SoCaST*: Personalized Event Recommendations for Event-based Social Networks
A Multi-Criteria Decision Making Approach
OGUNDELE, Tunde J.; CHOW, Chi-Yin; ZHANG, Jia-Dong

Published in:
IEEE Access

Published: 01/01/2018

Document Version:
Final Published version, also known as Publisher's PDF, Publisher's Final version or Version of Record

Publication record in CityU Scholars:
Go to record

Published version (DOI):
10.1109/ACCESS.2018.2832543

Publication details:

Citing this paper
Please note that where the full-text provided on CityU Scholars is the Post-print version (also known as Accepted Author Manuscript, Peer-reviewed or Author Final version), it may differ from the Final Published version. When citing, ensure that you check and use the publisher's definitive version for pagination and other details.

General rights
Copyright for the publications made accessible via the CityU Scholars portal is retained by the author(s) and/or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights. Users may not further distribute the material or use it for any profit-making activity or commercial gain.

Publisher permission
Permission for previously published items are in accordance with publisher's copyright policies sourced from the SHERPA RoMEO database. Links to full text versions (either Published or Post-print) are only available if corresponding publishers allow open access.

Take down policy
Contact lbscholars@cityu.edu.hk if you believe that this document breaches copyright and provide us with details. We will remove access to the work immediately and investigate your claim.

Download date: 19/12/2022
SoCaST*: Personalized Event Recommendations for Event-Based Social Networks: A Multi-Criteria Decision Making Approach

TUNDE JOSEPH OGUNDELE, (Student Member, IEEE), CHI-YIN CHOW, (Senior Member, IEEE), AND JIA-DONG ZHANG, (Member, IEEE)
Department of Computer Science, City University of Hong Kong, Hong Kong
Corresponding author: Chi-Yin Chow (chiychow@cityu.edu.hk)

This work was supported in part by the National Natural Science Foundation of China under Grant 61772445 and in part by the City University of Hong Kong under Grant 7004684, Grant 7004890, Grant 9678132, and Grant 9680221.

ABSTRACT In event-based social networks, user preferences mined from the influences of geographical locations, event categories, and social and temporal preferences have been exploited for event recommendations by assuming that each of these influences has the same weight for all users. However, in the reality, a user would have different degrees of importance for these influences on deciding whether to participate in an event. In this paper, we propose a personalized event recommendation framework called SoCaST*, which employs the multi-criteria decision making approach to rank events. In SoCaST*, preference models are built to compute geographical, categorical, social, and temporal influences, and a personalized weight is estimated for each criterion (i.e., each influence). By utilizing the personalized criterion’s weight, dominance intensity measures (i.e., dominating and dominated measures) are computed for alternatives (i.e., candidate events) of each criterion, and the set of alternatives is ranked based on the estimated dominance intensity measures to recommend $k$ top-ranked events. Extensive experiments are conducted based on two large real-world data sets collected from Meetup.com to evaluate the performance of SoCaST*. Experimental results show that SoCaST* performs better than the state-of-the-art techniques designed for event recommendations.

INDEX TERMS Event recommendations, multi-criteria decision making, event-based social networks.

I. INTRODUCTION
Technological advancement and increasing use of smart phones paved the way for event-based social networks (EBSNs) (e.g., Meetup.com). Meetup.com has attracted millions of users worldwide. Specifically, Meetup.com has over 35 million users and 300,000 groups in 182 countries since 2002 [1]. In general, an EBSN allows individuals with common interests to join groups and get together in a certain location and at a particular time [2] to participate in events (e.g., party, boat trip, and hiking), as depicted in Figure 1. Because of the enormous number of available events, the task of searching for events that best match a user’s preference is essential, difficult, and time consuming in EBSNs. To this end, an event recommender system helps users filter overwhelming information from community-contributed data and provides relevant recommendations, which make EBSNs more attractive to them.

The key objective of the problem of personalized event recommendations in EBSNs is to model a user’s preferences based on her historical attendance records to recommend events. The existing event recommendation techniques [2]–[9] model user preferences based on the straightforward integration of some of the following criteria. (1) Geographical influence. The geographical location information (i.e., latitude and longitude coordinates) of an
event plays a major role on users’ attendance behaviors in EBSNs. For example, some users prefer attending events that are close to their residences or offices. The technique of using the one-dimensional geographical distance between locations is utilized to mine the geographical influence; unfortunately, this technique does not reflect the two-dimensional geographical attributes associated with locations. (2) Content/categorical influence. Whenever events are created, they are usually accompanied by information (i.e., text) and keywords (e.g., fishing, bar, and hiking) describing them and users would check if events matching their interests can be found. A dictionary-based similarity computational approach is employed on the description of events [3], [4]. Conversely, the work [10] shows that users have different biases on event categories; and thus, the categories of points of interest (POIs) are better used to model user preferences on the event nature. (3) Social influence. In the real world, people establish relationships and interact with one another, so also in EBSNs. These social relationships are in form of groups, i.e., a user joins groups that match her interests; and thus, a user’s preference can be influenced by the group hosting the event or the members of that group. (4) Temporal influence. Time is an important factor that influences users’ attendance behaviours in EBSNs. Techniques such as Gauss formula [3] and cosine function [4] have been employed to estimate the temporal influence of upcoming events on a user from her event attendance records.

Existing techniques used for fusing different influences, such as geographical, categorical, social, and temporal influences, have two key limitations: (1) The same weight for all the influences. The assumption of existing fusion techniques for all the influences having the same importance to all users does not truly reflect their preferences in EBSNs. In the reality, users give high priorities or weights to some criteria but low priorities to others, and such weights vary from users to users. For example, a user who is a foodie first checks if the event category is related to food before checking other criteria such as the location or time of the event. Thus, we can say that the user gives a higher priority to the event category and the chance for her to attend the event may increase if it falls in her interested categories. (2) Aggregate evaluation. Existing event recommendation techniques employ a fusion method (e.g., the product or sum rule) to calculate an aggregate score for a candidate event to predict the chance for a user to attend it. After that, they rank all the candidate events based on their aggregate scores and recommend the top ranked events to the user. In this paper, the MCDM-based technique is used to provide a more sophisticated framework for ranking alternatives (i.e., candidate events) based on different criteria instead of an aggregate score to improve the quality of decisions (i.e., event recommendations) made on a set of conflicting criteria.

Real-World Motivating Examples: As our previous findings indicated that geographical, categorical, and temporal influences can significantly affect users’ preferences or attendance behaviors in EBSNs [2], [9], in this paper we conduct analysis on the data set crawled from Meetup.com to show that users have different degrees of importance to different influences. Three users are selected from Meetup.com New York data set in which Users 1, 2, and 3 have attended 387, 217, and 306 events, respectively. Obviously, these three users weigh differently on the geographical, categorical, and temporal preferences, as depicted in Figures 2 and 3. User 1 who likes recreation and language events and prefers attending events in a certain region at night time. User 2 does not have any particular preference on the event location and the days of the week, but she likes attending events for socializing after midnight, i.e., the weight of the categorical influence should be higher than the weights of the geographical and temporal influences. User 3 is interested at events about food and health mostly during early hours of the day and has a strong preference on a few locations. The summary of the three users’ personalized weights on geographical, categorical, and temporal influences are listed in Table 1.

### Table 1. The weights on the geographical, categorical and temporal influences for Users 1, 2, and 3.

<table>
<thead>
<tr>
<th>User</th>
<th>Geographical</th>
<th>Categorical</th>
<th>Temporal</th>
</tr>
</thead>
<tbody>
<tr>
<td>User 1</td>
<td>Medium</td>
<td>Medium</td>
<td>High</td>
</tr>
<tr>
<td>User 2</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>User 3</td>
<td>High</td>
<td>Medium</td>
<td>Medium</td>
</tr>
</tbody>
</table>
In this paper, we propose a new event recommendation framework, called SoCaST*, to evaluate events based on their geographical, categorical, social, and temporal influences through the multi-criteria decision making (MCDM) technique [11]–[14]. SoCaST* addresses three key challenges that include (1) modeling of the geographical, categorical, social, and temporal influences of events on users; (2) estimating a user’s personalized weights on each influence through a distance-based method [15]; and (3) evaluating alternatives (i.e., events) through the MCDM-based technique known as dominance intensity measures [11] and ranking the alternatives in order to recommend the $k$ top-ranked events (i.e., best $k$ ranked events) for a user.

In SoCaST*, the geographical, categorical, social, and temporal influences are modeled to estimate a user’s preferences on upcoming events from past events participated by the user. Specifically, we learn and model the personalized two-dimensional geographical influence through an adaptive kernel density estimation (KDE) method, the categorical influence based on the event category and its popularity, the social influence as the relevance of a group to a user and her friends, and the temporal influence by employing the KDE method. After modeling the influences, SoCaST* takes each influence as an input criterion to estimate a user’s personalized weights on each criterion through an objective distance-based method that reflects the user’s perception of each influence because the significance of a criterion is a direct function of the information conveyed by it [15]. Then, a personalized event recommendation framework based on the dominance intensity measuring technique is used to build preference models for each criterion. Finally, a sorted list for top-$k$ recommendations is generated by computing dominance intensity of each event as an alternative. To the best of our knowledge, SoCaST* is the first event recommendation framework to employ personalized weighted geographical, categorical, social, and temporal influences for recommendations through the MCDM-based technique.

In this paper, we significantly extend our previous work [2], [9] by proposing a more sophisticated method for evaluating and ranking events through the MCDM-based technique; the new method utilizes personalized weights on the estimated influences. The contributions of this paper are summarized as follows:

- We propose SoCaST*, a MCDM-based recommendation framework that utilizes geographical, categorical, social, and temporal influences with their personalized weights for recommending events to users in EBSNs. The influences are estimated for an individual user based...
on her historical event attendance records, event information, and social relationships. (Section III)

- We consider each estimated influence as a criterion to determine a user’s personal weights on the criteria by using the objective distance-based method that reflects the user’s rating of each criterion. (Section IV)

- The MCDM-based technique, called dominance intensity measures, is developed for evaluating and ranking events for a user based on her preferences that are modeled by the estimated influences and personal weights. (Section IV)

- 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* provides better quality of event recommendations than the state-of-the-art event recommendation techniques, in terms of both precision and recall. (Sections V and VI)

The rest of this paper is organized as follows. Section II highlights related work. The system model and user preference models of SoCaST* are presented in Section III. Section IV describes our approach for weight estimation and personalized event recommendation framework. Experiment settings are given in Section V and experimental results are analyzed in Section VI. Finally, we conclude this paper in Section VII.

II. RELATED WORK

In this section, we highlight relevant research work in three categories, namely, conventional recommender systems, event recommender systems, and the applications of MCDM-based techniques in recommendations.

A. CONVENTIONAL RECOMMENDER SYSTEMS

Conventional recommender systems analyze the historical ratings of users on items (or locations) to develop models that can predict users’ ratings on new items (or POIs) and recommend top-ranked items (i.e., items with the highest predicted ratings) to users. In general, geographical, content/categorical, social and temporal preferences of users have been often considered in recommender systems.

1) GEOGRAPHICAL INFLUENCES

Geographical influences have been widely explored in conventional recommender systems to improve the quality of item (or location) recommendations. Some studies [16]–[18] used the collaborative filtering technique to estimate the geographical influence on users. The collaborative filtering technique has an underlying assumption that similar users would like similar items. Other techniques that have been exploited for estimation of geographical influences are sentiment analysis [19] that processes ratings in form of likes and dislikes and the one-dimensional geographical distance between check-in locations of users as distributions [20]. Also, the KDE technique was used to model the two-dimensional geographical check-in distribution of locations over the latitude and longitude coordinates for each user [21], [22].

2) CONTENT AND CATEGORICAL INFLUENCES

The content information about items and POIs has been exploited in conventional recommender systems. Basically, the content-based methods compute the similarity between items (i.e., unseen items and rated items) by comparing their features and/or description for recommendations [23], [24]. For the categorical information about items and POIs, the probability approach [10] and the term frequency-inverse document frequency (TF-IDF) approach [21] were proposed for determining user preferences on the categorical information.

3) SOCIAL INFLUENCES

The social links among users have been considered to improve the quality of recommendations and alleviate the data sparsity and cold-start problems. The computation of social influences in conventional recommender systems is based on the assumption that friends have common interests and like to visit common POIs. Friend-based collaborative filtering is the most popular technique for estimating social influences in social networks [10], [18], [25], [26].

4) TEMPORAL INFLUENCES

As time plays an important role in activities of users, temporal influences have been widely considered for recommendations [27]–[31]. The probabilistic tensor factorization algorithm [32] and collaborative filtering algorithm [20], [33] were employed to model users’ check-in time slices to venue and estimate the temporal influence. Modeling time drifting patterns in user behaviors (i.e., the popularity of an item and the baseline ratings of a user change over time) was also used to improve recommendation accuracy [33].

B. EVENT RECOMMENDATIONS

In this section, we describe various techniques used to model users’ preferences for event recommendations in EBSNs.

1) GEOGRAPHICAL INFLUENCES

The geographical location is an important preference of users in EBSNs. The geographical influence was modeled through the Gauss formula to compute distance similarities between event locations [3] and the KDE method based on the one-dimensional distance distributions [4]. Other methods which have been used for estimating the geographical influence include K-mean [5], KDE on two-dimensional geographical location coordinates [2], and adaptive KDE on two-dimensional geographical location coordinates [9].

2) CONTENT AND CATEGORICAL INFLUENCES

The category or description of events significantly influences whether users want to attend them. Related techniques include content-based similarity method [34], TF-IDF

References models of SoCaST* are presented in Section III. High
to related work. The system model and user pref-
also the KDE technique was used to model
distance between check-in locations of users as distri-
of likes and dislikes and the one-dimensional geographi-
are sentiment analysis [19] that processes ratings in form
have been exploited for estimation of geographical influ-
collaborative filtering technique has an underlying assumption that similar
geographical influence on users. The collaborative filter-
ing technique has an underlying assumption that similar users would like similar items. Other techniques that have been exploited for estimation of geographical influences are sentiment analysis [19] that processes ratings in form of likes and dislikes and the one-dimensional geographical distance between check-in locations of users as distributions [20]. Also, the KDE technique was used to model the two-dimensional geographical check-in distribution of locations over the latitude and longitude coordinates for each user [21], [22].

2) CONTENT AND CATEGORICAL INFLUENCES

The content information about items and POIs has been exploited in conventional recommender systems. Basically, the content-based methods compute the similarity between items (i.e., unseen items and rated items) by comparing their features and/or description for recommendations [23], [24]. For the categorical information about items and POIs, the probability approach [10] and the term frequency-inverse document frequency (TF-IDF) approach [21] were proposed for determining user preferences on the categorical information.

3) SOCIAL INFLUENCES

The social links among users have been considered to improve the quality of recommendations and alleviate the data sparsity and cold-start problems. The computation of social influences in conventional recommender systems is based on the assumption that friends have common interests and like to visit common POIs. Friend-based collaborative filtering is the most popular technique for estimating social influences in social networks [10], [18], [25], [26].

4) TEMPORAL INFLUENCES

As time plays an important role in activities of users, temporal influences have been widely considered for recommendations [27]–[31]. The probabilistic tensor factorization algorithm [32] and collaborative filtering algorithm [20], [33] were employed to model users’ check-in time slices to venue and estimate the temporal influence. Modeling time drifting patterns in user behaviors (i.e., the popularity of an item and the baseline ratings of a user change over time) was also used to improve recommendation accuracy [33].

B. EVENT RECOMMENDATIONS

In this section, we describe various techniques used to model users’ preferences for event recommendations in EBSNs.

1) GEOGRAPHICAL INFLUENCES

The geographical location is an important preference of users in EBSNs. The geographical influence was modeled through the Gauss formula to compute distance similarities between event locations [3] and the KDE method based on the one-dimensional distance distributions [4]. Other methods which have been used for estimating the geographical influence include K-mean [5], KDE on two-dimensional geographical location coordinates [2], and adaptive KDE on two-dimensional geographical location coordinates [9].

2) CONTENT AND CATEGORICAL INFLUENCES

The category or description of events significantly influences whether users want to attend them. Related techniques include content-based similarity method [34], TF-IDF
method [4], latent Dirichlet allocation (LDA) method [7], and modeling category popularity with user categorical preferences [9].

3) SOCIAL INFLUENCES
In EBSNs, social relationships are in form of groups [7], [35] or conventional friendships [8], [34]. Other methods used to compute social influence of events include employing both friendships and group membership [6], participation records in a group [4], and the relevance of the group to the user and its members [2].

4) TEMPORAL INFLUENCES
Users usually attend events that fit their schedules and habits. Temporal influences were considered for event recommendations by using various techniques including the Gauss formula for modeling temporal influences based on the hour of the day and the day of the week of events [3], the cosine function for computing temporal influences based on a $24 \times 7$ matrix [4], and the KDE method for estimating temporal influences [9].

C. MULTI-CRITERIA DECISION MAKING (MCDM) IN RECOMMENDATIONS
Recommender systems usually mine user preferences based on various criteria, such as geographical, categorical, social and temporal criteria. In general, a model is built to mine user preferences for each criterion, and the model is used to predict the relevance score of a new item (or POI) for a user. The relevance scores of all the criteria are combined into a unified preference score. Finally, the top-$k$ items with the highest preference scores are recommended to the user. The most simple and common methods to integrate the relevance scores of various criteria into a single preference score are the sum and product rules [2], [9], [10], [36]–[38]. However, such integration techniques assume that all the criteria have the same importance to the user; and thus, it is essential to employ more sophisticated techniques to consider multiple criteria with different weights for providing better quality of recommendations.

The common approach used for dealing with the multi-criteria problem in the area of databases is skyline queries [39]–[42]. Skyline queries usually assume that all criteria have the same importance. They are estimated by iterative ordering alternatives to select the best ones that are not dominated by any other. Since skyline queries have no trade-off mechanism, some of the best alternatives may be good in terms of one criterion, but are bad for other criteria, i.e., the selected alternatives may not be the best across all criteria.

MCDM is a well-known branch of decision making [14], [43]. MCDM evaluates and ranks a set of alternatives based on multiple conflicting criteria and selects the best ones by aggregating them with a trade-off mechanism in an iterative manner [44], [45]. There are two major types of evaluation methods for MCDM, namely, cardinal and ordinal. The cardinal evaluation aggregates the preferences of an alternative over all the criteria by using the weighted sum method or other hybrid methods [46], while the ordinal evaluation aggregates the preferences on the alternatives’ positions in the preference preorders [21], [47].

III. SYSTEM MODEL AND USER PREFERENCES MODELING
In this section, we present the system model of SoCaST* and its user preference models for the geographical, categorical, social, and temporal influences. The key terms and problem of SoCaST* are defined in this section and Table 2 lists the key notations used in this paper.

### TABLE 2. Key notations and their meanings.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E$</td>
<td>Set of events $E = {e_1, e_2, \ldots, e_{</td>
</tr>
<tr>
<td>$U$</td>
<td>Set of users $U = {u_1, u_2, \ldots, u_{</td>
</tr>
<tr>
<td>$G$</td>
<td>Set of groups $G = {g_1, g_2, \ldots, g_{</td>
</tr>
<tr>
<td>$C$</td>
<td>Set of categories of events $C = {c_1, c_2, \ldots, c_{</td>
</tr>
<tr>
<td>$E(u), C(u)$</td>
<td>Sets of events and categories of events attended by user $u$</td>
</tr>
<tr>
<td>$E(g), U(g)$</td>
<td>Sets of events and users of group $g$</td>
</tr>
<tr>
<td>$E(c)$</td>
<td>Sets of events associated with category $c$</td>
</tr>
<tr>
<td>$G(u)$</td>
<td>Set of groups joined by user $u$</td>
</tr>
<tr>
<td>$L(u)$</td>
<td>Set of past event locations visited by user $u$</td>
</tr>
<tr>
<td>$T$</td>
<td>Set of time intervals $T = {00:00, 00:01, \ldots, 23:59}$</td>
</tr>
<tr>
<td>$T(u)$</td>
<td>Set of time of events attended by user $u$</td>
</tr>
<tr>
<td>$\text{link}(u_i, u_j)$</td>
<td>If $u_i$ and $u_j$ are in a common group (i.e., have a social link), $\text{link}(u_i, u_j) = 1$; otherwise, $\text{link}(u_i, u_j) = 0$</td>
</tr>
<tr>
<td>$E^*(u)$</td>
<td>Set of candidate events for user $u$, $E^*(u) = {e_1, e_2, \ldots, e_{</td>
</tr>
<tr>
<td>$R$</td>
<td>Set of criteria $R = {r_1, r_2, \ldots, r_{</td>
</tr>
<tr>
<td>$e_{ik}$</td>
<td>The influence score of criterion $r_k$ for event $e_i$</td>
</tr>
<tr>
<td>$w_k$</td>
<td>Weight of a criterion $r_k$</td>
</tr>
<tr>
<td>$M_L(</td>
<td>U</td>
</tr>
<tr>
<td>$M_P(</td>
<td>U</td>
</tr>
<tr>
<td>$M_T(</td>
<td>U</td>
</tr>
<tr>
<td>$M_C(</td>
<td>U</td>
</tr>
</tbody>
</table>

**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 date and time, location, and category of $e$, respectively, and $U(e)$ is a set of users committed to join $e$. 

VOLUME 6, 2018 27583
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., \( \text{link}(u_i, u_j) = 1 \) if both \( u_i \) and \( u_j \) are in a common group (i.e., \( u_i, u_j \in U(g) \) and \( g \in G \)); otherwise, \( \text{link}(u_i, u_j) = 0 \), i.e., \( u_i \) and \( u_j \) do not have any common group.

Problem Definition: 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. SoCaST* is designed to recommend events to users in an EBSN. 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. For user \( u \) with a set of events \( E(u) \) that were attended by \( u \) or \( u \) committed to attend and a set of candidate events \( E^*(u) \) that \( u \) can attend in the future, SoCaST* determines the geographical influence \( GI(e, u) \), categorical influence \( CI(e, u) \), social influence \( SI(e, u) \), and temporal influence \( TI(e, u) \) of \( E^*(u) \) on \( u \), ranks \( E^*(u) \), and recommends the \( k \) top-ranked candidate events to \( u \). The key objective of SoCaST* is to maximize the quality of event recommendations through mining user preferences and MCDM-based techniques.

A. GEOGRAPHICAL INFLUENCE MODELING

In SoCaST*, we model the geographical influence of events through an adaptive KDE method based on personalized two-dimensional location distribution of a user. The estimation of geographical influence has three steps, namely, pilot estimation, local bandwidth determination, and adaptive kernel estimation.

Step 1 (Pilot Estimation): The pilot estimation is computed based on the densities from KDE with a fixed bandwidth. Since we are considering each event’s location as a pair of latitude and longitude \((\text{lat}, \text{lon})\), the KDE function for \( L(u) \) (i.e., the set of locations visited by \( u \) to attend events) of user \( u \) is defined as:

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

where \( l_i = (\text{lat}_i, \text{lon}_i)^T \) is a two-dimensional location vector, \( \sigma \) is the bandwidth, \( N \) is the number of locations in \( L(u) \), and \( K(.) \) is the Gaussian kernel function [48] that is given as:

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

The bandwidth for the Gaussian kernel is defined as:

\[
\sigma = N^{-\frac{1}{6}} \sqrt{\frac{\hat{\sigma}^T \hat{\sigma}}{2}},
\]

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

Step 2 (Local Bandwidth Determination): This step combines the features of nearest neighbour and kernel approaches [49] by increasing or decreasing the bandwidth in an area with low or high data density, respectively [48]. The basic ideas of the adaptive KDE method are to remove noise in the area where the distribution of data is sparse, but recover details in the area where the distribution of data is dense. This step first determines the initial estimation to generate a series of bandwidths that correspond to the observations; this estimation gives an overview of the density [49]. Then, it constructs the adaptive estimation from the bandwidths. The local bandwidth weight is inversely proportional to the initial density estimation and is determined by:

\[
\omega_i = \left( \frac{\hat{g}}{f(x_i)} \right)^2,
\]

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

\[
\hat{g} = \left( \prod_{i=1}^{n} \hat{f}(x_i) \right)^{\frac{1}{n}}.
\]

Step 3 (Adaptive Kernel Estimation): Having computed the local bandwidth \( \omega_i \) by Equation (4), the probability of \( u \) attending an event at location \( e.l \) on user \( u \) is computed as follows:

\[
f_A(l|L(u)) = \frac{1}{N \sigma^2} \sum_{l_i \in L(u)} \frac{m_i(u, l_i)}{\omega_i} \cdot K \left( \frac{1 - l_i}{\omega_i \sigma} \right).
\]

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

\[
GI(e, u) = f_A(e.l|L(u)).
\]

B. CATEGORICAL INFLUENCE MODELING

In EBSNs, each event is usually associated with at least one category. The participation frequency of an event’s category for a user represents her preference on the event category. The estimation of the categorical influence of an event category on a user involves the relevance of the event category to the user and the popularity of the event category.

1) CATEGORY RELEVANCE TO A USER

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

\[
\hat{c}(e.c, u) = \text{tf-idf}(u, e.c),
\]

where tf-idf\((u, e.c)\) is the TF-IDF weight [50] of user \( u \) for event category \( e.c \), that is computed by:

\[
\text{tf-idf}(u, c) = \frac{m_c(u, c)}{|E(u)|} \cdot \log \frac{|E|}{|E(c)|},
\]

where \( m_c(u, c) \) is the number of events associated with category \( c \) attended by \( u \), \( |E(u)| \) is the total number of events attended by \( u \), \( |E| \) is the total number of events, and \( E(c) \) is the set of events associated with category \( c \).
2) CATEGORY POPULARITY

In EBSNs, the participants of an event affect whether a user attends the event. In other words, there is a higher chance for the user attending the event if more group members committed to attend it; and hence, the categorical influence model should consider the popularity of an event category. Given group $g$ for event $e$ and a set of users of $g$ (i.e., $U(g)$), the category popularity of $c$ in $g$ is given by:

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

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

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

C. SOCIAL INFLUENCE MODELING

In EBSNs, users can join groups and participate in events created by their groups. Since the members in the same group usually share the common interest and they may meet each other in the group’s events, they are considered as friends (i.e., a social link is established between every pair of members in a group). The social influence of a group to a user is modelled as the relevance of the group to the user and her friends.

1) GROUP RELEVANCE TO A USER

Here, we estimate the relevance of group $g$ to user $u$ among all the groups joined by $u$ (i.e., $G(u)$) as follows:

$$\hat{s}(u, g) = \frac{m_p(u, g)}{\sum_{g_i \in G(u)} m_p(u, g_i)}. \quad (12)$$

where $m_p(u, g)$ is the number of events created by $g$ and attended by $u$.

2) SOCIAL GROUP RELEVANCE

The relevance of group $u$ to user $u$’s friends is called as social group relevance. Note that the friendship between $u_i$ and $u_j$ is defined by $\text{link}(u_i, u_j) = 1$, as in Definition 3. The social group relevance that is basically measured by the similarity of the group relevance to $u$’s friends is computed by:

$$s_g(u, g) = \frac{\sum_{u_i \in U(g) \land u_i \neq u} \text{sim}(u, u_i) \cdot m_p(u_i, g)}{\sum_{u_i \in U(g) \land u_i \neq u} \text{sim}(u, u_i)}. \quad (13)$$

where $\text{sim}(u_i, u_j) = \frac{\sum_{g_k \in G(u_i) \land g_k \in G(u_j)} m_p(u_i, g_k) \cdot m_p(u_j, g_k)}{\sum_{g_k \in G(u_i)} m_p(u_i, g_k) \cdot \sum_{g_k \in G(u_j)} m_p(u_j, g_k)}$ calculates the similarity between users $u_i$ and $u_j$. Then, $s_g(u, g)$ is normalized as:

$$\hat{s}_g(u, g) = \frac{s_g(u, g)}{\sum_{g \in G(u)} s_g(u, g)}. \quad (14)$$

Finally, the social influence of group $g$ on user $u$ is computed by:

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

D. TEMPORAL INFLUENCE MODELING

Discretization of continuous time into time slices would cause information loss. To this end, we employ the KDE technique to compute the continuous time probability density of events to infer the temporal influence of events on a user. Note that $T(u)$ denotes the set of the time of the events attended by user $u$. The continuous time probability $f(t|T(u))$ is computed by:

$$f(t|T(u)) = \frac{1}{|D(u)|} \sum_{t_i \in T(u)} m_1(u, t_i) \cdot K\left(\frac{t - t_i}{\sigma}\right), \quad (16)$$

where $\sigma$ is the bandwidth, $K(.)$ is the kernel function (Equation (2)) that satisfies the condition:

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

and $t \ominus t_i$ is the time difference between two instances of time (i.e., $t$ and $t_i$). The time difference is calculated by [51]:

$$t \ominus t_i = \begin{cases} |t - t_i|, & \text{if } |t - t_i| \leq 12:00; \\
24:00 - |t - t_i|, & \text{if } |t - t_i| > 12:00. \end{cases} \quad (18)$$

Therefore, the temporal influence is determined as:

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

IV. MULTI-CRITERIA DECISION MAKING FOR EVENT RECOMMENDATIONS

In this section, we describe the multi-criteria decision making (MCDM) technique for SoCaST*. Basically, each estimated influence is considered as a criterion for MCDM and the influences are conflicting criteria. This section is organized as follows: Sections IV-A and IV-B first present the preliminaries of the proposed MCDM technique and the personalized weight estimation method for criteria, respectively, and then the technical details of our MCDM technique are described in Section IV-C. Finally, we give an example to illustrate our MCDM technique designed for SoCaST* in Section IV-D.

A. PRELIMINARIES OF MCDM

The geographical, categorical, social, and temporal influences of candidate events $e \in E^*(u)$ on user $u$ are estimated by $GI(e, u), CI(e, u), SI(e, u)$, and $TI(e, u)$, respectively. The key terms and concepts of MCDM are defined as follows.

**Definition 4 (Candidate Events):** A set of candidate events $E^*(u) = \{e_1, e_2, \cdots, e_{|E^*|}\}$ is a set of alternatives that could be recommended to user $u$.

**Definition 5 (Criteria):** There are four criteria $R = \{r_1, r_2, r_3, r_4\}$ that represent the geographical, categorical, social, and temporal influences, respectively. For candidate event $e_i$, the four influence scores of $e_i$ for user $u$ are denoted as $e_{i1} = GI(e_i, u), e_{i2} = CI(e_i, u), e_{i3} = SI(e_i, u)$, and $e_{i4} = TI(e_i, u)$.

**Definition 6 (Weights):** A set of weights for user $u$ is denoted as $W(u) = \{w_1, w_2, w_3, w_4\}$ in which each weight $w_k$ indicates the importance of criterion $r_k$ to $u$. 
Definition 7 (Ordinal Relations): Given a set of candidate events for user $u$, $E^*(u) = \{e_1, e_2, \ldots, e_{|E^*(u)|}\}$, the following ordinal relations can be defined.

- $e_i > e_j$ (Dominating): for two candidate events $e_i$ and $e_j$, $e_i$ is preferred to $e_j$ if $PM_{ij} \geq 0$ where $PM_{ij}$ is their paired-dominance measure, as described in Section IV-C.
- $e_i < e_j$ (Dominated): for two candidate events $e_i$ and $e_j$, $e_j$ is preferred to $e_i$ if $PM_{ij} < 0$.

B. USER PERSONALIZED WEIGHT ESTIMATION

It is essential to mine users’ preferences on each criterion because they may want to prioritize some criteria over some other criteria. For example, a foodie will check if an event is related to dining before checking other criteria; thus, the categorical criterion is the most important one. Based on each user’s historical event attendance records, we can mine her preferences on different criteria and compute different weights on them (i.e., $w_k$ for each criterion $r_k$ in Equation (22)). Basically, our proposed user personalized weight estimation employs the distance-based method to determine a user’s preferences on each criterion based on her attended events.

Algorithm 1 User Personalized Weight Estimation

1: function WeightEstimation ($u$, $R$, $E(u)$)
2: //Target user: $u$; a set of criteria: $R = \{r_1, r_2, r_3, r_4\}$; and a set of events attended by $u$: $E(u)$
3: Split $E(u)$ into training set $E_t(u)$ and evaluation set $E_e(u)$;
4: for each event $e_i \in E_t(u)$ do
5: for each criterion $r_k \in R$ do
6: Compute influence score $e_{ik}$ based on $E_t(u)$;
7: Normalize $e_{ik}$ to $e'_{ik} = \frac{e_{ik} - \min_{e_{ik}}}{\max_{e_{ik}} - \min_{e_{ik}}}$;
8: end for
9: end for
10: for each criterion $r_k \in R$ do
11: $U_k^+ = \max_{e_{ik} \in E_t(u)} (e'_{ik})$;
12: $U_k^- = \min_{e_{ik} \in E_t(u)} (e'_{ik})$;
13: $d_k^+ = \sqrt{\sum_{i=1}^{|E_t(u)|} (e'_{ik} - U_k^+)^2}$;
14: $d_k^- = \sqrt{\sum_{i=1}^{|E_t(u)|} (e'_{ik} - U_k^-)^2}$;
15: $w_k = \frac{\xi_k}{\sum_{k=1}^{4} \xi_k}$, where $\xi_k = \frac{d_k^+}{d_k^+ + d_k^-}$;
16: end for
17: return $W(u) = \{w_1, w_2, w_3, w_4\}$;

Algorithm 1 depicts the pseudo code for the user personalized weight estimation. Given user $u$ and the set of events attended by $u$ (i.e., $E(u)$), $E(u)$ is split into a training set $E_t(u)$ (i.e., 60% of events in $E(u)$) and an evaluation set $E_e(u)$ (i.e., 40% of events in $E(u)$). The influence score of each criterion is computed for each event $e_i \in E_t(u)$ based on the events in $E_t(u)$; and hence, we have $e_{ik} = \{e_{ik}, e_{ik}, \ldots, e_{ik}\}_{i \in E^*(u)}$ for each criterion $r_k \in R$. To remove the effect of data magnitude, these influence scores are normalized into $[0, 1]$ interval. The normalized evaluation value for the influence score $e_{ik}$ is given as:

$$e'_{ik} = \frac{e_{ik} - \min_{e_{ik}}}{\max_{e_{ik}} - \min_{e_{ik}}},$$

where $\min_{e_{ik}}$ and $\max_{e_{ik}}$ are the minimum and maximum influence scores for criterion $r_k$, respectively (Lines 4 to 9 in Algorithm 1).

The distance-based method [15] is based on optimistic and pessimistic utility values because they reflect how a user responds to things in her perceived environment and can be used to determine event predictions [52]. Since each criterion $r_k$ is positive criterion (i.e., the higher is the better), the distances between candidate events are given by:

$$d_k^+ = \sqrt{\sum_{i=1}^{|E(u)|} (e'_{ik} - U_k^+)^2} \quad \text{and} \quad d_k^- = \sqrt{\sum_{i=1}^{|E(u)|} (e'_{ik} - U_k^-)^2},$$

where $U_k^+ = \max_{e_{ik} \in E(u)} (e'_{ik})$ and $U_k^- = \min_{e_{ik} \in E(u)} (e'_{ik})$ are the criterion’s optimistic and pessimistic utility values (i.e., the maximum and minimum influence scores of criterion $r_k$ for the events in $E(u)$), respectively. Finally, the user personalized weight for criterion $r_k$ is computed by:

$$w_k = \frac{\xi_k}{\sum_{k=1}^{4} \xi_k},$$

where $\xi_k = \frac{d_k^+}{d_k^+ + d_k^-}$ is a dispersion measurement (Lines 10 to 16 in Algorithm 1). As a result, we have the personalized weights on each criterion $r_k \in R$ for user $u$, $W(u) = \{w_1, w_2, w_3, w_4\}$.

C. DOMINANCE INTENSITY

In this section, the key objective is to rank the best candidate events based on their influence scores because the $k$ top-ranked candidate events are recommended to the user. The underlying technique of the alternative ranking is based on the dominance intensity model consists of two major parts, dominance measures (i.e., dominating and dominated measures) and dominance intensity estimation.

1) DOMINANCE MEASURES

The dominance measure technique employed in SoCaST* is called pairwise dominance [11], [53] and it is used to compare ordered pairs of candidate events $e_i, e_j \in E^*(u)$, such that

$$PM_{ij} = \min_{r_k \in R} \{w_k(e_{ik} - e_{jk})\}.$$  

If $PM_{ij} \geq 0$, candidate event $e_i$ dominates $e_j$; otherwise, $e_j$ dominates $e_i$, as described in Definition 7. The pairwise dominance method can further prioritize competitive alternatives leading to recommendation of the best alternatives. This is the advantage for the pairwise dominance method over the absolute dominance method that is based on boundaries difference (i.e., upper and lower bounds) and often results in many non-dominated alternatives [53].
2) DOMINANCE INTENSITY ESTIMATION

In SoCaST*, the dominance intensity is dependent on the dominating intensity ratio \( I^+ \) and dominated intensity ratio \( I^- \) (Algorithm 2). The dominating intensity ratio is given as:

\[
I^+ = \frac{Q_i^{+\star}}{Q_i^{\star\star} - Q_i^{\star<}}, \tag{23}
\]

where \( Q_i^{+\star} \) and \( Q_i^{\star<} \) are the dominating indices (Lines 12 to 16 in Algorithm 2). These indices are given as:

\[
Q_i^{+\star} = \sum_{j=1/|E^*(u)|}^{j \neq i} PM_{ij} \text{ for } PM_{ij} \geq 0; \quad \text{and} \quad Q_i^{\star<} = \sum_{j=1/|E^*(u)|}^{j \neq i} PM_{ij} \text{ for } PM_{ij} < 0. \tag{24}
\]

Note that \( Q^+ = Q_i^{+\star} + Q_i^{\star<} \) and \( 0 \leq I^+ \leq 1 \). Likewise, the dominated intensity ratio is given as:

\[
I^- = \frac{Q_i^{-\star}}{Q_i^{\star\star} - Q_i^{\star<}}, \tag{26}
\]

where \( Q_i^{-\star} \) and \( Q_i^{\star<} \) are the dominated indices (Lines 19 to 23 in Algorithm 2) that are calculated as:

\[
Q_i^{-\star} = \sum_{j=1/|E^*(u)|}^{j \neq i} PM_{ij} \text{ for } PM_{ij} \geq 0; \quad \text{and} \quad Q_i^{\star<} = \sum_{j=1/|E^*(u)|}^{j \neq i} PM_{ij} \text{ for } PM_{ij} < 0. \tag{27}
\]

Note that \( Q^- = Q_i^{-\star} + Q_i^{\star<} \) and \( 0 \leq I^- \leq 1 \). In SoCaST*, to estimate the overall dominance intensity for each candidate event, we subtract the alternative dominated ratio from its corresponding dominating ratio. Therefore, the overall dominance intensity \( D_i \) of candidate event \( e_i \) is computed by:

\[
D_i = I^+ - I^-. \tag{29}
\]

where \(-1 \leq D_i \leq 1 \) (Line 28 in Algorithm 2).

3) EVENT RECOMMENDATIONS

Algorithm 3 gives the pseudo code for SoCaST*. A set of candidate events for user \( u \) is considered as a set of candidate events \( E^*(u) \). Algorithm 1 is executed to determine a set of personalized weights for the criteria and Algorithm 2 is performed to compute the dominance intensity for each candidate event in \( E^*(u) \) (Lines 6 to 7 in Algorithm 3). Finally, the candidate events in \( E^*(u) \) are sorted by their overall dominance intensity values in decreasing order, and the \( k \) top-ranked candidate events are recommended to \( u \) (Lines 8 to 9 in Algorithm 3).

D. EXAMPLE

In this section, we use an example to illustrate SoCaST*, in which we assume the set of personalized weights for user \( u \) of the geographical influence (\( r_1 \)), categorical influence (\( r_2 \)), social influence (\( r_3 \)) and temporal influence (\( r_4 \)) is \( W(u) = \{0.21, 0.47, 0.17, 0.15\} \), a set of candidate events is
Next, we compute the dominating indices $Q^e_{-}$ of the candidate events, as depicted in Table 3. For instance, we have the dominance matrix as follows:

$$PM_{ij} = \begin{bmatrix} r_1 & r_2 & r_3 & r_4 \\ e_1 & 0.1623 & 0.0423 & 0.3295 & 0.3860 \\ e_2 & 0.4935 & 0.9604 & 0.7571 & 0.3049 \\ e_3 & 0.2786 & 0.5430 & 0.4752 & 0.1560 \\ e_4 & 0.5891 & 0.3272 & 0.7565 & 0.4162 \end{bmatrix}$$

By computing the paired-dominance in Equation (22), we have the dominance matrix as follows:

$$PM_{ij} = \begin{bmatrix} e_1 & e_2 & e_3 & e_4 \\ e_1 & -0.4315 & -0.2353 & -0.1339 \\ e_2 & 0.0223 & -0.0201 \\ e_3 & -0.0345 & -0.1962 \\ e_4 & 0.0045 & -0.2976 & -0.1014 \end{bmatrix}$$

For example, $PM_{12} = (0.21 \times (0.1623 - 0.4935), 0.47 \times (0.0423 - 0.9604), 0.17 \times (0.3295 - 0.7571), 0.15 \times (0.386 - 0.3049)) = \min(-0.0696, -0.4315, -0.0727, 0.0122) = -0.4315$. Next, we compute the dominating indices $Q^+_{e}$ and $Q^-_{e}$ and the dominated indices $Q^<_{e}$ and $Q^>_{e}$ for all the candidate events, as depicted in Table 3. For instance, the dominating indices of $e_1$ are $Q^+_{e_1} = 0$ and $Q^-_{e_1} = (-0.4315) + (-0.2353) + (-0.1339) = -0.8007$, while the dominated indices of $e_1$ are $Q^<_{e_1} = 0.0045$ and $Q^>_{e_1} = (-0.0122) + (-0.0345) = -0.0467$.

After the dominating and dominated intensity ratios (i.e., $I^+$ and $I^-$) are determined for the candidate events by Equations (23) and (26), respectively, the overall dominance intensity $D_i$ for each candidate event $e_i$ is computed. Finally, the candidate events are sorted by their overall dominance intensity values in decreasing order, i.e., $\{e_2, e_4, e_3, e_1\}$. In case of $k = 2$, $e_2$ and $e_4$ are recommended to $u$.

### Table 3. Dominance measures.

<table>
<thead>
<tr>
<th>Candidate events</th>
<th>$Q^+_{e}$</th>
<th>$Q^-_{e}$</th>
<th>$I^+$</th>
<th>$I^-$</th>
<th>$D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_1$</td>
<td>0</td>
<td>-0.8007</td>
<td>0.0885</td>
<td>0.0885</td>
<td></td>
</tr>
<tr>
<td>$e_2$</td>
<td>0.0223</td>
<td>-0.3022</td>
<td>0.4092</td>
<td>0.4092</td>
<td></td>
</tr>
<tr>
<td>$e_3$</td>
<td>0</td>
<td>-0.2959</td>
<td>0.0622</td>
<td>0.0622</td>
<td></td>
</tr>
<tr>
<td>$e_4$</td>
<td>0.0045</td>
<td>-0.3990</td>
<td>0.0112</td>
<td>0.0112</td>
<td></td>
</tr>
</tbody>
</table>

### Table 4. The statistical summary of the data sets.

<table>
<thead>
<tr>
<th></th>
<th>Members</th>
<th>Events</th>
<th>Groups</th>
<th>Venues</th>
<th>RSVP</th>
</tr>
</thead>
<tbody>
<tr>
<td>SF</td>
<td>324,485</td>
<td>169,120</td>
<td>11,367</td>
<td>18,126</td>
<td>1,665,591</td>
</tr>
<tr>
<td>NY</td>
<td>910,938</td>
<td>432,938</td>
<td>13,906</td>
<td>35,623</td>
<td>3,614,720</td>
</tr>
</tbody>
</table>

In our previous work, SoCaST utilizes user preferences on geographical, categorical, social and temporal influences to recommend events by integrating these influences using the product rule based on an assumption that all user preferences have the same weight. Specifically, SoCaST models the personalized two-dimensional geographical influence on a user with an adaptive KDE method; the categorical preference is modeled through the TF-IDF method; and the social influence is modeled as the group relevance to a user and her friends while the temporal influence is modeled through the KDE method.

### B. EVALUATED EVENT RECOMMENDATION TECHNIQUE

We compare our SoCaST* with the following state-of-the-art event recommendation methods as baselines:

- **SoCaST [9]**: In our previous work, SoCaST utilizes user preferences on geographical, categorical, social and temporal influences to recommend events by integrating these influences using the product rule based on an assumption that all user preferences have the same weight. Specifically, SoCaST models the personalized two-dimensional geographical influence on a user with an adaptive KDE method; the categorical preference is modeled through the TF-IDF method; and the social influence is modeled as the group relevance to a user and her friends while the temporal influence is modeled through the KDE method.

- **SKYLINE [41]**: This MCDM-based method continuously orders alternative values to obtain the best ones which are regarded as the most preferred alternatives. It employs the Pareto dominance method to eliminate alternatives which are dominated by other alternatives.

- **PAAT [3]**: This method uses a singular value decomposition with a multi-factor neighborhood to predict events attendance for members by exploiting time, distance and event content.

- **CAER [4]**: This technique considers event description, time, social group, and geographical location to rank and recommend relevant events to users.

- **CFM [33]**: CFM exploits the collaborative filtering method and dynamics of temporal influences on user ratings for event recommendations.

- **SRE [8]**: This technique exploits the selection of a group of friends to provide recommendations based on the social influence in EBSNs.
C. PERFORMANCE METRICS
Generally, every recommendation method computes the score of every candidate event for a user, ranks events based on their scores and returns the certain number of events with the highest scores to the user as a recommendation result. To evaluate the accuracy of recommendation results, it is important to find the number of recommended events that a user has visited in the testing data set. To this end, two standard metrics, namely, precision and recall [22], are employed to measure the quality of recommendations results.

 Precision is the ratio of the number of events that have been attended by a user to the total number of recommended $k$ top-ranked events, i.e.,

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

and recall is the ratio of the number of recommended $k$ top-ranked events that have been attended by a user to the total number of events attended by that user, i.e.,

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

D. PARAMETER SETTINGS
In the experiments, the number of recommended events (i.e., $k$) varies from 2 to 50 for regular users, while it increases from 1 to 10 for cold-start users, since they have attended smaller numbers of events. The training set (i.e., 60% events data with the earliest timestamps of a user) is also used for learning personalized weights on criteria.

VI. EXPERIMENTAL RESULTS
In this section, we evaluate the performance of our SoCaST* for both regular and cold-start users by comparing it with the state-of-the-art techniques in terms of precision and recall based on two large real-world data sets (Section VI-A). In addition, the effects of using personalized weights and the proposed MCDM-based technique in our SoCaST* are analyzed in Sections VI-C and VI-D, respectively.

A. COMPARISONS OF EVENT RECOMMENDATION TECHNIQUES FOR REGULAR USERS
In the experiments, users attended more than 10 events in the training set are considered as regular users. Fig. 4 depicts the experimental results of our SoCaST* and baselines with respect the various top-$k$ values (i.e., from 2 to 50) for both the NY and SF data sets. The performance of each evaluated technique is summarized as follows:

SRE: SRE gives the worst performance among all the baseline techniques because it considers only the social influence in its model.

CFM: As CFM employs the collaborative social and temporal influences for event recommendations, it performs better than SRE.

CAER: CAER incorporates the geographical (i.e., distance between home and event locations), social and temporal influences into its model, so it achieves higher accuracy than SRE and CFM.

PAAT: PAAT models the one-dimensional distance similarity as the geographical influence and fuses it with
content, social and temporal influences for event recommendations; and thus, its accuracy is better than CAER, CFM and SRE.

SKYLINE: The SKYLINE query is a multi-criteria technique based on alternatives comparisons to get the best ones for recommendations. Since it assumes the same weight for all the criteria (including the distance between home and event locations), the selected alternatives may not represent the best across all the criteria. Its performance is better than PAAT but worse than SoCaST and SoCaST*.

SoCaST: SoCaST which aggregates the geographical influence based on the two-dimensional geographical location coordinates with the categorical, social and temporal influences to provide event recommendations for users. It achieves the best performance among all the baselines.

SoCaST*: Our SoCaST* is the best technique in terms of both precision and recall. It performs better than SoCaST because it considers personalized weights on criteria and utilizes the MCDM technique to determine the best events for recommendations.

B. COMPARISONS OF EVENT RECOMMENDATION TECHNIQUES FOR COLD-START USERS
In this section, we evaluate the event recommendation techniques for the cold-start users who have attended less than 10 events in the training set (Fig. 5). Experimental results show that SoCaST* also performs better than all other techniques in terms of both precision and recall. Since the cold-start users have attended much fewer events than the regular users in the testing set, the precision of the cold-start users is lower than that of regular users and the recall of the cold-start users is higher than that of regular users.

C. EFFECT OF PERSONALIZED WEIGHTS ON THE CRITERIA
The effect of the personalized weights on the criteria in our SoCaST* is evaluated in this section. We compare our SoCaST* with the MCDM-based technique with the same weight for all the criteria, denoted as SoCaST* w/o PW. In SoCaST* w/o PW, the weights for the geographical, categorical, social and temporal influences are set to 0.25.

Figs. 6 and 7 depict the experimental results for the regular
and cold-start users, respectively. Our SoCaST\textsuperscript{*} performs better than SoCaST\textsuperscript{*} w/o PW in terms of both precision and recall for both the regular and cold-start users. As a result, the user usually has personal preferences on different criteria, and thus, the personalized weights on the criteria are very important for event recommendations.

D. COMPARISONS OF FUSION TECHNIQUES

In this section, we compare the performance of the proposed MCDM-based event recommendation technique in SoCaST\textsuperscript{*} with the product rule (Product rule with PW) and sum rule (Sum rule with PW) fusion methods that aggregate the geographical, categorical, social, and temporal influences into a single rating score for each candidate event and recommend the \( k \) top-ranked candidate events with the highest rating scores to the user. Both the product and sum rule methods use the personalized weights on the criteria because we want to focus on the fusion method. Figs. 8 and 9 depict that our SoCaST\textsuperscript{*} performs better than the product and sum rule methods in terms of both precision and recall. The improvement of SoCaST\textsuperscript{*} is more significant for the cold-start users. As a result, the experimental results show that the MCDM-based technique is more effective than the simple product and sum rule fusion methods.

VII. CONCLUSION

In this paper, we have proposed a new event recommendation framework called SoCaST\textsuperscript{*} for event-based social networks. SoCaST\textsuperscript{*} first models geographical, categorical, social, and temporal influences of events on users based on their historical attendance records. Then, it learns the user’s personalized weights on these four criteria. Finally, the multi-criteria decision making (MCDM) technique is used to rank candidate events and the \( k \) top-ranked events are recommended to the user. Experiments were conducted on two large publicly available Meetup.com data sets in New York and San Francisco, USA to evaluate the performance of SoCaST\textsuperscript{*} for both the regular and cold-start users. Experimental results show that SoCaST\textsuperscript{*} outperforms the state-of-the-art techniques in terms of both precision and recall and the use of personalized weights and MCDM-based technique effectively improves the accuracy of SoCaST\textsuperscript{*}.

REFERENCES


