SoCaST: Exploiting Social, Categorical and Spatio-Temporal Preferences for Personalized Event Recommendations

Tunde J. Ogundele Chi-Yin Chow Jia-Dong Zhang
Department of Computer Science, City University of Hong Kong, Hong Kong
Email: jogundele3-c@my.cityu.edu.hk, chiychow@cityu.edu.hk, jiadongzh@gmail.com

***Abstract*—In event-based social networks, an event recommender helps users to discover events that align with their preferences from a large number of upcoming events. In this paper, we propose a personalized event recommender called SoCaST based on the geographical, categorical, social and temporal influences of events on users to provide event recommendations. SoCaST uses an adaptive Kernel Density Estimation (KDE) to model the personalized two-dimensional geographical location. The categorical influence indicates how an event category is relevant to a user and its popularity, while the social influence is modeled as the relevance of a group to a user and her friends. Furthermore, geographical, categorical, and social influences are fused with the temporal influence which is modeled through the KDE method to generate event recommendations. Performance evaluation of SoCaST is conducted by using two large-scale Meetup.com data sets. Experimental results show that SoCaST provides better event recommendations than the state-of-the-art recommendation techniques.**

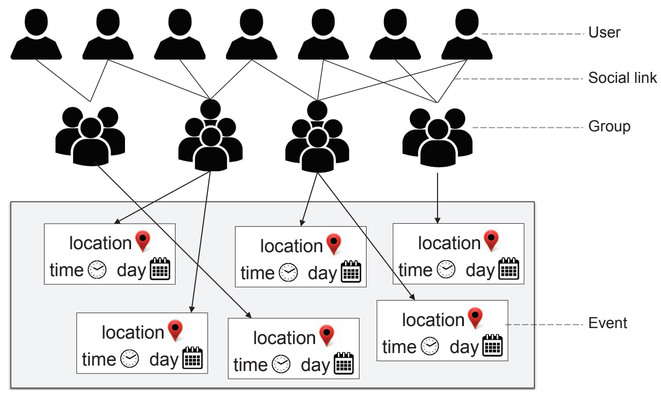


Fig. 1. An event-based social network

I. Introduction

The task of searching for events that best match a user's preference has become a difficult task in event-based social networks (EBSNs) with a large number of available events, e.g., Meetup.com publishes a monthly average of 585,041 events to individuals and groups [1]. In an EBSN, as depicted in Fig. 1, users can join various groups and group organizers can create events (e.g., party, boat trip, and hiking) for their members. One of the most important functions in EBSNs is event recommendation which models a user's preferences based on her attendance history to recommend upcoming events to her.

Most of the existing event recommendation techniques model user preferences based on some of the following influences. (1) **The content influence.** An event can usually be described by keywords (e.g., hiking, bar, and fishing) that characterize it and most time a user always check to see if they match her interests. Studies [2], [3] employed dictionary-based similarity computational approach on event description. However, the category of events is better and easier than keywords for people to search relevant events [4]. (2) **The geographical influence.** The geographical location of an event has a significant influence on attendance behaviors of a user. Techniques such as Kernel Density Estimation (KDE) [3], Gauss formula [2], and K-means [5] have been exploited to estimate the geographical preference of a user. Unfortunately, these studies used one-dimensional geographical distance between locations which does not reflect the geographical attributes associated with locations. (3) **The social influence.** In EBSNs, social relationships are in the form of groups; thus, most existing techniques estimate social influence based on the level of attendance in a group [3], group membership [6], friendship [7], [8], and aggregation of friendship and group membership [9]. (4) **The temporal influence.** Time has a significant influence on user's attendance in EBSNs. Various techniques (e.g., Gauss formula [2] and cosine function [3]) were used to estimate temporal influence for event recommendations on time slices from past events participated by the user.

Real-world motivating examples. Many studies have verified that geographical and temporal influences have significant effects on user preferences in location-based applications (e.g., [10] and [11]). In this work, we focus on the new categorical influence for EBSNs. We use real world examples to illustrate how the categorical influence affects the behaviors of users in EBSNs. We conducted analysis to show the behaviors and preferences of users in relation to their categorical preferences on data sets crawled from Meetup.com. Three users are selected in Meetup.com New York data set in which Users 1, 2, and 3 have attended 387, 217, and 306 events, respectively. Fig 2 depicts the distributions of the three users' category preferences. User 1 likes events in recreation and language categories; User 2 loves to attend events for

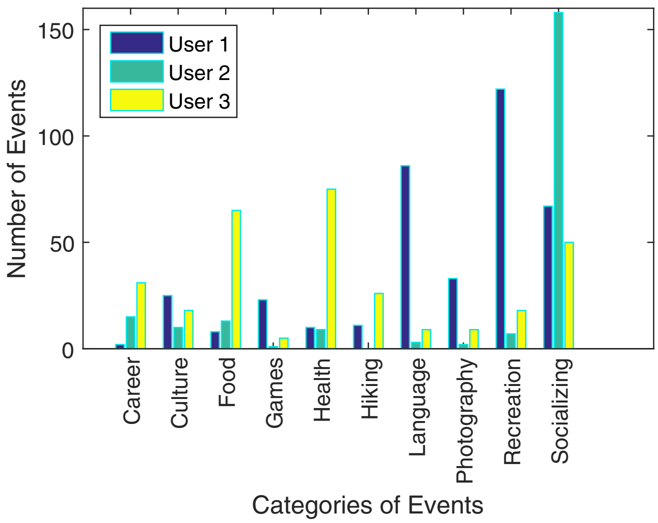


Fig. 2. The categories of events attended by the three selected users in the Meetup.com New York data set

socializing; and User 3 is interested in events about food and health. Therefore, this analysis shows that users in EBSNs have different individual preferences on event categories.

In this paper, we propose an event recommendation framework, called SoCaST, for EBSNs to recommend events for users based on the social, categorical, and spatio-temporal influences. (1) **Geographical influence modeling.** We learn the geographical influence of events by employing an adaptive KDE technique [12], [13] to model the personalized two-dimensional locations (i.e., longitude-latitude coordinates) distributions of each individual user. Our geographical influence model with adaptive KDE bandwidth gives more accurate predictions than the existing methods based on K-mean clustering [5] or one-dimensional distance [2], [3]. (2) **Categorical influence modeling.** Each user's preferences on event categories are represented as a vector consisting of all event categories, and then we adopt the term frequencyinverse document frequency (TF-IDF) technique to estimate a user's preferences on event categories. Our approach is different from the existing studies [2], [3] which focused on the content influence extracted from events' description. (3) **Social influence modeling.** We consider the relevance of the group to a user and other group members to model the social influence. Our method is different from the existing ones which only depend on the user's preference [7] or her friends' preferences [3]. (4) **Temporal influence modeling.** The temporal influence of an event is modeled by employing the KDE technique to estimate the time probability density of a user participating in the event based on the temporal information of past events attended by users.

The main contributions of this paper are summarized as follows:

- The geographical influence model of SoCaST using the adaptive KDE technique for the two-dimensional geographical information of event locations is presented in Section IV-A.
- We use the TF-IDF technique to estimate the categorical influence of events on users based their attendance history. (Section IV-B)
- For the social influence, we consider the relevance of the group to a user and other group members. We model the temporal influence by employing the KDE technique to determine the probability density function to capture the user's preferences on temporal factors. (Sections IV-C and IV-D)
- The geographical, categorical, social and temporal influences are integrated together for personalized event recommendations to the target user, as discussed in Section IV-E.
- Extensive experiments are conducted on two large Meetup.com data sets in New York and San Francisco, USA, to evaluate the performance of SoCaST. Experimental results show that SoCaST performs better than the state-of-the-art event recommendation techniques and the geographical, categorical, social and temporal influences are important user preferences for event recommendations. (Sections V and VI)

The rest of this paper is organized as follows. Section II highlights related work. The system model of SoCaST is defined in Section III. Section IV describes the modeling of geographical, categorical, social, and temporal influences for SoCaST. Performance evaluation and experimental results are explained in Sections V and VI, respectively. Finally, we conclude this paper in Section VII.

II. Related Work

In this section, we highlight related work grouped them into two categories, namely, conventional and event recommender systems.

A. *Conventional Recommender Systems*

Recommender systems analyze users' historical ratings on items to build models that predicts users' ratings on new items and recommend the top items with the highest predicted ratings to users. The geographical, categorical, social and temporal influences are usually considered in recommender systems.

Geographical influences. To exploit geographical influences in location recommendations, collaborative filtering techniques are used to estimate the user's preference on a location [14], [15] and sentiment analysis techniques are also employed to process tips in forms of likes and dislikes which indicate checked-in user's ratings to venues in location-based social networks [16]. With the assumption that the probability of a user visiting a location is a function of its distance, all location check-in history records are transformed into geographical distance distributions that are used to estimate the probability of the user visiting a new location [17]. Furthermore, the KDE method can be used to model the geographical check-in distribution of locations for each user over the latitude and longitude coordinates [18], [19].

Categorical Influences. Information extracted from items' descriptions or users' preferences can be exploited to provide recommendations to users. Conventional content-based filtering techniques use item's attribute to match user's preferences [20], [21] or determine the similarity between different items' attributes [22]. Other techniques that have been exploited include sentiment analysis [23] and TF-IDF [18].

Social influences. With the growth of social networks such as Twitter and Facebook, social links among users have been utilized to improve the quality of recommendations and alleviate the data sparsity and cold-start problems. Friends visit more common locations than other people, and thus friend-based collaborative filtering is used to recommend new locations to a user based on her friends' visited locations [24], [25], [26], [4].

Temporal influences. The role of temporal influence as it affects people's activities and preferences has been widely studied [27], [28], [29], [10]. Every user's checked-in time is split into eight equal-interval periods in a day and employed probabilistic tensor factorization algorithm to model the time influence on the user [30]. Similarity, time is split into hours of a day and stored as an additional dimension to a user-venue matrix [17], i.e., a user-time-venue cube, in which a cell is set to one for a visited venue at a particular time; otherwise, the cell is set to zero; the cube is then fed into a collaborative filtering algorithm. Modeling time drifting patterns in user behavior (i.e., the popularity of an item and the baseline ratings of a user change over time) was considered to improve the quality of recommendations [31].

B. Event Recommender Systems in EBSNs

In this section, we distinguish SoCaST from existing event recommendation techniques proposed for EBSNs.

Geographical influences. To model the geographical influence of an event, Gauss formula is used to compute the similarity between the distance of an event location and a user's attendance history [2]. Similarly, the distance of the target event on the Gaussian distribution of the previously attended events is modeled through the KDE method [3]. With the assumption of knowing users' current locations, the distance between the target user and other users that have attended events together is used to model the geographical location influence [6]. In contrast, K-means and Gaussian distributions are used to compute the probability of the location of an event in a region based on a user's historical attended event locations to predict her regional preference on an event location [5]. In SoCaST, the geographical influence of events on a user is modeled by using the adaptive KDE method over two-dimensional latitude and longitude coordinates of event locations in ESBNs.

Categorical influences. In EBSNs, an event is associated with description and category types that help users decide whether to attend or not. Content-based influences have been widely explored, e.g., content-based similarity method [32], TF-IDF technique [3], and latent Dirichlet allocation (LDA) technique [2] are used to compute the content influence. In this work, we consider categories of events and determine how an event category is relevant to a user and her friends in a group based on their historical attendance records.

Social influences. The social relationships or links are in the form of groups in EBSNs, i.e., a member has a social link with each of other members in a group. Thus, the simplest way of modeling the social influence of events on users is based on the group membership [6], [33]. In contrast, friendships without considering group membership were be employed [7], [8], [32]. Also, social influences can be modeled through both friendships and group membership [9], or participation records in a group [3]. The methods considering friendships and attendance records usually result in low-quality event recommendations due to the cold-start user problem (i.e., new members in a group have not had any strong relationship with other group members or attendance record yet). In SoCaST, we model the social influence of events in a group based on the relevance of the group to a user and its members.

Temporal influences. The time of an event has a great impact on a user decision whether to attend the event. Gauss formula is employed to model temporal influences based on the hour of the day and the day of the week of events [2]. The event time can also be split into 24×7 matrix and the cosine function is employed to compute the temporal influence [3]. In SoCaST, the time probability density of an event is estimated based on the temporal intervals of past events participated by users through the KDE method.

III. System Model

The key terms and problem of SoCaST are defined in this section. Table I lists key notations used in this paper.

Definition 1: Event. An event $e \in E$ is defined as a tuple of $e = (e.t, e.l, e.c)$ where $e.t$, $e.l$, and $e.c$ indicate the time, location, and category of e, respectively, and $U(e)$ is a set of users committed to join e.

Definition 2: Group. A group $g \in G$ has its members $U(g)$ and events $E(g)$.

Definition 3: Social Link. Given two users $u_i, u_j \in U$, $u_i \neq u_j$, a social link (or friendship) is established between u_i and u_j (i.e., $link(u_i, u_j) = 1$), if both u_i and u_j are in a common group (i.e., $u_i, u_j \in g$ and $g \in G$); otherwise, $link(u_i, u_j) = 0$, i.e., u_i and u_j do not have any common group.

SoCaST is designed to recommend events to users in an EBSN, given a set of users $U = \{u_1, u_2, \ldots, u_{|U|}\}$, a set of groups $G = \{g_1, g_2, \ldots, g_{|G|}\}$, and a set of events $E = \{e_1, e_2, \ldots, e_{|E|}\}$, where $|U|$, $|G|$, and $|E|$ are the total number of users, groups, and events, respectively. Each user u_i can join multiple groups $G(u_i)$ (i.e., $u_i \in U$ and $G(u_i) \subseteq G$) as their members, each group g_j can create events $E(g_j)$ (i.e., $g_j \in G$ and $E(g_j) \subseteq E$) for its members $U(g_j)$ (i.e., $U(g_j) \subseteq U$) to attend.

Problem definition. Given a user u in group g, the goal is to determine the relevance scores of geographical, categorical, social, and temporal influences of events in $E(g)$ based on u's

TABLE I
NOTATIONS AND THEIR MEANING

Symbols	Meaning
E	Set of events $E = \{e_1, e_2, \ldots, e_{\|E\|}\}$
U	Set of users $U = \{u_1, u_2, \ldots, u_{\|U\|}\}$
G	Set of groups $G = \{g_1, g_2, \ldots, g_{\|G\|}\}$
C	Set of categories of events $C = \{c_1, c_2, \ldots, c_{\|C\|}\}$
$E(g)$, $U(g)$	Sets of events and users of group g
$G(u)$	Set of groups joined by user u
L	Set of past event locations
$L(u)$	Set of past event locations visited by u
T	Set of time intervals $T = \{00:00, 00:01, \ldots, 23:59\}$
$D(u)$	Set of time sample of events participated by user u
$link(u_i, u_j)$	if u_i and u_j are in a common group (i.e., have a social link), $link(u_i, u_j) = 1$; otherwise, $link(u_i, u_j) = 0$
$M_L(\|U\|, \|L\|)$	Visit matrix: $m_l(u, l)$ is the frequency of user u visiting location l to attend events
$M_P(\|U\|, \|G\|)$	Group participation matrix: $m_p(u, g)$ is the number of g's events attended by user u
$M_T(\|U\|, \|T\|)$	Time matrix: $m_t(u, t)$ is the number of events scheduled at $t \in T$ and attended by u
$M_C(\|U\|, \|C\|)$	Category preference matrix: $m_c(u, c)$ is the number of c's events participated by user u

preferences, and recommend the top-k events with the highest relevance scores to u.

IV. Modeling of Preferences for Recommendations

In this section, we present how a user's preferences are modeled with respect to the geographical, categorical, social, and temporal influences (Sections IV-A to IV-D), and integrate them to provide event recommendations for the user (Section IV-E).

A. Geographical Influence Modeling

To model the geographical influence of events, we use the adaptive KDE method to model the personalized two-dimensional location distribution of a user. We compute the geographical influence using the adaptive kernel density estimation technique in three steps which include pilot estimation, local bandwidth determination, and adaptive kernel estimation [4].

Step 1: Pilot estimation. In our SoCaST, the pilot estimation is computed based on the densities from KDE with fixed bandwidth. As we consider an event's location as a pair of latitude and longitude (lat, lon), the KDE function for $L(u)$ (i.e., the set of locations visited by u to attend events) of user u is defined as:

$$\hat{f}(l) = \frac{1}{N\sigma^2} \sum_{l_i \in L(u)} m_l(u, l_i) \cdot K\left(\frac{l - l_i}{\sigma}\right), \qquad (1)$$

where $l_i = (lat_i, lon_i)^T$ is a two-dimensional location vector, σ is the bandwidth, N is the number of locations in the samples, and $K(.)$ is the Gaussian kernel function [12] that is given as:

$$K(x) = \frac{1}{2\pi} e^{-\frac{1}{2}x^T x}. \qquad (2)$$

Algorithm 1 The computation of $f_A(l|L(u))$ using the adaptive KDE method.

Input: A set of locations visited by user u to attend events: $L(u) = \{l_1, l_2, \cdots, l_{|L(u)|}\}$.
Output: For each upcoming event, the probability (i.e., $f_A(l|L(u))$) of u visiting l to attend it.

$\hat{f}(l) \leftarrow 0$ // Initialize estimates $\hat{f}(l)$
Compute a pilot estimation $\hat{f}(l)$ that satisfies $\hat{f}(l_i) > 0$ as given in Equation 1
for each upcoming event at location l that u can attend **do**
  Compute the geometric mean $\hat{g}$ by Equation 5
  Compute the local bandwidth weight w_i by Equation 4
  for each $l_i \in L(u)$ **do**
    $z \leftarrow z + \frac{m_l(u,l_i)}{w_i} \cdot K(\frac{l-l_i}{w_i \sigma})$
  end for
  $f_A(l|L(u)) \leftarrow \frac{z}{N\sigma^2}$
end for

The bandwidth for the Gaussian kernel is defined as:

$$\sigma = N^{-\frac{1}{6}} \sqrt{\frac{\hat{\sigma}^T \hat{\sigma}}{2}}, \qquad (3)$$

where $\hat{\sigma}$ is the sample standard deviation.

Step 2: Local bandwidth determination. This step combines the features of nearest neighbour and kernel approaches [34] by increasing or decreasing the bandwidth in an area where the data density is low or high, respectively [12]. This is because the main idea of the adaptive KDE technique is to remove noise in an area where the distribution of data is sparse, but recover details in another area where data are dense. We first determine the initial estimation to derive a series of bandwidths that correspond to the observations; this estimation gives an overview of the density [34], and then construct the adaptive estimation from the bandwidths. The local bandwidth weight is inversely proportional to the initial density estimation and can be computed by:

$$w_i = \left\{ \frac{\hat{g}}{\hat{f}(x_i)} \right\}^{\frac{1}{2}}, \qquad (4)$$

where $\hat{g}$ is the geometric mean of the observations $\hat{f}(x_i)$ that is given as:

$$\hat{g} = \left\{ \prod_{i=1}^{n} \hat{f}(x_i) \right\}^{\frac{1}{n}}. \qquad (5)$$

Step 3: Adaptive kernel estimation. With the local bandwidth w_i computed by Equation 4, this step computes the probability of user u attending an event at location l as below:

$$f_A(l|L(u)) = \frac{1}{N\sigma^2} \sum_{l_i \in L(u)} \frac{m_l(u, l_i)}{w_i} \cdot K(\frac{l - l_i}{w_i \sigma}). \qquad (6)$$

Algorithm 1 outlines the process for estimating $f_A(l|L(u))$ through Equation 6. The local bandwidth weight is dependent on the pilot estimation that is computed once by using

Equation 1 for all n upcoming events. After that, for each upcoming event at location l that user u can attend (Lines 3 to 10 in Algorithm 1), we compute local bandwidth w_i by using Equations 4 and 5 (Lines 4 and 5), and then compute the probability of u visiting location l to attend the corresponding event based on Equation 6 (Line 9).

As a result, the geographical influence of event e at location $e.l$ on user u is computed by:

$$GI(e,u) = f_A(e.l|L(u)). \tag{7}$$

B. Categorical Influence Modeling

The category of an event plays a significance factor for a user to decide whether to attend an event in EBSNs. For example, in Meetup.com, each event is associated with at least one category. The frequency of an event's category attended by a user indicates her preferences on event categories. The categorical influence involves the user's preferences on the event category and category popularity.

Category relevance to a user. Let $C(u)$ be a set of categories of events attended by user u. The relevance of the category of event e to u is given as:

$$\hat{c}(e.c,u) = \text{tf-idf}(u, e.c), \tag{8}$$

where tf-idf$(u, e.c)$ is the term frequency-inverse document frequency weight [35] of user u for event's category $e.c$, that is computed by:

$$\text{tf-idf}(u,c) = \frac{m_c(u,c)}{|E(u)|} \cdot \log\frac{|E|}{|E(c)|}, \tag{9}$$

where $m_c(u,c)$ is the number of c's events attended by u, $|E(u)|$ is the total number of events attended by u, and $E(c)$ is the set of events associated with category c.

Category popularity. In EBSNs, a user joins other members of a group to participate in an event. The perception of the members about the event's category is important and thus our categorical influence model takes the category popularity into consideration. Given group g for event e and a set of users of group g (i.e., $U(g)$), the category popularity of c in group g is given by:

$$\hat{c}_p(e.c,g) = \frac{\sum_{u_i \in U(g)} \text{tf-idf}(u_i, e.c)}{\max_{c_j \in C}\left\{\sum_{u_i \in U(g)} \text{tf-idf}(u_i, e.c_j)\right\}}. \tag{10}$$

Finally, the categorical influence of event e associated with category $e.c$ on u in group u is estimated as:

$$CI(u,c,g)) = [\hat{c}(e.c,u) + \hat{c}_p(e.c,g)]/2. \tag{11}$$

C. Social Influence Modeling

Each user can join at least one group and participate in events created and published by their corresponding groups. This means that when a user joins more than one group, she can have some preferred groups, namely, the relevance of a group to a user. In addition, as all the members in a group are considered as friends, the relevance of a group to a user's friend is also involved in the social influence model.

Group relevance to a user. Let $G(u)$ be a set of groups that user u joins for event participation. The group relevance of u is given as:

$$\hat{s}(u,g) = \frac{m_p(u,g)}{\max_{g_i \in G(u)}\{m_p(u,g_i)\}}, \tag{12}$$

where $m_p(u,g)$ is the number of g's events participated by u.

Social group relevance. To compute social group relevance, the similarity of the group relevance of user u's friends in group g (Definition 3) is estimated as follows:

$$s_g(u,g) = \frac{\sum_{u_i' \in U(g) \wedge u_i' \neq u} sim(u,u_i') \cdot m_p(u_i',g)}{\sum_{u_i' \in U(g) \wedge u_i' \neq u} sim(u,u_i')}, \tag{13}$$

where $sim(u_i,u_j) = \frac{\sum\limits_{g_k \in G(u_i)} m_p(u_i,g_k)\cdot m_p(u_j,g_k)}{\sqrt{\sum\limits_{g_k \in G(u_i)} m_p(u_i,g_k)^2}\cdot\sqrt{\sum\limits_{g_k \in G(u_i)} m_p(u_j,g_k)^2}}$ calculates the similarity between users u_i and u_j. Then, $s(u,g)$ is normalized as:

$$\hat{s}_g(u,g) = \frac{s(u,g)}{\max_{g_i \in G(u)}\{s(u,g_i)\}}. \tag{14}$$

Finally, the social influence of group g on user u is determined as:

$$SI(u,g) = [\hat{s}(u,g) + \hat{s}_g(u,g)]/2. \tag{15}$$

D. Temporal Influence Modeling

In SoCaST, rather than breaking down the continuous time into discrete time slices which can result into time information loss, temporal influence on a user can be modeled by employing the non-parametric KDE technique to estimate the time probability density of an event.

Let $D(u)$ be the time sample drawn from distribution with an unknown probability density $f(t|D(u))$ that user u attended. The $f(t|D(u))$ is given by:

$$f(t|D(u)) = \frac{1}{|D(u)|\sigma}\sum_{t_i \in T(u)} m_t(u,t_i) \cdot K(\frac{t \ominus t_i}{\sigma}), \tag{16}$$

where σ is the bandwidth, $K(.)$ is the kernel function given in Equation 2 and satisfy the condition:

$$\forall x, K(x) \geq 0 \text{ and } \int_{+\infty}^{-\infty} K(x)dx = 1. \tag{17}$$

$t \ominus t_i$ is the time difference between two instances of time (i.e., t and t_i). We adopt the time difference technique of [11] and it is given by:

$$t \ominus t_i = \begin{cases} |t - t_i|, & |t - ti| \leq 12:00; \\ 24:00 - |t - t_i|, & |t - ti| > 12:00. \end{cases} \tag{18}$$

Thus, the temporal influence is determined as:

$$TI(e,u) = \int_{t \in T} f(t|T(u))dt. \tag{19}$$

E. Event Recommendations

In SoCaST, we combine the geographical, categorical, social and temporal influences into unified preference score $s_p(e,u,g)$ for user u with respect to event e using the product rule:

$$s_p(e,u,g) = GI(e,u) \cdot CI(u,c) \cdot SI(u,g) \cdot TI(e,u). \quad (20)$$

For each upcoming event e in group g that can be attended by user u, we compute its preference score (i.e., $s_p(e,u,g)$. The top-k events with highest preference scores are recommended to u.

V. Experiment Settings

In this section, we describe the data sets, baseline techniques, and performance metrics.

Data sets. All experiments are conducted on Meetup.com [1] event data sets in New York (NY) and San Francisco (SF), USA. Table II gives the statistical summary of the data sets. We divide each of the data sets into training and testing sets based on timestamps, i.e., 60% event data with earliest timestamps are considered as the training set and the remaining event data are constituted as the testing set.

TABLE II
THE STATISTICAL SUMMARY OF THE DATA SETS

	Members	Events	Groups	Venue	RSVP
SF	324,485	169,120	11,367	18,126	1,665,591
NY	910,938	432,938	13,906	35,623	3,614,720

Evaluated event recommendation techniques. Our proposed SoCaST is compared with the state-of-the-art event recommendation techniques as listed below:

- SRE: This technique exploits social influences of users (i.e., friends) in EBSNs for event recommendations [8].
- CFM: CFM exploits the collaborative filtering technique and dynamics on temporal influences for users [31].
- CAER: This technique considers event description, time, social group, and geographical location to rank and recommend events to users [3].
- PAAT: This method uses singular value decomposition with a multi-factor neighborhood to predict events attendance for members by exploiting time, distance and event description [2].

Performance metrics. In event recommendations, it is important to find the number of recommended events that a target user actually visited in her testing data set. In this regard, the standard metrics, namely, precision and recall [19], are used to measure the quality of recommendations of SoCaST. Precision is the ratio of the number of events that a user attended to the total number of recommended top-k events, i.e.,

$$\text{Precision} = \frac{\text{No. of the attended top-}k\text{ events}}{\text{No. of the top-}k\text{ events}},$$

and recall is the ratio of the number of recommended top-k events that user u attended to the total number of events attended by u, i.e.,

$$\text{Recall} = \frac{\text{No. of the top-}k\text{ events attended by }u}{\text{No. of events attended by }u}.$$

Furthermore, the number of recommended events (i.e., top-k) varies from 2 to 50 for regular users, but 1 to 10 for cold-start users due to the small total number of events attended by them.

VI. Experimental Results

This section analyzes the experimental results and evaluates the performance of our SoCaST by comparing its quality of event recommendations with the baselines in Sections VI-A and VI-B). The contributions of the geographical, categorical, social, and temporal influences of SoCaST are studied in Section VI-C.

A. Comparison of Event Recommendation Techniques

Fig. 3 depicts the quality of event recommendations in the NY and SF data sets. SoCaST gives the best performance in terms of both the precision and recall for all values of k.

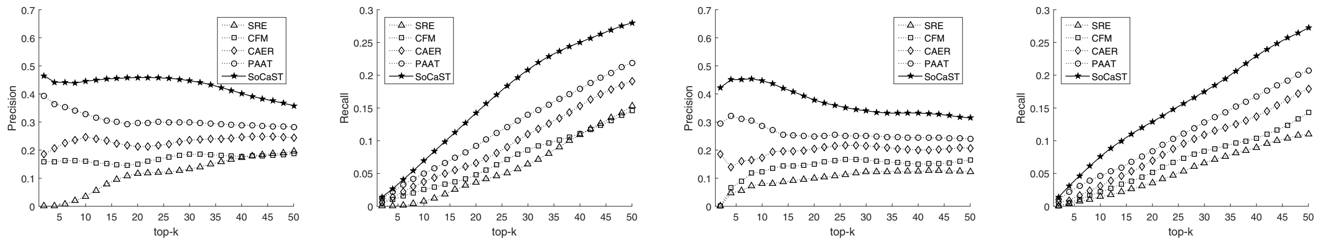
SRE. Since SRE exploits only social influence, it gives the worst performance.

CFM. CFM considers the collaborative social and temporal influences; and thus it performs better than SRE. However, its performance is worse than CAER, PAAT and SoCaST because it ignores the geographical influence of events in EBSNs.

CAER. CAER performs better than SRE and CFM. CAER computes the geographical influence by modeling the distance between the pair of home and event locations and integrates it with social and temporal influences. However, its performance is worse than SoCaST.

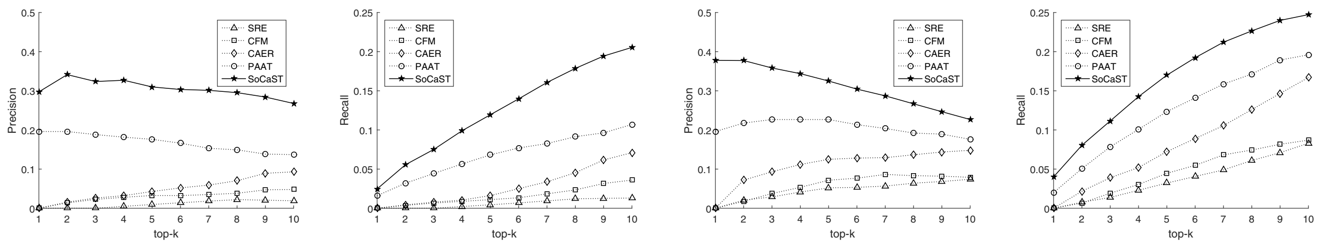
PAAT. PAAT uses distance similarity to estimate the geographical influence of an event and aggregates it with content, social and temporal influences for recommendations. Although it performs better than other baselines, its recommendation quality is still worse than SoCaST.

SoCaST. Our SoCaST achieves the best performance in terms of both the precision and recall. The reason is that SoCaST has more sophisticated geographical, categorical, social, and temporal influence models: (1) SoCaST models the geographical influence by exploiting the personalized two-dimensional geographical locations through the adaptive KDE technique. (2) SoCaST models the categorical influence based on the relevance of an event category to a user and her friends. (3) SoCaST models the social influence based on the relevance of a group to a user and her friends. (4) For the temporal influence, SoCaST further considers the time probability density function of events for a target user.



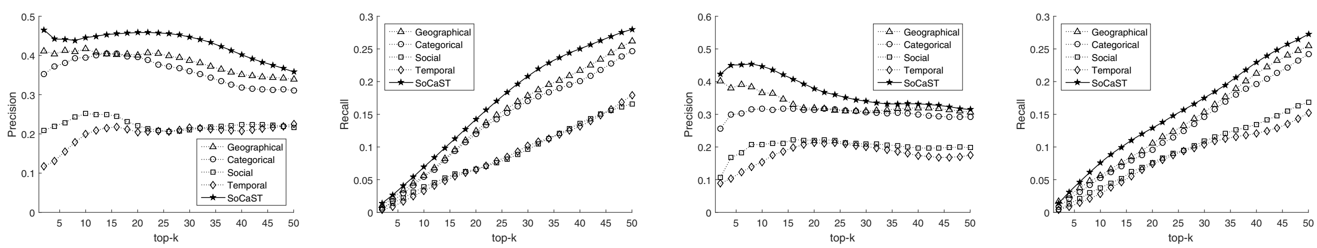
(a) New York (Precision) (b) New York (Recall) (c) San Francisco (Precision) (d) San Francisco (Recall)

Fig. 3. Performance of event recommendation methods



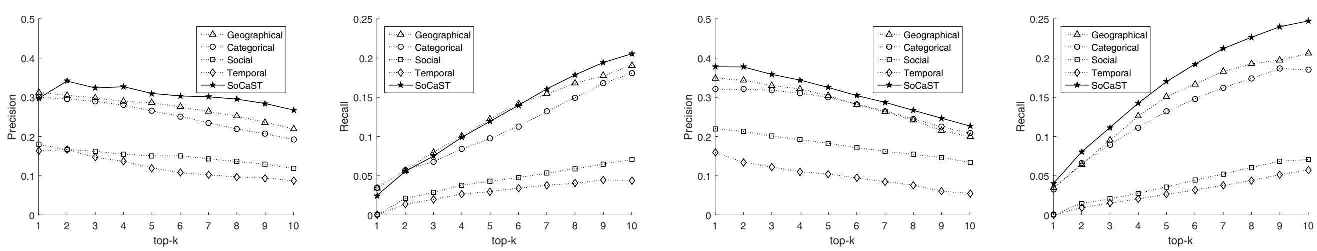
(a) New York (Precision) (b) New York (Recall) (c) San Francisco (Precision) (d) San Francisco (Recall)

Fig. 4. Performance of event recommendation methods for cold-start users



(a) New York (Precision) (b) New York (Recall) (c) San Francisco (Precision) (d) San Francisco (Recall)

Fig. 5. Performance of the geographical, social, and temporal influences of SoCaST



(a) New York (Precision) (b) New York (Recall) (c) San Francisco (Precision) (d) San Francisco (Recall)

Fig. 6. Performance of the geographical, social, and temporal influences of SoCaST for cold-start users

B. Comparison of Event Recommendation Techniques for Cold-Start Users

Users with only a few attendance records are considered as the cold-start users and the performance of SoCaST for cold-start users is depicted in Fig. 4. Experimental results show that SoCaST also performs better than all the baselines in terms of both the precision and recall. Therefore, we can conclude that SoCaST provides higher quality of event recommendations than the evaluated techniques for both regular users (Fig. 3) and cold-start users.

C. Contribution of the Geographical, Categorical, Social and Temporal Influences in SoCaST

The contribution of geographical, categorical, social and temporal influences of SoCaST based on Equations 7, 11, 15, and 19, respectively, for regular users (Fig. 5) and cold-start users (Fig. 6) is studied. The experimental results show that (1) each influence plays a significant role in SoCaST for event recommendations, (2) the geographical and categorical influences are more important than the other two influences, and (3) the social influence is more important than the temporal influence. The results give us an invaluable insight that the influences do not have the same importance; and thus, one of our research directions is to design an adaptive fusion method for the influences.

VII. Conclusion

In this paper, we proposed the event recommendation framework called SoCaST that recommends events for users in EBSNs based on the geographical, categorical, social and temporal influences of events on users. SoCaST employs the adaptive Kernel Density Estimation (KDE) on the two-dimensional geographical locations of events to compute the geographical influence of events to a user. The categorical influence model considers the relevance of an event category to a user and the category popularity, while we estimate the social influence of an event on a user in a group based on the relevance of the group to the user and her friends. We develop the time probability density function to model the temporal influence of events on users. Two large-scale real data sets in New York and San Francisco crawled from Meetup.com are used to evaluate the performance of SoCaST. Experimental results show that SoCaST performs better than the state-of-the-art event recommendation techniques in terms of both precision and recall for both regular and cold-start users. Our future work will investigate how to find out personalized weights for different influences.

References

[1] Meetup.com, *About Meetup*, 2017 (accessed on February 12, 2017). [Online]. Available: http://www.meetup.com/about/

[2] R. Du, Z. Yu, T. Mei, Z. Wang, Z. Wang, and B. Guo, "Predicting activity attendance in event-based social networks," in *Proceedings of ACM UbiComp*, 2014.

[3] A. Q. Macedo, L. B. Marinho, and R. L. Santos, "Context-aware event recommendation in event-based social networks," in *Proceedings of ACM RecSys*, 2015.

[4] J.-D. Zhang and C.-Y. Chow, "GeoSoCa: Exploiting geographical, social and categorical correlations for point-of-interest recommendations," in *Proceedings of ACM SIGIR*, 2015.

[5] Z. Qiao, P. Zhang, Y. Cao, C. Zhou, L. Guo, and B. Fang, "Combining heterogenous social and geographical information for event recommendation," in *Proceedings of AAAI*, 2014.

[6] E. M. Daly and W. Geyer, "Effective event discovery: Using location and social information for scoping event recommendations," in *Proceedings of ACM RecSys*, 2011.

[7] Y. Zhang, H. Wu, V. Sorathia, and V. K. Prasanna, "Event recommendation in social networks with linked data enablement," in *Proceedings of EIS*, 2013.

[8] T. D. Pessemier, J. Minnaert, K. Vanhecke, S. Dooms, and L. Martens, "Social recommendations for events," in *Proceedings of CEUR*, 2013.

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

[10] J.-D. Zhang and C.-Y. Chow, "Spatiotemporal sequential influence modeling for location recommendations: A gravity-based approach," *ACM Transactions on Intelligent Systems and Technology*, vol. 7, no. 1, pp. 11:1–11:25, 2015.

[11] ——, "TICRec: A probabilistic framework to utilize temporal influence correlations for time-aware location recommendations," *IEEE Transactions on Services Computing*, vol. 9, no. 4, pp. 633–646, 2016.

[12] B. Silverman, *Density estimation for statistics and data analysis*. London: Chapman and Hall, 1986.

[13] J. Simonoff, *Smoothing Methods in Statistics*. Springer, 1996.

[14] J. J. Levandoski, M. Sarwat, A. Eldawy, and M. F. Mokbel, "LARS: A location-aware recommender system," in *Proceedings of IEEE ICDE*, 2012.

[15] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee, "Exploiting geographical influence for collaborative point-of-interest recommendation," in *Proceedings of ACM SIGIR*, 2011.

[16] D. Yang, D. Zhang, Z. Yu, and Z. Yu, "Fine-grained preference-aware location search leveraging crowdsourced digital footprints from LBSNs," in *Proceedings of ACM UbiComp*, 2013.

[17] Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. M. Thalmann, "Time-aware point-of-interest recommendation," in *Proceedings of ACM SIGIR*, 2013.

[18] Y. Lyu, C.-Y. Chow, R. Wang, and V. C. S. Lee, "Using multi-criteria decision making for personalized point-of-interest recommendations," in *Proceedings of ACM SIGSPATIAL*, 2014.

[19] J.-D. Zhang and C.-Y. Chow, "CoRe: Exploiting the personalized influence of two-dimensional geographic coordinates for location recommendations," *Information Sciences*, vol. 293, pp. 163 – 181, 2015.

[20] J. Dez, J. J. del Coz, O. Luaces, and A. Bahamonde, "Clustering people according to their preference criteria," *Expert Systems with Applications*, vol. 34, no. 2, pp. 1274 – 1284, 2008.

[21] S. Jiang, X. Qian, J. Shen, Y. Fu, and T. Mei, "Author topic model-based collaborative filtering for personalized POI recommendations," *IEEE Transactions on Multimedia*, vol. 17, no. 6, pp. 907–918, 2015.

[22] R. Dong, M. P. O'Mahony, M. Schaal, K. McCarthy, and B. Smyth, "Combining similarity and sentiment in opinion mining for product recommendation," *Journal of Intelligent Information Systems*, vol. 46, no. 2, pp. 285–312, 2016.

[23] A. Yessenalina, Y. Yue, and C. Cardie, "Multi-level structured models for document-level sentiment classification," in *Proceedings of EMNLP*, 2010.

[24] J. Bao, Y. Zheng, and M. F. Mokbel, "Location-based and preference-aware recommendation using sparse geo-social networking data," in *Proceedings of ACM SIGSPATIAL*, 2012.

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

[26] H. Wang, M. Terrovitis, and N. Mamoulis, "Location recommendation in location-based social networks using user check-in data," in *Proceedings of ACM SIGSPATIAL*, 2013.

[27] M. Gruteser and D. Grunwald, "Anonymous usage of location-based services through spatial and temporal cloaking," in *Proceedings of ACM MobiSys*, 2003.

[28] K. Zhang, Q. Jin, K. Pelechrinis, and T. Lappas, "On the importance of temporal dynamics in modeling urban activity," in *Proceedings of ACM SIGKDD UrbComp Workshop*, 2013.

[29] S. Bannur and O. Alonso, "Analyzing temporal characteristics of check-in data," in *Proceedings of WWW*, 2014.

[30] Y.-L. Zhao, L. Nie, X. Wang, and T.-S. Chua, "Personalized recommendations of locally interesting venues to tourists via cross-region community matching," *ACM Transactions on Intelligent Systems and Technology*, vol. 5, no. 3, pp. 50:1–50:26, 2014.

[31] Y. Koren, "Collaborative filtering with temporal dynamics," in *Proceedings of ACM SIGKDD*, 2009.

[32] M. Kayaalp, T. Özyer, and S. T. Özyer, "A collaborative and content based event recommendation system integrated with data collection scrapers and services at a social networking site," in *Proceedings of ASONAM*, 2009.

[33] X. Liu, Q. He, Y. Tian, W.-C. Lee, J. Mcpherson, and J. Han, "Event-based social networks : Linking the online and offline social worlds," in *Proceedings of ACM SIGKDD*, 2012.

[34] I. H. Salgado-Ugarte, M. A. Perez-Hernandez *et al.*, "Exploring the use of variable bandwidth kernel density estimators," *Stata Journal*, vol. 3, no. 2, pp. 133–147, 2003.

[35] H. C. Wu, R. W. P. Luk, K. F. Wong, and K. L. Kwok, "Interpreting tf-idf term weights as making relevance decisions," *ACM Transactions on Information Systems*, vol. 26, no. 3, pp. 13:1–13:37, 2008.