. Rouncefield, and J. Hannon. Experiences with a social travel information system. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pages 173–182. ACM, 2013.

[10] K. Järvelin and J. Kekäläinen. Cumulated gain-based evaluation of ir techniques. *ACM Transactions on Information Systems (TOIS)*, 20(4):422–446, 2002.

[11] S. Khetarpaul, R. Chauhan, S. Gupta, L. V. Subramaniam, and U. Nambiar. Mining gps data to determine interesting locations. In *Proceedings of the 8th International Workshop on Information Integration on the Web: in conjunction with WWW 2011*, page 8. ACM, 2011.

[12] A. Majid, L. Chen, H. T. Mirza, I. Hussain, and G. Chen. A system for mining interesting tourist locations and travel sequences from public geo-tagged photos. *Data & Knowledge Engineering*, 95:66–86, 2015.

[13] C. D. Manning, P. Raghavan, H. Schütze, et al. *Introduction to information retrieval*, volume 1. Cambridge university press Cambridge, 2008.

[14] S. Purushotham, Y. Liu, and C.-C. J. Kuo. Collaborative topic regression with social matrix factorization for recommendation systems. *arXiv preprint arXiv:1206.4684*, 2012.

[15] S. Tiwari and S. Kaushik. Mining popular places in a geospatial region based on gps data using semantic information. In *Databases in Networked Information Systems*, pages 262–276. Springer, 2013.

[16] D. Wang, T. Abdelzaher, and L. Kaplan. *Social Sensing: Building Reliable Systems on Unreliable Data*. Morgan Kaufmann, 2015.

[17] D. Wang, T. Abdelzaher, L. Kaplan, R. Ganti, S. Hu, and H. Liu. Exploitation of physical constraints for reliable social sensing. In *The IEEE 34th Real-Time Systems Symposium (RTSS'13)*, 2013.

[18] D. Wang, M. T. Al Amin, T. Abdelzaher, D. Roth, C. R. Voss, L. M. Kaplan, S. Tratz, J. Laoudi, and D. Briesch. Provenance-assisted classification in social networks. *Selected Topics in Signal Processing, IEEE Journal of*, 8(4):624–637, 2014.

[19] D. Wang, M. T. Amin, S. Li, T. Abdelzaher, L. Kaplan, S. Gu, C. Pan, H. Liu, C. C. Aggarwal, R. Ganti, et al. Using humans as sensors: an estimation-theoretic perspective. In *Proceedings of the 13th international symposium on Information processing in sensor networks*, pages 35–46. IEEE Press, 2014.

[20] D. Wang, L. Kaplan, and T. Abdelzaher. Maximum likelihood analysis of conflicting observations in social sensing. *ACM Transactions on Sensor Networks (ToSN)*, Vol. 10, No. 2, Article 30, January, 2014.

[21] D. Wang, L. Kaplan, T. Abdelzaher, and C. C. Aggarwal. On scalability and robustness limitations of real and asymptotic confidence bounds in social sensing. In *The 9th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON 12)*, June 2012.

[22] D. Wang, L. Kaplan, H. Le, and T. Abdelzaher. On truth discovery in social sensing: A maximum likelihood estimation approach. In *The 11th ACM/IEEE Conference on Information Processing in Sensor Networks (IPSN 12)*, April 2012.

[23] D. Wang, K. Qiu, and L.-c. Wang. Design of dba algorithm in epon upstream channel in support of sla. *JOURNAL-CHINA INSTITUTE OF COMMUNICATIONS*, 26(6):87, 2005.

[24] J. Xie, B. K. Szymanski, and X. Liu. Slpa: Uncovering overlapping communities in social networks via a speaker-listener interaction dynamic process. In *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on*, pages 344–349. IEEE, 2011.

[25] H. Yin, Y. Sun, B. Cui, Z. Hu, and L. Chen. Lcars: a location-content-aware recommender system. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 221–229. ACM, 2013.

[26] N. J. Yuan, Y. Zheng, X. Xie, Y. Wang, K. Zheng, and H. Xiong. Discovering urban functional zones using latent activity trajectories. *Knowledge and Data Engineering, IEEE Transactions on*, 27(3):712–725, 2015.

[27] J. Zhang, X. Kong, and P. S. Yu. Transferring heterogeneous links across location-based social networks. In *Proceedings of the 7th ACM international conference on Web search and data mining*, pages 303–312. ACM, 2014.

[28] Y. Zheng and X. Xie. Learning travel recommendations from user-generated gps traces. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(1):2, 2011.

[29] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma. Mining interesting locations and travel sequences from gps trajectories. In *Proceedings of the 18th international conference on World wide web*, pages 791–800. ACM, 2009.