Information Sciences 480 (2019) 90–108

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

A statistical approach to participant selection in location-based social networks for offline event marketing

Yuxin Liu[a], Anfeng Liu[a,b], Xiao Liu[a,*], Xiaodi Huang[c]

[a] School of Information Science and Engineering, Central South University, ChangSha 410083 China
[b] Jiangsu High Technology Research Key Laboratory for Wireless Sensor Networks, Nanjing, Jiangsu 210003, China
[c] School of Computing and Mathematics, Charles Sturt University, Albury, NSW 2640, Australia

ARTICLE INFO

Article history:
Received 7 September 2018
Revised 9 November 2018
Accepted 13 December 2018
Available online 18 December 2018

Keywords:
Location-based social networks
Marketing cost
Offline event marketing
Statistical inference approach
Participant selection

ABSTRACT

Offline event marketing has become increasingly popular. As a large amount of data from location-based social networks (LBSNs), such as Foursquare, Gowalla, and Facebook, becomes available, how to make use of these data to analyze users' social behaviors is an important issue for offline event marketing. To provide some valuable guidance for businesses, this paper presents a statistical inference approach to optimally selecting participants who have a high probability of visiting an offline event. Technically, we formulate participant selection as a constraint optimization problem. In particular, our marketing cost model takes into account key factors such as distance, loyalty influence, and recommendation index. In addition, four participant-based strategies and a detailed algorithm are presented. Experiments on real-world datasets have demonstrated the effectiveness and efficiency of our proposed approach and the quantitative model.

1. Introduction

Markets that rely on media channels without being connected to the World Wide Web are called "offline" markets [17,19,20,24,29]. In other words, offline marketing utilizes offline media channels to create awareness of a company's products and services. Offline media campaigns include radio and print advertising, telemarketing, and television ads [29]. Given the Internet's tremendous rise in popularity [1,8,23], however, a company can accumulate data on a large number of users and merchandise through devices such as mobile phones [4,6,21], sensing devices [5,8,9,32], and RFID [10,11,20]. These data have a wealth of marketing value by which to improve the effectiveness of advertising. In fact, offline event marketing and associated strategies have received increasing attention [17,18].

An offline marketing event usually contains the following attributes [22,30].

a) *Location:* The address of a company or a store that holds the event in terms of its longitude and latitude. The place for holding an event is usually a business venue where products or services can be consumed by customers.
b) *Scale:* The scale of an event is measured by the number of customers that are allowed to participate in the event. Sometimes, the more participants attending the event, the better. However, given the capacity of a venue, only a limited number of participants can be served. This work aims to invite a limited number of customers who are highly likely to

* Corresponding author.
E-mail address: xiaoliu@csu.edu.cn (X. Liu).

https://doi.org/10.1016/j.ins.2018.12.028

participate in an event at a venue. At the same time, the expected number of visiting participants can equal a planned scale, called a scale goal.

c) *Served items:* Products or services in an event as well as their corresponding prices.
d) *Others:* A brief description of an event, such as the event duration, or some principles in selecting the participants to whom to present the event.

Except for advertising campaigns and creating promotional materials, the basic targets for any offline marketing strategy are to increase online traffic, overall sales, and profits [14,16]. For this purpose, various strategies are available. Among them, three representative strategies are described as follows:

a) Direct mail - one of the most commonly used mediums in offline marketing campaigns. Mailing lists based on demographic data should include customers who are most likely purchase the products. For instance, insurance companies may desire lists of customers who have recently purchased a new car or other things. This strategy only considers one factor. In fact, many factors can affect customers purchasing products, such as the travel distance of the customers purchasing the products and their interests.
b) Discount pricing - Another way of attracting customers is to offer price discounts, such as coupons. Department stores, supermarkets and companies often insert their coupons into local newspapers or leaflets to advertise the sales of their products and services. Although offering discounts will increase the desire for customer consumption, this strategy is not feasible for consumers who do not need such products or services. This scheme increases some unnecessary costs of advertising.
c) Loyalty programs - Loyal customers are an important resource for companies. To keep these customers and attract new customers, some preferential policies are introduced to encourage them to patronize businesses more frequently [32]. For example, a pet store may reward customers with a can of dog food on their tenth visit. Although marketers provide some preferential policies for encouraging customers, the probability of customers visiting companies may also be lower due to the long travel distance.

The essential problem of offline event marketing is how to propagate and advertise an event at a minimized marketing cost while fulfilling the scale and item coverage constraints. In other words, participants with a larger probability of attending an event are selected so that the resources invested in the event can be consumed as much as possible [7]. Some simple strategies, such as “nearest people first” (NF), random selection (RS), or first come first served (FCFS), cannot guarantee that the marketing cost is minimized. Instead, these strategies merely guarantee that the number of invited participants will not exceed the volume of an event as planned. A better strategy is “most potential people first” (MF). This assumes that the more potential a participant has, the more likely he/she will attend the event. Thus, the problem becomes how to measure the potential of a participant [25–27]. For offline event marketing, several important factors should be considered when applying the potential influence in participant selection: 1) distance: people are inclined to visit venues that are closer to their home locations. Thus, a far distance may cause a low visiting probability [15]; 2) item coverage: it would be better if all product items served by a venue could be consumed by the customers. Therefore, each item should be liked by the participants; 3) loyalty influence: if a customer is very loyal to a venue, then advertising to him/her is necessary because he/she is likely to come to the venue again [13]; and 4) influences from friends (called recommendation index): if a customer has many friends with considerable loyalty to a venue, then he/she has a high potential of visiting the venue because his/her friends may recommend the venue to him/her [25]. Our approach takes into account these factors for offline event marketing.

The main contributions of this paper are as follows.

(1) We form offline event marketing as a constraint optimization problem.
(2) We introduce the statistical inference approach by considering the distance factor, loyalty influence and recommendation index. Based on this approach, an effective algorithm is given.
(3) Four strategies that select customers based on three factors (distance, loyalty influence and recommendation index) are proposed:
(4) We conducted extensive experiments, which demonstrate the performance of the proposed algorithm and four strategies. The results show that strategy 3 based on unidirectional loyalty local influences and strategy 4 based on bidirectional loyalty local influences have better performance.

The rest of this paper is organized as follows. Section 2 reviews some related work. The system model and the problem of offline event marketing are well defined in Section 3. Section 4 presents an algorithm for minimizing the marketing cost. The experimental results are reported in Section 5. Finally, we conclude this paper in Section 6.

2. Related work

The availability of location-based social networks (LBSNs) offers a good opportunity to analyze user social behaviors [2,14,15,29,30,33]. There is a body of literature on studying the location of users and personal preferences in LBSNs [30,33]. We review some of the research relevant to our work in the following.

Location-based social network. Studying how users share their locations in the real world by collecting traces, the authors [14] presented a large-scale quantitative data analysis on user profiles [17], updated activities, mobility characteristics [19], social graphs, and attribute correlations [2,14].

Exploiting the social and geographical characteristics of users and locations, Ye et al. [29] presented a system for location recommendation services on a large scale. Through the spatial-social analysis of a large dataset collected from Foursquare, this work demonstrates that strong social and geospatial ties exist among users and their visited locations. For location recommendation, the authors present a *friend-based collaborative filtering* (FCF) approach based on collaborative filtering of locations among social friends. Moreover, a variant of the FCF is proposed based on heuristics derived from geospatial characteristics in the Foursquare dataset. The findings of this work confirm the important roles of user locations and their friends in the analysis of human social behaviors [23].

Trust and distrust relationships. Ma et al. [13] systematically exploited the effectiveness of incorporating trust and distrust information into a recommender system. The proposed approach is based on matrix factorization with a regularization term as the constraint of trust and distrust relations between users. This work reports a finding that distrust information plays a role that is at least as important as trust information. In offline marketing, a high visiting frequency of a customer to a venue can be regarded as his/her trust for the venue [13,17]. In contrast, other venues serving the same products and services are viewed as distrust information [13]. Additionally, trust and distrust information can be used in commercial competition analysis [11,12].

Participant selection approach for marketing events. As our most closely related work, Yu et al. [30] presented a framework that aims to maximize the marketing effect (the number of customers that are influenced in terms of visiting the venue) by carefully selecting invitees to a sponsored offline event by leveraging location-based social networks. The work first casts participant selection as a multiple-objective optimization problem and then introduces a metaheuristics algorithm based on simulated annealing (SA) to obtain a participant team, which is a nearly optimal solution in all possible participant combinations. Specifically, this work considers distance, overlapping social influence, and item coverage in leveraging LBSNs. This work demonstrates that overlapping social influence is more important than distance in participant selection, and thus, more friendship information can improve the prediction result, and less historical data are required. The main idea of our approach to maximize the marketing effect considers four factors: (a) the distance factor, (b) loyalty influence, (c) the recommendation index, and (d) the item coverage constraint. Ref. [29] proposed a friend-based collaborative filtering (FCF) approach for location recommendation based on collaborative ratings of places made by social friends. Ref. [29] mainly considered the distance and did not consider the loyalty influence and the item coverage constraint. In [30], the participant selection only considered the distance and overlapping social influence. If the distance of the participants the venues is closer and the social influence of the participants is larger, then the participants are selected. However, this scheme does not consider the loyalty influence of the participants to the venues, and the recommendation index does not consider the process of selecting participants. As is known, most information of participants comes from the recommendations of his/her friends, and it is most important to consider the recommendations of friends. For example, if participant A likes to visit venue 1, then he/she will recommend the information to his/her friends. Our approach considers the four factors comprehensively to select the participants; thus, the performance can be improved.

Differing from the above work, our inference approach is capable of predicting the probability of customers visiting a given event from social network datasets. In particular, those who are likely to attend an event will be selected as participants from a given set of customers. We propose an algorithm for participant selection with the time complexity of $nlog_2n$.

3. System model and offline marketing constraints

3.1. The system model

We list the notations used in this paper in Table 1. A set $Set_c = \{c_1,\ c_2,\ \ldots, c_{N_c}\}$ denotes the universal set of customers, and $Set_v = \{v_1,\ v_2,\ \ldots, v_{N_v}\}$ denotes a universal venue set, where N_c and N_v are the total number of customers and venues, respectively. Each customer or venue is associated with a set of tags. The universal tag set is $Set_t = \{t_1,\ t_2,\ \ldots, t_{N_t}\}$, where N_t is the tag number. The tags of venue v_i denoted as T_{v_i} indicate the products or services the venue usually supplies, which is represented as $T_{v_i} = t_1(v_i),\ t_2(v_i),\ \ldots, t_{N_t}(v_i)$, where $t_j(v_i) = 1$ means that venue v_i has this product or service, and $t_j(v_i) = 0$ otherwise, $i = 1,\ 2,\ \ldots,\ N_v$ and $j = 1,\ 2,\ \ldots,\ N_t$. The tag of customer c_k represented as T_{c_k} depicts the product or service the customer prefers. We have $T_{c_k} = t_1(c_k),\ t_2(c_k),\ \ldots, t_{N_t}(c_k)$, where $t_j(c_k) = 1$ if customer c_k likes this product or service, and $t_j(c_k) = 0$, otherwise, where $k = 1,\ 2,\ \ldots,\ N_c$. Let $d(c_i, v_j)$ denote the distance from the home location of customer c_i to venue v_j, whereas $v(c_i, v_j)$ is whether c_i has visited v_j $(v(c_i,\ v_j) = \{1\ if\ visited,\ 0\ otherwise\})$, and $f(c_i, v_j)$ is the monthly number of visits of c_i's friend to venue v_j. A customer c_j is considered a friend of customer c_i if c_i follows c_j on social media, such as Twitter, but it does not mean c_i is c_j's friend unless c_j also follows c_i. Generally, three factors affect the visiting probability of a customer to a venue: the distance between the customer and the venue, the loyalty of the customer to the venue, and the influences from the customer's friends, such as their assessments or recommendations. $p_d(c_i, v_j)$ measures the impact of the location with distance $d(c_i, v_j)$. The average times that customer c_i, visits venue v_j per month is denoted as $t(c_i, v_j)$, and the corresponding value of loyalty is denoted as $l(t(c_i, v_j))$. The loyalty can be seen as the visiting probability of the customer if he/she is invited to the marketing event regarding his/her history visiting frequency. The impact of c_i's friends on venue v_j is denoted as $p_f(c_i, v_j)$. Finally, a solution (i.e., a team of invited participants) of

Table 1
Notations.

Symbol	Description
c_i	a participant, $c_i \in Set_c$
v_i	a venue, $v_i \in Set_v$
t_i	a tag, $t_i \in Set_t$
N_{SCALE}	scale of the event
$N_f(c_i)$	number of c_i's friends
$d(c_i, v_j)$	the distance between customer c_i and venue v_j
$v(c_i, v_j)$	whether c_i has visited v_j
$t(c_i, v_j)$	the average times that customer c_i visits venue v_j per month
$f(c_i, v_j)$	number of visits of c_i's friend to v_j monthly
$p_d(c_i, v_j)$	visiting probability of c_i to v_j with distance $d(c_i, v_j)$
$p_f(c_i, v_j)$	recommendation index of c_i to v_j
$l(t)$	loyalty of a customer to a venue with number of visits t
S	a solution (i.e., a team of invited participants)

the proposed scheme is $S = \{s_1,\ s_2,\ \ldots, s_{N_c}\}$, where $s_i = 1$ if c_i should be invited, and $s_i = 0$ otherwise. Among all possible selections of participants, an available solution should satisfy all constraints, while the optimal solution is the one that minimizes the marketing cost from all available solutions.

3.2. Offline marketing constraints

A marketer should invite a participant team that can minimize the marketing cost (i.e., the selected participants are more likely to visit the venue than other customers so that the resources will not be wasted and more profits can be made). At the same time, both the scale constraint and item coverage constraint can be satisfied. We formulate the constraints below.

1) Scale constraint: the expected number of visiting participants should equal the planned scale of an event:

$$\sum_{i=1}^{N_c} (s_i \times p_{c_i}) = N_{SCALE}, \tag{1}$$

where p_{c_i} is the probability of c_i to visit the venue.
The reason for this is that if the expected number of visiting participants is larger than the planned scale of an event, the marketer must pay a much higher cost. If not, the performance will not be maximized.

2) Coverage constraint of an item: the venue tags should be covered by the tags of the participants:

$$\forall_{i=1}^{N_t} t_i,\ \ \sum_{j=1}^{N_c} \left[s_i \times p_{c_i} \times t_i\left(c_j\right) \right] \geq t_i(v_k). \tag{2}$$

This implies that all events should be propagated and advertised as soon as possible.

3) To save resources, the marketing cost of a solution should be minimized. The marketing cost minimizing constraint: let $MC(S)$ be the marketing cost of a solution S; then, the target of the proposed statistical inference approach is to minimize the value:

$$\min\,(MC(S)). \tag{3}$$

4. Quantification of marketing cost

This section discusses the method of LBSN raw data preprocessing, and the calculation method for three potential factors to quantify the visiting probability of customers to the venues. Then, according to the potential factors, this section provides four offline event marketing schemes and provides a detailed discussion of the scheme based on bidirectional loyalty and the location of the venue.

4.1. LBSN raw data preprocessing

The LBSN dataset may contain information about the preferences and habits of the customers. However, they cannot be directly used in the applications since they are usually out-of-order and unstructured. Therefore, it is necessary to process the raw data. The processed data should reflect the friendship and visiting frequencies of customers, the distance between a given customer and a given venue, and the item coverage.

The main information contained in the LBSNs is as follows.

a) The comments (made by the customers who visited a venue) and the location (the longitude and latitude) of each venue.

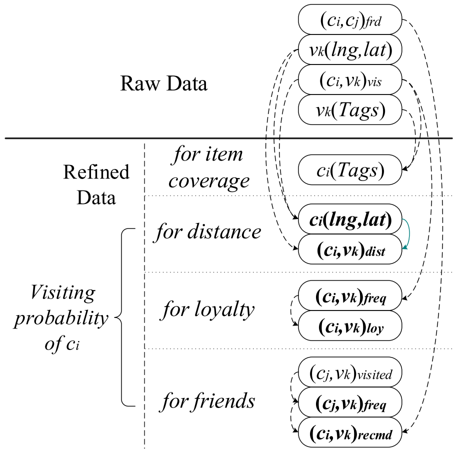

Fig. 1. Data preprocessing of the raw data.

b) Venue tags added by customers that show the products or services offered by the venue or their styles. The information hidden behind the tags is the products and services offered by the venue.
c) Check-in record. A particular location of a customer is known to an application as a result of him/her using a mobile website, text messaging, or a device-specific application. We assume that a customer who purchased products or received services from a venue has a check-in record of the venue.

Fig. 1 shows the correspondence between the source raw data and the refined data. Each venue has its own tag set, $v_k(Tags)$, which indicates the items it serves. $v_k(lng, lat)$ is the location of each venue. $(c_i, c_j)_{frd}$ is the one-sided friendship between customer c_i and c_j where c_i follows c_j. $(c_i, v_k)_{vis}$ represents that c_i checked-in at v_k a certain number of times in the past five months. The friendship among the customers can be easily obtained by the following lists of customers in their Twitter profiles, and the total number of visits divided by five months yields the visiting frequency. Next, we discuss how to determine the distance and the item coverage from the collected raw data.

To calculate the distance between the home of customer c_i and venue v_j, i.e., $d(c_i, v_j)$, we need to obtain both the locations of c_i and v_j. However, the raw data only contain the locations of the venues, which are the check-in places. According to [18], the geographical center of a customer's most frequent check-ins can be regarded as his/her home location. This can be obtained by improving the recursive grid search [1]. First, find all venues that the customer has visited, and for each of these venues, set it as the center of a circle with a radius of 50 km. Then, select the circle that contains the most check-in records from among these circles and make each visited venue within this selected circle the center of a new circle whose radius is 5 km. Again, choose the circle with the most check-ins and then set the radius to 0.5 km [30]. Finally, calculate the geographical center of the check-ins within this area and set it as the location of the customer's home. Given the latitudes and longitudes of the two places, $d(c_i, v_j)$ is then easily calculated as the distance between two points on the spherical surface, i.e., the surface of the Earth.

For meeting an item coverage constraint, the items that customers like should be obtained, which are not directly given in the raw data. Generally, the items that a customer prefers are reflected in the tags of the venues that the customer has visited. If a customer visits at least n (a threshold) venues with the same tag, then we can determine that this tag is reasonably liked by the customer. For instance, customer c_i has visited four venues, v_1, v_2, v_3 and v_4, which have $v_1(Tags) = 1,\ 1,\ 0,\ 0$, $v_2(Tags) = 1,\ 1,\ 1,\ 0$, $v_3(Tags) = 1,\ 1,\ 0,\ 1$ and $v_4(Tags) = 0,\ 1,\ 1,\ 0$ of items t_1, t_2, t_3 and t_4, respectively. If we set the threshold $n = 3$, then c_i is assumed to like item v_1 and v_2, i.e., $c_i(Tags) = 1,\ 1,\ 0,\ 0$. All tags of a venue are assumed to be covered by the invited participants.

4.2. Potential factors

There are many kinds of costs for a venue to hold an offline marketing event, such as constant expenses for the event, advertising expenses, and working hours. In this paper, we use the expected number of participants needed to meet the scale constraint to measure the marketing cost. The rationale is that venues need to spend more money and time, i.e., larger marketing cost, on invitation requirements with an increasing number of participants (other expenses can be seen as a constant). In this case, the number of selected participants can indicate the marketing cost. Hence, our goal is to select the smallest participant team with the largest probabilities of the customers so that the marketing cost can be minimized [3].

We introduce three potential factors to quantify the visiting probability of a particular customer to a certain venue: the *distance factor*, the *loyalty influence*, and the *recommendation index*. Note that all of these models are statistical models so that their parameters are learned from the real dataset collected from the location-based social networking sites (LBSNs) [30].

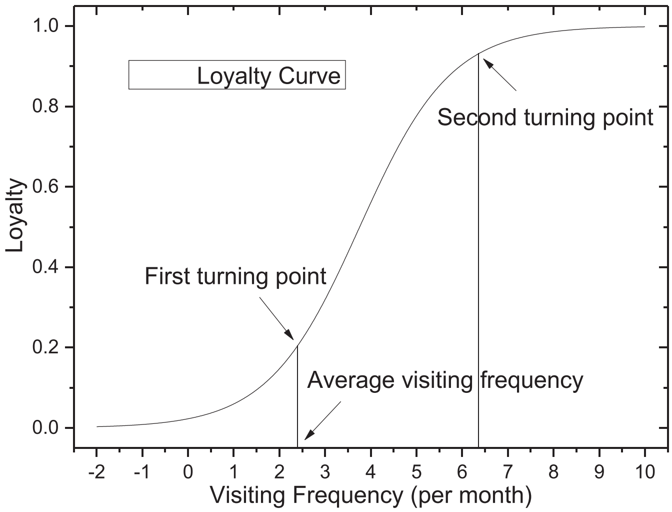

Fig. 2. Loyalty versus visiting frequency.

4.2.1. Distance factor

The probability of a customer visiting a venue normally decreases with the increasing distance between a customer's home and the location of the venue. This means that local customers have a higher probability of visiting a local event. According to Ref. [31], the relation between the probabilities and distances can be quantified as a power law; that is, $p_d(c_i,\ v_j) = \alpha_d \times d(c_i,\ v_j)^{\beta_d}$, where α_d and β_d are the parameters of the power distribution. Note that $d(c_i,\ v_j)$ needs to be discretized, or the number of cases with an exact $d(c_i,\ v_j)$ will be only one. We choose 1 km as an interval; that is, $d(c_i,\ v_j) \in [0,\ 1]$ is reset to $d(c_i,\ v_j) = 1$, and likewise, $d(c_i,\ v_j) \in (1,\ 2]$ is reset to $d(c_i,\ v_j) = 2$, and so on. The exact value of $p_d(c_i, v_j)$ can be learned by the ratio of the following two cases:

$$p_d\left(c_i,\ v_j\right) = \alpha_d \times d\left(c_i,\ v_j\right)^{\beta_d} = \frac{\#case\ with\ d(c_n,\ v_m) = d\left(c_i,\ v_j\right),\ v(c_n,\ v_m) = 1}{\#case\ with\ d(c_n,\ v_m) = d\left(c_i,\ v_j\right)}, \tag{4}$$

where $\#case\ with\ d(c_n,\ v_m) = d(c_i,\ v_j),\ v(c_i,\ v_j) = 1$ means the number of cases in which a customer, c_n, has visited a venue, v_m, and the distance between them is $d(c_i, v_j)$, i.e., $d(c_n,\ v_m) = d(c_i,\ v_j)$, and $\#case\ with\ d(c_n,\ v_m) = d(c_i,\ v_j)$ is the number of cases in which the distance between c_n and v_m equals $d(c_i, v_j)$.

4.2.2. Loyalty influence

If a customer c_i visits a venue v_k at a high frequency, then we say that this customer has considerable loyalty to the venue. Our target is c_i because he/she is more likely to visit the venue again if he/she is invited. However, if c_i is a friend of another customer c_j, who seldom or never visits v_k, i.e., c_j has a low loyalty to v_k, and then we say that c_j has a good potential to visit v_k because c_i may post some positive comments on the products or services of v_k in the social media, or give v_k a high rating, or directly recommend v_k to c_j. Therefore, customers who have many friends with considerable loyalties to a venue should be selected as participants to the venue.

We quantify the loyalty of a customer according to his/her frequency of visiting the venue per month, which should not be related proportionally to each other. Therefore, we choose the logistic function:

$$l\left(t\left(c_i,\ v_j\right)\right) = \frac{1}{1 + e^{-\left(t\left(c_i,\ v_j\right) - \varepsilon\right)}}, \tag{5}$$

where $t(c_i, v_j)$ is the average number of times that customer c_i visits venue v_j per month. Fig. 2 shows the loyalty versus the number of visits per month. When a customer visits a venue fewer than 2.439 times monthly, the loyalty increases slowly because these frequencies are not high enough to demonstrate that he/she is loyal to the venue. When the frequency is over 2.439 times a month, the loyalty increases considerably faster. The reason we choose 2.439 times per month as a turning point is that it is the average visiting frequency of a customer to a venue in our collected dataset. It may vary in different situations, but we consider the average number of visits as a good choice. Because the loyalty has to be kept below 1, the growth rate decreases again at the other turning point, where the frequency is 6.39 times a month, which is sufficient to determine that he/she has considerable loyalty to a venue. In our dataset, the number of customers who visit a certain venue over 6.39 times monthly only represent 0.31% of the customers. The exact value of parameter ε should be $\varepsilon = 3.756$. The proof is given below:

Proof. Consider the standard logistic function with parameter $\varepsilon = 0$, which yields $f(x) = 1/1 + e^{-x}$. Its derivative is $df(x)/dx = f(x)(1 - f(x))$. The function has the property of $1 - f(x) = f(-x)$. The two turning points are the locations where the second derivative of the function are the largest and the smallest values. Based on the above mathematical properties, we can easily calculate the second derivative as: $d(df(x)/dx)/dx = f(x)f(-x)f(-x) - f(x)f(x)f(-x)$. The result shows that the first turning point is $x = -1.317$, and the second is $x = 1.317$. The parameter ε actually represents the midpoint of the logistic function. If the first turning point is 2.439, then ε should be $2.439 - (-1.317) = 3.756$. Similarly, the second turning point can also be easily calculated as 6.39. □

The loyalty of a customer is simply regarded as the probability of the customer visiting the venue if invited based on his/her historical visiting records. We set a loyalty inferior threshold, l^{th}, to classify how loyal a customer is to a venue. If the loyalty of a customer satisfies

$$l\left(t\left(c_i,\ v_j\right)\right) \geq l^{th}, \tag{6}$$

then we say that this customer is loyal to the venue, i.e., he/she is likely to visit the venue. Note that we also need to discretize $t(c_i, v_j)$ for the same reason and here, once per month is the interval. For $t(c_i, v_j) \in (0, 1]$, it will be reset to $t(c_i, v_j) \in (0, 1]$, and so on. Specifically, if a customer never visits a venue, we have $t(c_i,\ v_j) = 0$.

4.2.3. Recommendation index

In addition to a customer's loyalty to a venue, the probability of his/her visiting the venue is also adjusted by his/her friends' loyalties. For example, if a customer has a friend who is very loyal to a certain venue, then this friend may post some positive information about the products or services of the venue in social media, which may be read by the customer, or the friend may give the venue a high rating, or directly recommend the venue to the customer if they know each other [1,28]. We call this kind of potential influence the *Recommendation Index* (RI). We use a coefficient $\tau(t(c, v))$ to measure the degree of such an influence from RI, where $t(c, v)$ is the number of visits of the customer's friend per month. That is, considering the influences from the friends' loyalties, we calculate the probability that a customer c_i will visit venue v_j as:

$$p_f\left(c_i,\ v_j\right) = \sum_{k=1}^{N_f(c_i)} \tau\left(t\left(c_k^i,\ v_j\right)\right) \tag{7}$$

where c^i is a customer who is a friend of c_i. Additionally, we model the recommendation index in the form of a power law, which is $\tau(t(c,\ v)) = \alpha_\tau \times t(c,\ v)^{\beta_\tau}$, where α_τ and β_τ are the power distribution parameters. The value of parameter $\tau(t)$ is learned through the ratio of the following two cases:

$$\tau(t) = \alpha_\tau \times t^{\beta_\tau} = \frac{\#case\ with\ v\left(c_i,\ v_j\right) = 1,\ f\left(c_i,\ v_j\right) = t}{\#case\ with\ f\left(c_i,\ v_j\right) = t} \tag{8}$$

where $f(c_i,\ v_j)$ is the number of visits of c_i's friend, and $\#case\ with\ v(c_i,\ v_j) = 1,\ f(c_i,\ v_j) = t$ represents the number of cases in which a customer c_i has visited a venue v_j at least once, and c_i has a friend who has visited v_j t times, and $\#case\ with\ f(c_i,\ v_j) = t$ is the number of cases in which c_i has a friend whose number of visits to venue v_j equals t.

We explain these four factors by four toy examples as illustrated in Fig. 3. A dotted line connecting a venue and a customer denotes the distance between them, which is marked as d_i (i is the index of the customer), whereas a solid line between any two customers is their friendship relationship. The monthly number of visits of a customer to the venue are denoted as t_i. In Fig. 3(a), c_1 is closer to the venue than c_2, so c_1 has a higher probability of visiting and thus he/she will be invited. However, in Fig. 3(b), c_2 visits the venue more often than c_1 even he/she is further away, which indicates that c_2 favors the venue, so he/she is more likely to visit the venue again if invited. In this case, c_1 is not necessarily prioritized over c_2. Furthermore, in Fig. 3(c), although c_1 visits the venue frequently (which cannot guarantee he/she will come if invited), c_2 has more friends that really like the venue. In this case, c_2 can be encouraged by his/her friends to visit the venue (through recommendations or ratings), so c_2 may have a higher probability of visiting than c_1. Fig. 3(d) shows the item coverage constraint. The venue serves three types of food, namely, seafood, chicken, and steak. Given that c_1 likes seafood, c_2 likes chicken and steak, and c_3 likes steak, who should be invited given the participant quota of two? Although c_2 and c_3 both have larger visiting frequencies, $S = \{c_1,\ c_2\}$ seems to be a better choice because all the products of the venue can be consumed.

4.3. Marketing cost calculation

Based on the above three potential factors, we provide four offline event marketing strategies. These strategies are *Location Influences* (LOCI), *Loyalty Influences* (LOYI), *Unidirectional Friends Local Influences* (UFLI), and *Bidirectional Friends Local Influences* (BFLI). The quantification of a marketing cost by using each of these strategies will be given below.

4.3.1. Location influences

This strategy 1, based on location influences, is concerned only with the influences of the home locations of the customers. First, the visiting probabilities of the customers with different distances to a venue are learned from the dataset,

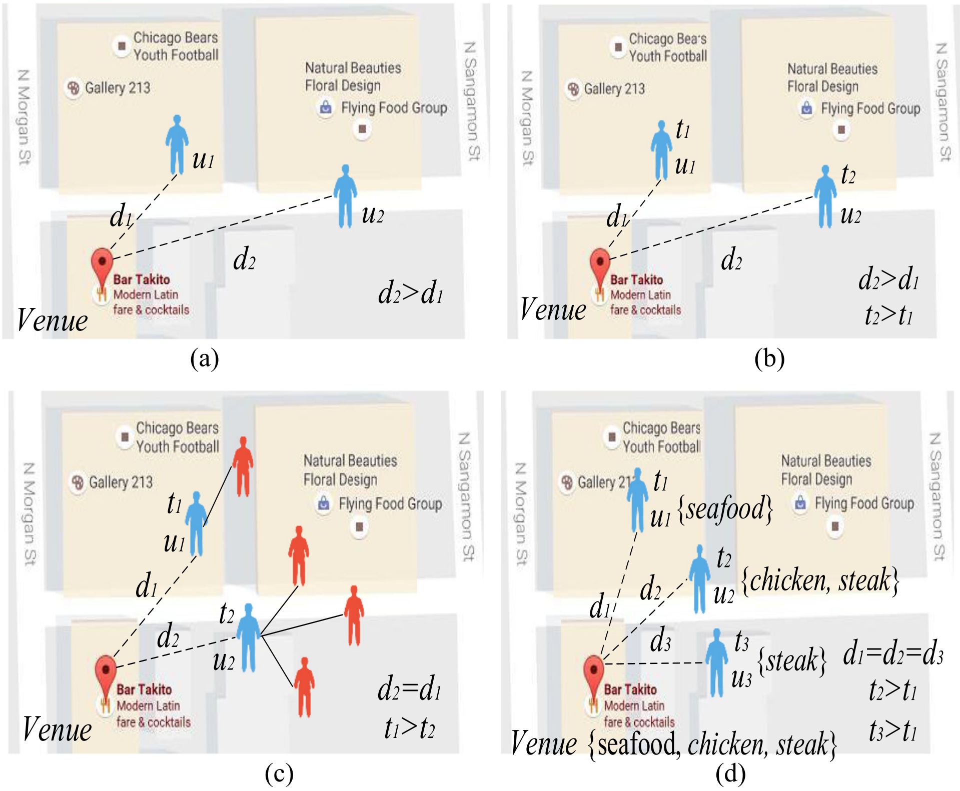

Fig. 3. Illustrations of three potential factors: (a) distance factor, (b) loyalty influence, (c) recommendation index, and (d) the item coverage constraint.

and then the probabilities are sorted in decreasing order. Finally, the top highest probabilities are selected until the expected number of visiting customers equals the planned event scale,

$$MC_{LOCI}(S) = \sum_{i=1}^{N_c} s_i, \quad \text{s.t. } \sum_{i=1}^{N_c} s_i \times p_d(c_i,\ v_j) = N_{SCALE} \quad \text{where } v_j \text{ is a venue.} \tag{9}$$

4.3.2. Loyalty influences

Strategy 2, based on location influences, focuses only on the influences of a customer's friends and his/her own loyalty to the venue. That is, with the influences of friends and loyalty, the customer who has a large probability to the venue will be selected as a participant. Note there may exit such a case: customer c_i has considerable loyalty to venue v_j, and simultaneously, he/she is also very loyal to another venue v_k; then, v_j has to compete with v_k for c_i if v_j has similar tags with v_k, i.e., they provide similar products or services. $\nabla[l(t(c_i, v_k))]$ in the formula equals $l(t(c_i, v_k))$ when v_j and v_k have similar tags, while it is 0 if they do not. The probability that v_j can win over v_k is $l(t(c_i,\ v_j))/\{l(t(c_i,\ v_j)) + \nabla[l(t(c_i,\ v_k))]\}$. Therefore, the marketing cost can be obtained by Eq. (10) as follows:

$$MC_{LOYI}(S) = \sum_{i=1}^{N_c} s_i, \quad \text{s.t. } \sum_{i=1}^{N_c} s_i \times \left[p_f(c_i,\ v_j) + \frac{l(t(c_i,\ v_j))^2}{l(t(c_i,\ v_j)) + \nabla[l(t(c_i,\ v_k))]} \right] = N_{SCALE} \tag{10}$$

4.3.3. Unidirectional loyalty of local influences

In practice, the probabilities of customers visiting are affected by both location and loyalty, which are combined into strategy 3. Compared with strategies 1 and 2, the calculated probabilities of visiting of strategy 3 on unidirectional loyalty

Table 2
Data statistics.

N_c	2,321
N_v	5,596
N_t	20
# total cases	12,988,316
# cases with $v(c_i,\ v_j) = 1$	82,428
average $d(c_i, v_j)$	2.32 km
Average $f(c_i, v_j)$	6.58
# cases with $v(c_i,\ v_j) = 0$	5,911,088
average $d(c_i, v_j)$	4.21 km
average # customers of a venue	2.95
average # venues of a customer	7.10
average # items of a venue	0.66
average # items of a customer	2.63
average # friends of a customer	19.51

local influences are larger, leading to fewer actual participants . Eq. (11) is the marketing cost of this strategy. Note that the case discussed in strategy 2 still exists in strategy 3. In fact, it is better if we take into account the loyalties of the customers to the competition venues, i.e., distrust information; that is, if a customer is very loyal to several competitive venues at the same time, then this customer should be excluded, which we will discuss in strategy 4.

$$MC_{UFLI}(S) = \sum_{i=1}^{N_c} s_i,$$
$$\sum_{i=1}^{N_c} s_j \times [p_d(c_i,\ v_j) + p_f(c_i,\ v_j) + \frac{l(t(c_i,\ v_j))^2}{l(t(c_i,\ v_j)) + \nabla[l(t(c_i,\ v_k))]}] = N_{SCALE}. \tag{11}$$

4.3.4. Bidirectional loyalty of local influences

Considering the competitive relationship among the venues, we introduce strategy 4 that considers the bidirectional loyalty local influences. If a customer is chosen to be invited by a host venue using strategy 3, then we search for the visited venues of the customer to which he/she is very loyal and match the tags between the host venue and the visited venues one by one. If any one of the visited venues is quite similar to the host venue (which indicates that there is a competitive relationship between them), then this customer's invitation will be canceled. From the data statistics in Table 2, we can see that the average frequency of venues visited by a customer is 7.10. This means that the comparisons among the venues will not be complicated and time-consuming. After eliminating those participants, the marketing cost of strategy 4 is obtained by Eq. (12):

$$MC_{BFLI}(S) = \sum_{i=1}^{N_c} s_i,$$
$$\sum_{i=1}^{N_c} s_j \times \{p_d(c_i,\ v_j) + p_f(c_i,\ v_j) + l(t(c_i,\ v_j))\} = N_{SCALE}. \tag{12}$$

4.4. Marketing cost minimization

Regardless of which quantitative model is adopted, a marking cost can be formalized as a multiple-objective optimization problem below:

$$\begin{cases} \min(MC(S)) \\ \sum_{i=1}^{N_c} (s_i \times p_{c_i}) = N_{SCALE} \\ \forall_{i=1}^{N_t} t_i,\ \sum_{j=1}^{N_c} \left[s_i \times p_{c_i} \times t_i\left(u_j\right)\right] \geq t_i(v_k) \end{cases}. \tag{13}$$

We present an algorithm that can quickly minimize the marketing cost with strategy 4, given as Algorithm 1. If a venue, called a host venue, wants to hold an event, the set of its associated customers will be searched for a participant team of members with higher visiting probabilities than other customers. For each customer, first, the following three visiting probabilities will be calculated:

1) Location-based visiting probability. This is determined by the distance between a customer's home address and a venue.
2) Loyalty-based visiting probability. This is determined by the visiting frequency (per month) of a customer to a host venue.

3) Friend-based visiting probability. This is determined by the number of friends a customer has and how loyal his/her friends are to the host venue.

Then, the historical venues that the customer has visited are identified, and the tags of the host venue are matched with the tags of the historical venues one by one. For the historical venue denoted as v_c, which has κ tags that are the same as the host venue indicating that they serve the same products or services, the loyalty of the customer to v_j will be computed, and if it satisfies Eq. (6), then this customer will be excluded in advance because the host venue has to compete for this customer with v_c. Finally, for the remaining customers, three kinds of visiting probabilities will be added up, and they will be sorted in descending order of the sum of the visiting probabilities. The customers with higher probabilities of visiting will be selected from the top to bottom as participants to the event until the expected number of visiting customers equals the planned event scale.

For example, consider customers set, Set_C; universal venue set, Set_V; expected visiting customer, $N_{EXP}=0$; participants team size, $S_p=0$ and a host venue v_h. For each customer c_i in Set_c, the system first calculates the location-based visiting probability of c_i, the loyalty-based visiting probability of c_i and the friend-based visiting probability of c_i. Second, for each historical venue v_i^h that c_i has visited, if venue v_i^h has the same κ tags as v_h, the system calculates the loyalty of customer c_i to venue v_i^h. If customer c_i is loyal to the venue, customer c_i is removed from Set_c. Third, the system adds up the three kinds of visiting probabilities of c_i, denoted as p_{c_i}, and sorts Set_c in the descending order of p_c. Fourth, the host v_h selects the customers following a sequence until the expected number of visiting participants equals the planned scale of the event. Thus, the selected customer set is the solution for the host venue.

The time complexity of the implementation of the four strategies is $nlog_2n$, which is superior to the previous work with NP-hard complexity, particularly for the very large size of Set_c. The time complexity of the algorithm is analyzed below.

Proof. Before starting the algorithm, we need to preprocess the dataset. Review the check-in records for one round to obtain a tuple of $c_i, v_j, t(c_i, v_j)$, indicating who (c_i) visited where (v_j) and $t(c_i, v_j)$ times per month. Next, quicksort (or another sort algorithm) the refined tuples in ascending order of the ID of customers (0 to N_c-1), then use a two-dimensional array to store the sorted tuples. Then, we can quickly identify a certain customer's check-in records through the index of the above two-dimensional array, i.e., identifying the check-ins of a customer is $O(1)$. Likewise, the social relationships of the customers are stored in a two-dimensional array, so that finding the friends of a customer is also $O(1)$. Assume that there are N_{ci} check-ins thus far such that the time complexity is $O(N_{ci}+N_c+N_{ci}log_2N_{ci}) \rightarrow O(N_{ci}log_2N_{ci})$. Note that the number of a customer's friends and the venues he/she visited are both constants (denoted as c_f and c_v, respectively); thus, the time needed for computing Eq. (7) (excluding $\tau(t)$) and the location of a customer is $O(1)$. Based on the above discussion, we can draw the conclusion that calculating the location-based and loyalty-based visiting probability are all $O(1)$. Here, we discuss the time needed for the friend-based visiting probability. For every t of the coefficient $\tau(t)$, finding the cases with $f(c, v)=t$ is $O(c_fc_vN_c) \rightarrow O(N_c)$, and finding the cases with $\nu(c,\ v)=1$ are $O(1)$. Additionally, t is limited, so that the time complexity for the friend-based visiting probability is $O(N_c)$. However, the calculation of $\tau(t)$ can be performed in the preprocessing phase, so the fifth line of Algorithm 1 is $O(1)$. For the 15*th* line, we can also use a quicksort. The inner for-loop of the first for-loop is $O(1)$, and the second for-loop is $O(N_c)$; thus, according to the rules of time complexity calculation, the final time complexity of Algorithm 1 is $N_{ci}log_2N_{ci} \rightarrow nlog_2n$. □

5. Experiments

The experimental evaluation of the proposed scheme is provided in this section. The real-world dataset extracted from LBSNs is used to evaluate the effectiveness of the statistical inference scheme. Our target is to maintain an optimal participant set to minimize the marketing cost among the collected data while fulfilling the constraints. We use a Core2 Duo 2.4 GHz laptop with Windows 10 and 2 GB memory to record the actual time needed for the implementation of the four strategies.

5.1. Datasets

As one of the most popular location-based services, Foursquare accumulated over 55 million registered members worldwide as of October 2014. The dataset used in this paper is obtained from Foursquare by using the publicly available Foursquare APIs, which grant access to the company's database of locations as well as information on venue check-ins. In particular, the dataset contains 151589 check-in records made by $N_c=2321$ Foursquare users in $N_\nu=5596$ venues, spanning a period of approximately five months from August 2010 to January 2011. The selected Foursquare users and venues are located in an approximately square area due to their relatively higher density, whose longitudes and latitudes of the four corners are (1.2431, 103.8275), (1.3879, 103.9995), (1.4650, 103.8154), and (1.3382, 103.6471). A customer is considered only if his/her home location is within the above square area and has visited at least one venue in the past five months. Customers' friendships are obtained from their Twitter profiles. Note that we can only access those check-ins that users explicitly choose to share on Foursquare, although users have the possibility to set this option as the default for all their check-ins, so the visiting probability of a certain customer calculated within this scheme may differ from the actual probability.

To meet the item coverage constraint, each venue should be tagged with some items. We choose tags (items) that can clearly distinguish the classifications of the products or services a venue offers instead of those that describe the venue, such

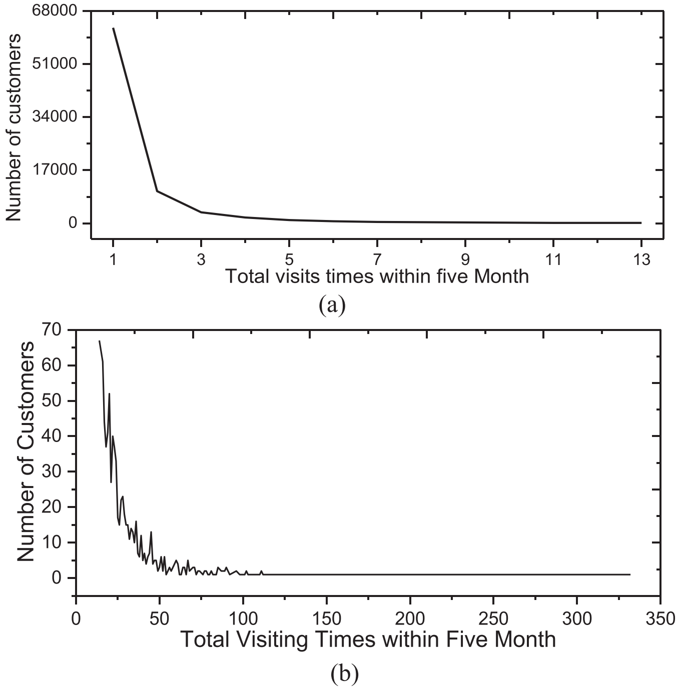

Fig. 4. Distribution of customers' number of visits.

as seafood and burger, but not coffee or tea because these are commonly offered by many venues even if their products or services are quite different. However, the tags that are too specialized should be discarded for simplicity, so only the tags that are owned by at least five venues will be chosen. The filtered tag set is Set_t={sandwich, bakery, burger, pizza, bistro, dessert, bbq, beer, seafood, chicken, superstore, steak, gym, fruits, noodles, cheese, tapas, pet, salad, and ice cream}. The tags of venues added by the visitors not in Set_t are removed.

Table 2 presents the dataset statistics. The total possible visiting cases are 12988316, and 82428 cases have $v(c_i,\ v_j) = 1$, making the average visiting probability $82,428/12,988,316 = 0.006$. Among the visited cases, the average $d(c_i, v_j)$ is 2.32 km, and the friends' average number of visits is 6.58. For the unvisited cases, the average $d(c_i, v_j)$ is 4.21 km. Each venue has 2.95 visitors and serves an average of 0.66 items, whereas each customer visits 7.10 venues, likes 2.63 items, and follows 19.51 users (friends) on average.

Note that the numbers of user check-ins vary significantly. In particular, the probability distribution of the number of visits exhibits a heavy tail, with approximately 50% of users having fewer than 10 check-ins. A similar pattern for the number of check-ins in each place arises, with only 10% of places having more than 10 check-ins.

Additionally, the total number of visits over the past five months exhibits a heavy tail, with approximately 50% of the users having fewer than 10 check-ins. A similar pattern on the number of check-ins for each place arises with only 10% of places having more than 10 check-ins. This phenomenon is illustrated in Figs. 4 and 5.

5.2. Curve fitting

As mentioned before, the distance factor and recommendation index are learned from the training through a reasoning ratio of two different kinds of cases filtered by specific rules. The learned results are listed in Table 3. The obtained empirical probabilities of $p_d(c, v)$ show an increasing tendency with a decreasing distance. This validates the existence of the distance factor. The recommendation index, $\tau(t)$, of a customer generally increases with the number of visits of his/her friends, which is due to their positive influences. The fitting results of $p_d(c, v)$ and $\tau(t)$ are shown in Figs. 6 and 7, respectively.

5.3. Performance comparisons

Four strategies are used to minimize four different marketing costs, $MC(S)$. Each strategy results in a solution for a given venue. The performances of the four strategies are compared in this section. The first comparison is the value of $MC(S)$, which tells the marketer who and how many customers should be invited to meet the scale goal. The second comparison is the runtime of the algorithm, which is an important concern in any optimization problem [10]. The third comparison is item coverage, which determines whether the items of a venue are liked by all of the invited customers.

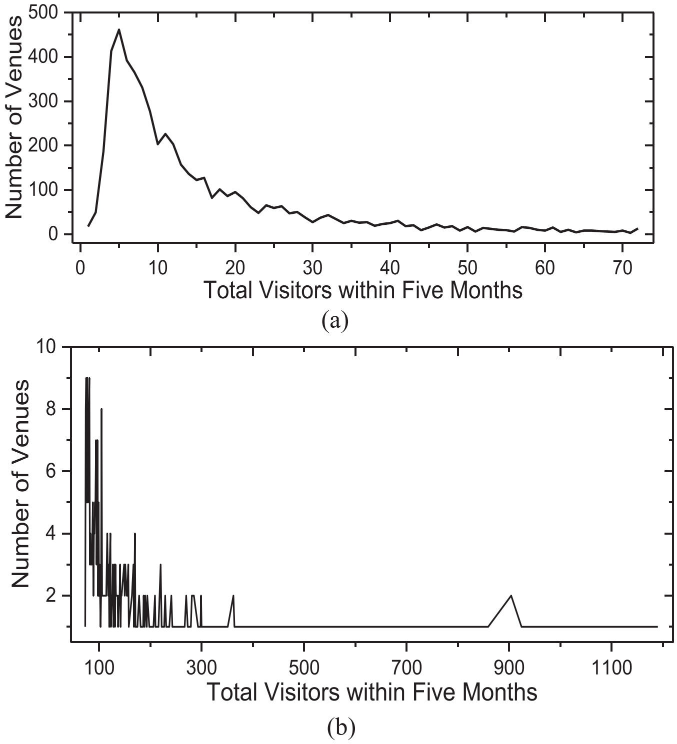

Fig. 5. Distribution of the venues' visitors.

Table 3
Statistical probabilities and fitted parameters.

d	$p_{d(c,v)}$	t	$p_{\tau(t)}$
1	**0.0175** (710/40,509)	1	**0.0296** (39,537/1,335,742)
2	**0.0057** (344/60,556)	5	**0.0306** (81/2629)
3	**0.0051** (238/46,621)	9	**0.0441** (28/632)
4	**0.0044** (178/40,323)	13	**0.0201** (4/177)
5	**0.0034** (117/34,044)	17	**0.0787** (9/114)
6	**0.0036** (79/22,184)	21	**0.0708** (6/83)
7	**0.0029** (48/16,350)	25	**0.0967** (6/54)
…	…	…	…
	$\alpha_d = 0.01686$ $\beta_d = -0.97462$		$\alpha_\tau = 0.00992$ $\beta_\tau = 0.61627$

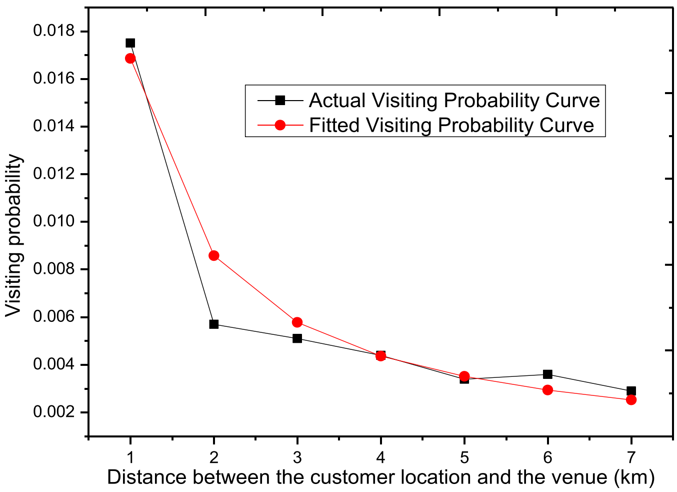

Fig. 6. Visiting probability versus distance.

Algorithm 1 Marketing cost minimization.

1) Input: Set_c a data set for location-based social networks.
Output: S_p –the sets of customers with the minimal cost.
Initial: Host venue, v_h; customers set, Set_c; expected visiting customer, $N_{EXP} = 0$; participants team size, $S_p = 0$.
2) **For** every customer, c_i, in Set_c, **Do**
3) Calculate the location-based visiting prob. of c_i.
4) Calculate the loyalty-based visiting prob. of c_i.
5) Calculate the friend-based visiting prob. of c_i.
6) **For** each historical venue v_i^h that c_i has visited, **Do**
7) **If** v_i^h has κ same tags with v_h, **Do**
8) Calculate the loyalty of c_i to v_i^h, $l(t(c_i, v_j))$
9) **If** $l(t(c_i, v_j)) \geq l^{th}$, **then**
10) Remove c_i from Set_c
11) **End if**
12) **End for**
13) Sum the three kinds of visiting probabilities of c_i, denoted as p_{c_i}
14) **End for**
15) Sort Set_c in the descending order of p_c
16) **For** every customer, c_i, in Set_c, **Do**
17) $N_{EXP} = N_{EXP} + p_{c_i}$, $S_p = S_p + 1$
18) Output c_i as one invited customer
19) **If** $N_{EXP} \geq N_{SCALE}$, **then**
20) Exit for
21) **End if**
22) **End For**
23) Output S_p

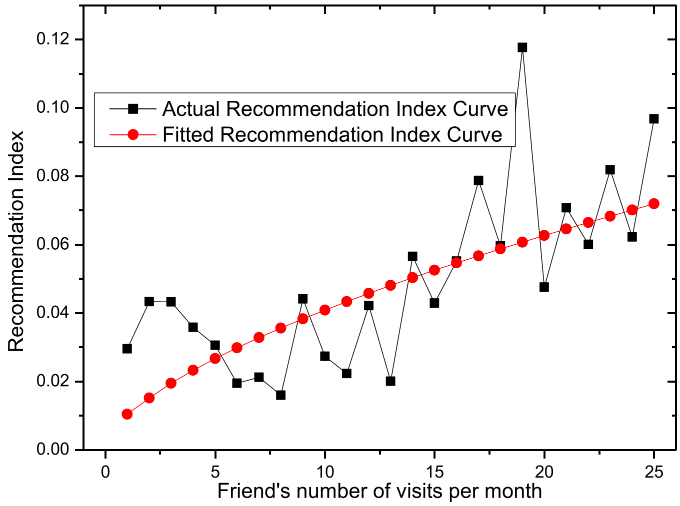

Fig. 7. Recommendation versus friend's number of visits per month.

For the comparisons among the different strategies [15], the default scale of an offline marketing event is set as $N_{SCALE} = 10$, which is typical for free trial advertising by a venue [30]. In addition, N_{SCALE} is set at 3, 8, 10, and 13 to determine its impact on each strategy.

5.3.1. Average marketing cost

Figs. 8–11 illustrate the marketing cost distribution of strategies 1 to 4. Strategy 1 allows most venues to reach the scale goal by inviting 600–1500 customers. In particular, two extremely large cases are not shown in this figure, which include 168 venues that need 593 participants and 3278 venues that need 2321 participants (i.e., all customers). They are also the smallest and largest sizes of the participant teams. Note that for those venues that need the whole customer set, they may not meet the scale constraint since they are in an area with too low of a density of customers. The range of the size of the participant team obtained by strategy 2 is much narrower and smaller than strategy 1, which distributes intensively from 500 to 630. However, strategy 3 and strategy 4 have even shorter participant teams, which are almost completely concentrated in the range of 300 to 370. Note that none of the venues needs 2321 participants, which means that all venues can actually reach their scale goals.

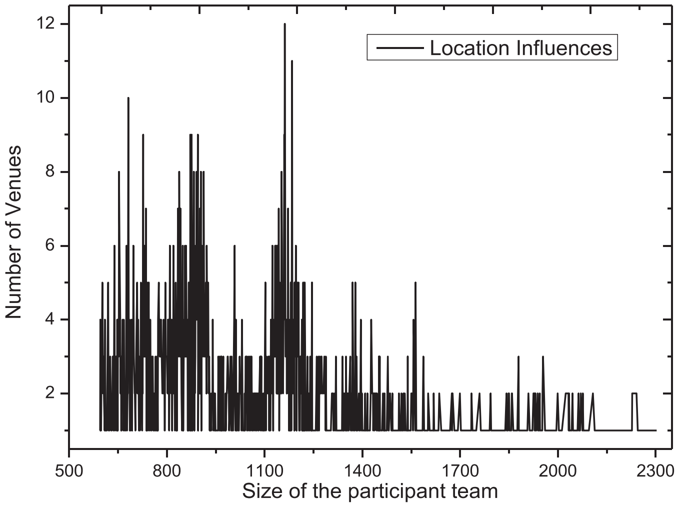

Fig. 8. Marketing cost distribution of strategy 1.

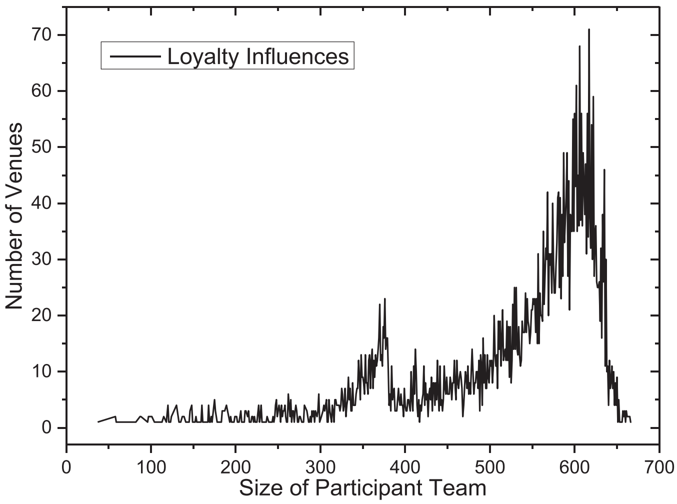

Fig. 9. Marketing cost distribution of strategy 2.

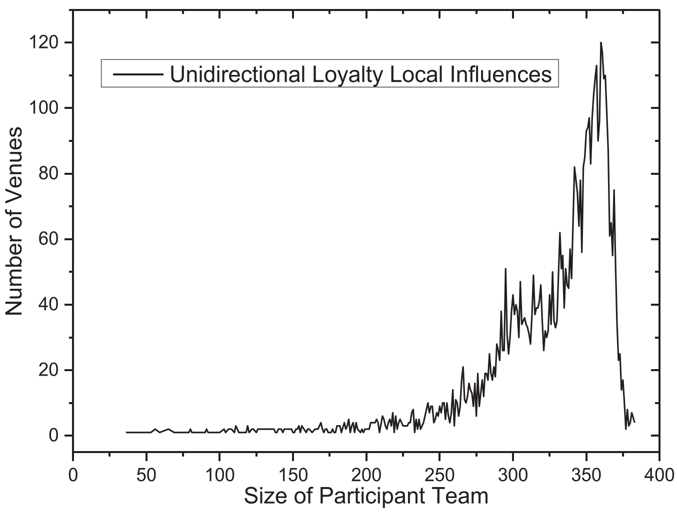

Fig. 10. Marketing cost distribution of strategy 3 based on unidirectional loyalty local influences.

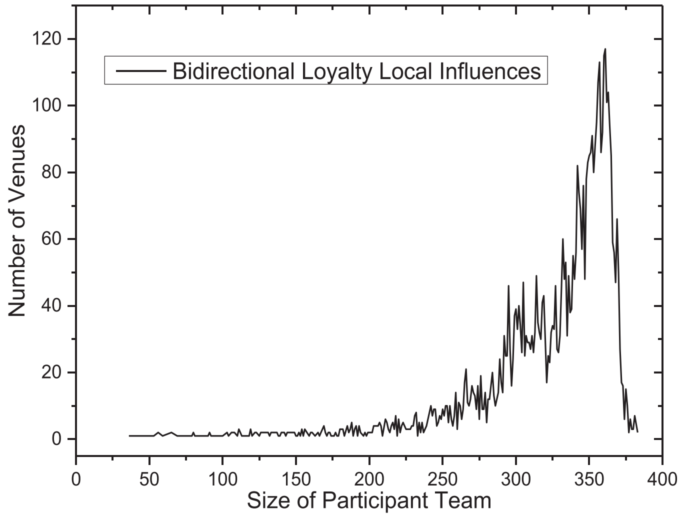

Fig. 11. Marketing cost distribution of strategy 4 on bidirectional loyalty local influences.

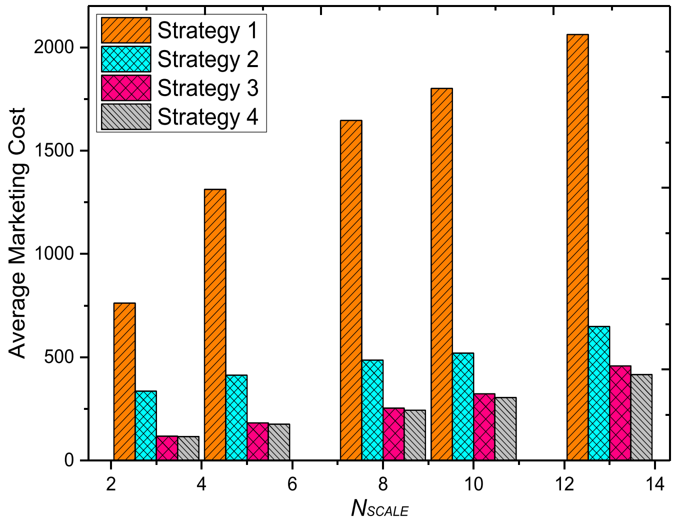

Fig. 12. Average marketing cost for strategies 1–4 in different settings of N_{SCALE}.

The average marketing costs of the four strategies are shown in Fig. 12. Strategy 1 always has a much larger participant team because it only considers the distance factor. Strategy 2 has the second largest team, which only considers the loyalty factor (i.e., the preferences of customers). We can draw the conclusion that the loyalty influence plays a more important role than the distance factor in our scheme. Strategy 3 and Strategy 4 have smaller teams for taking into account the above two factors, and they behave similarly. The reason for this is that our dataset has insufficient tags to clearly distinguish the products or services of different venues, and many venues have no tags. Therefore, the competition relationship cannot be easily identified.

5.3.2. Average time cost

Due to the efficiency of simple programming methods, all of the methods can obtain a solution within a short period of time for each venue. Strategy 1 is the simplest. Due to the pretreatment of the dataset, the required runtime for each venue to obtain a solution is readily available. Strategies 2–4 requires tens of seconds, as shown in Fig. 13. We can conclude that generally, more time is needed for considering more factors. However, we also find that the time cost remains nearly constant with an increasing N_{SCALE}. The reason for this is that N_{SCALE} only affects lines 15 to 22 of the algorithm that selects enough participant to reach the scale goal from a sorted array storing the expected visiting probabilities of all customers. The time for going through an array whose length is equal to (or shorter than) N_c one time is negligible. Moreover, we can see that the procedure added in strategy 4 is not time-consuming.

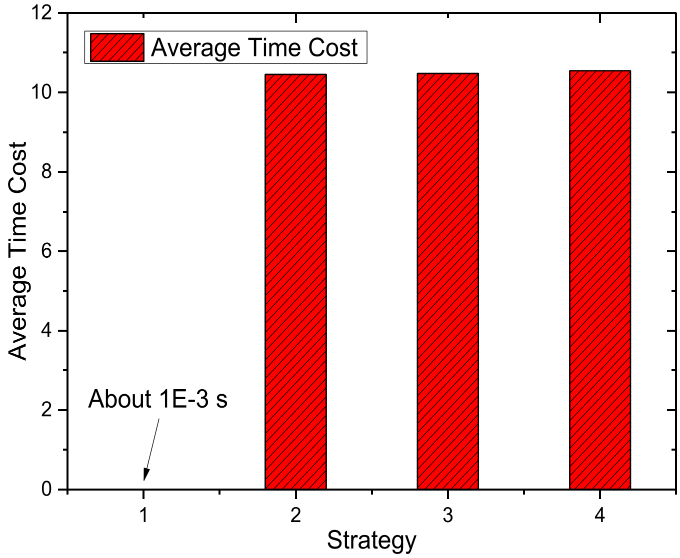

Fig. 13. Average time costs for different strategies.

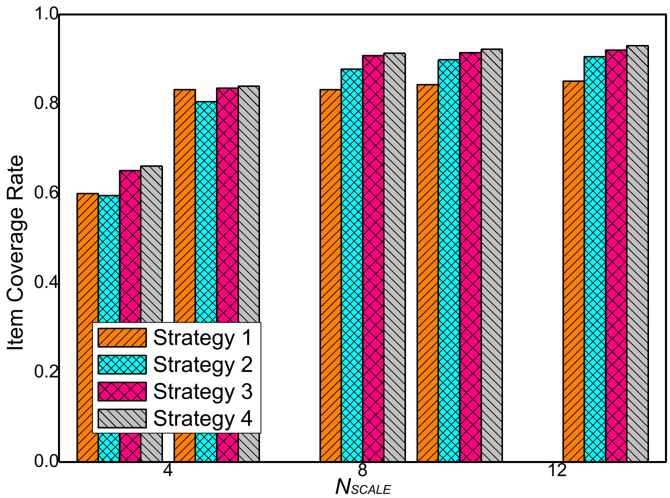

Fig. 14. Average item coverage rate for Strategies 1–4 in different settings of N_{SCALE}.

5.3.3. Average item coverage rate

The item coverage rate denoted as CR_{item} is used to measure the degree to which the expected number of items are covered by the participants:

$$CR_{item} = \frac{\sum_{k=0}^{N_c} \left(p_{c_k} \times \left| \left\{ t_i | s_k = 1,\ CR_{t_i} < 1,\ t_i(u_k) = 1,\ t_i\left(v_j\right) = 1 \right\} \right| \right)}{\left| \left\{ t_i | t_i\left(v_j\right) = 1 \right\} \right|} \tag{14}$$

where $|X|$ is the size of the set X, and CR_{t_i} is the cover rate of item t_i. Since an expected coverage rate of a certain item should be less than 1 (otherwise it will affect the rate of other items), the condition $CR_{t_i} < 1$ is required. For the right side of the formula, the denominator is the number of items owned by venue v_j, and the numerator is the expected number of v_j's items liked by the invited customers.

Note that an empty tag set may exist for a venue if no user has ever added a tag to it, or the added tags are not in our filtered tag set, Set_t. This kind of venue is considered to have no competitive relationship with any other venues in strategy 4. Additionally, the empty tag set for a customer may exist for his/her check-ins because it fails to meet the rule in the tail of Section III-B to identify the items a customer likes. These two cases are excluded from calculating CR_{item}.

There is no doubt that all four strategies can achieve a higher coverage rate with increasing N_{SCALE} because more customers are invited, causing each item to have more chances to be covered, as shown in Fig. 14. Note that in our dataset, the average item coverage rate of strategy 2, strategy 3, and strategy 4 are comparable to strategy 1 even though it has a much

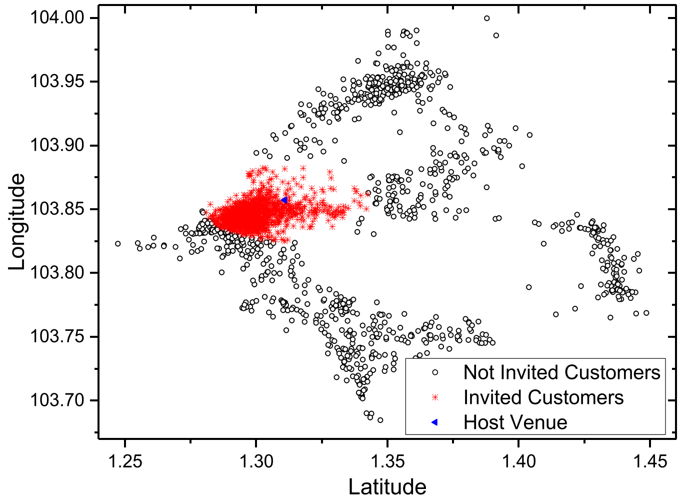

Fig. 15. Strategy 1: 1352 customers are invited.

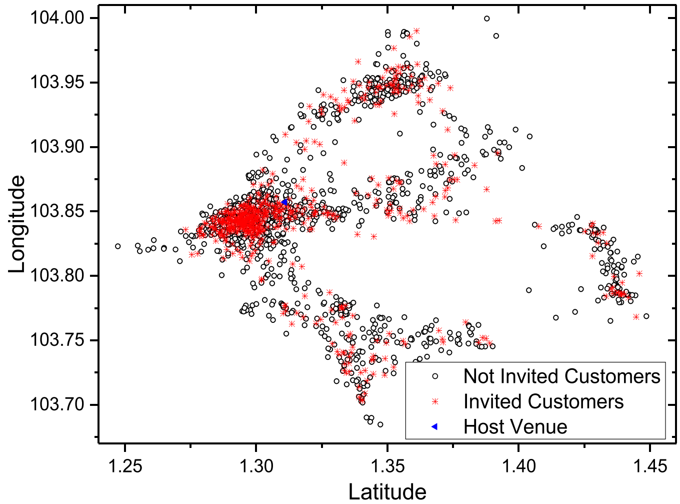

Fig. 16. Strategy 2: 596 customers are invited.

larger participant team, which demonstrates that considering the loyalty influence is helpful in improving the coverage rate. Strategy 3 and strategy 4 show great effectiveness, reaching as much as 92%–93%. However, none of them can guarantee that each item served by the venue is liked.

5.4. A case study

In this section, we apply our approach with the four strategies to a particular venue in our dataset as the case study of comparing their different results. The location of this venue, denoted as v_{4761}, is not in a too remote or too prosperous area (indicated by a blue triangle). Additionally, we set N_{SCALE} to 10. The solutions of the four strategies are clearly shown in Figs. 15 –18, where 1352, 596, 354, and 328 customers are invited to the offline marketing event, respectively.

Strategy 1 is concerned only about the distance factor. As a result, nearby customers have more probabilities to be selected, as depicted in Fig. 15. The local participants form a nearly elliptical area with the host venue situating in the center. In addition, the size of the participant team is very large, where 1353 customers are invited, and the expected visiting number is only $N_{SCALE} = 10$, i.e., the effectiveness is 0.0074 visitors per invitation.

In Fig. 16, we can see that participants do not necessarily gather together as a result of using strategy 2. The size of the participant team is much less than strategy 1.

Fig. 17 illustrates the distribution of participants with strategy 3. Compared to strategy 2, Fig. 17 shows a small location-oriented feature for adding the distance factor. Because there are insufficient tags of venues and customers, not many customers who are only loyal to v_{4761} are found. Therefore, the participant team of strategy 4 is quite similar to that of strategy

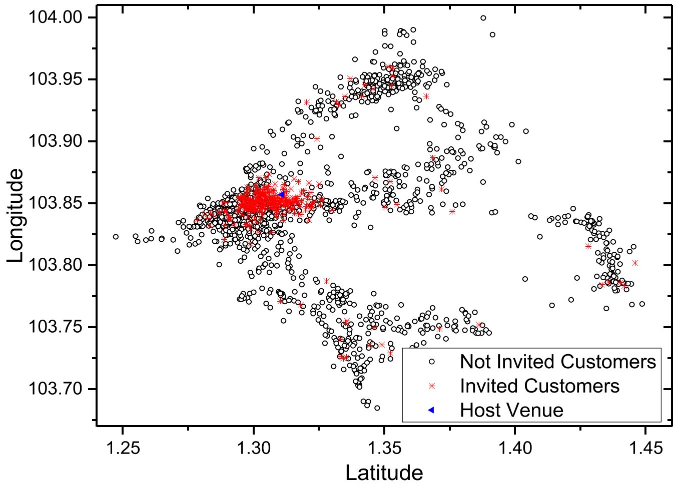

Fig. 17. Strategy 3: 354 customers are invited.

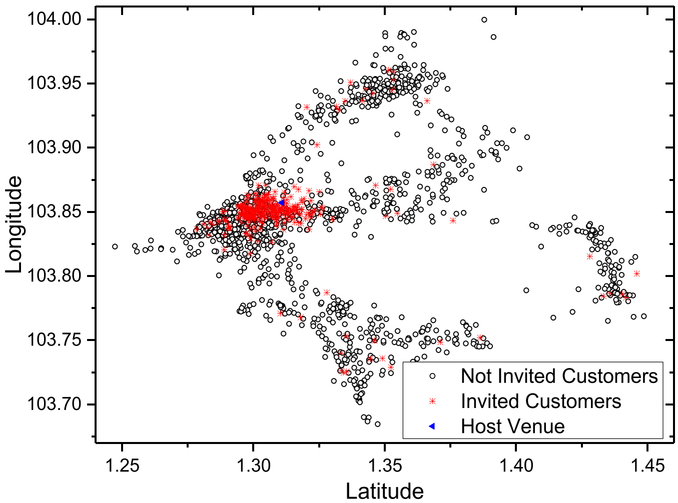

Fig. 18. Strategy 4: 328 customers are invited.

3, as shown in Fig. 18. However, the number of invited customers is less, which demonstrates that strategy 4 is more focused on being precise in selecting the participants.

Particularly, the venue v_{4761} serves five items: bbp, fruits, seafood, steak, and tapas. In particular, when using strategy 4 and N_{SCALE} is 10, the result shows these items are liked by 0.3209, 0.6458, 1.0, 1.0, and 0.8273 people, respectively, in all the participants of v_{4761}.

6. Conclusion

Offline event marketing has become increasingly popular. Marketers need support to improve the marketing effectiveness of sponsored offline events. In this paper, we have presented a statistical approach that is able to optimally invite customers with high visiting probabilities based on their historical check-in records. In particular, our approach takes into account three factors that affect the customers' probabilities of attending an event: *distance factor, loyalty influence*, and *recommendation index*. Given a friend set, scale, served items, and location, as well as the marketing cost, which is measured by how many customers are expected to be invited to meet the scale goal of a host venue, our approach can predict the marketing costs for an arbitrary customer set. Therefore, the participant team with the minimal marketing cost is selected from all the candidates.

To fully explore the influences of the three potential factors, we presented four strategies by considering their combinations. We conducted experiments on real-world raw data from Foursquare, a popular location-based social network (LBSN). By experiments, we validated our proposed algorithm and evaluated the four strategies in terms of marketing cost, runtime, and item coverage. The experimental results demonstrated the effectiveness of our approach and strategies. Our future work will focus on experimenting on more datasets and developing more strategies.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (61772554), and the Jiangsu High Technology Research Key Laboratoryfor Wireless Sensor Networks (NLBKF201804).

References

[1] L. Backstrom, E. Sun, C. Marlow, Find me if you can: improving geographical prediction with social and spatial proximity, in: Proceedings of the 19th international conference on World wide web, ACM, 2010, pp. 61–70.
[2] Z. Ding, X. Li, C. Jiang, M. Zhou, Objectives and state-of-the-art of location-based social network recommender systems, ACM Comput. Surv. (CSUR) 51 (1) (2018) 18.
[3] J. Jiang, B. An, Y. Jiang, D. Lin, Z. Bu, J. Cao, et al., Understanding crowdsourcing systems from a multiagent perspective and approach, ACM Trans. Autonom. Adapt. Syst. (TAAS) 13 (2) (2018) 8.
[4] M. Huang, A. Liu, M. Zhao, T. Wang, Multi working sets alternate covering scheme for continuous partial coverage in WSNs, Peer-to-Peer Netw. Appl. (2018), doi:10.1007/s12083-018-0647-z.
[5] M. Huang, A. Liu, N. Xiong, T. Wang, A.V. Vasilakos, A low-latency communication scheme for mobile wireless sensor control systems, IEEE Trans. Syst. Man Cybernet.-Syst. (2018), doi:10.1109/TSMC.2018.2833204.
[6] S. Kirkpatrick, Optimization by simulated annealing: quantitative studies,, J. Stat. Phys. 34 (5) (1984) 975–986.
[7] DH. Lee, P. Brusilovsky, How to measure information similarity in online social networks: a case study of Citeulike, Inform. Sci. 418 (2017) 46–60.
[8] T. Li, S. Tian, A. Liu, H. Liu, T. Pei, DDSV: optimizing delay and delivery ratio for multimedia big data collection in mobile sensing vehicles, IEEE Internet Things J. (2018), doi:10.1109/JIOT.2018.2847243.
[9] Z. Li, Y. Liu, M. Ma, A. Liu, X. Zhang, G. Luo, MSDG: a novel green data gathering scheme for wireless sensor networks, Comp. Netw. 142 (4) (2018) 223–239.
[10] Z. Li, Y. Liu, A. Liu, S. Wang, H. Liu, Minimizing convergecast time and energy consumption in green internet of things, IEEE Trans. Emerg. Topic. Comp. (2018), doi:10.1109/TETC.2018.2844282.
[11] X. Liu, Y. Liu, A. Liu, L. Yang, Defending on-off attacks using light probing messages in smart sensors for industrial communication systems, IEEE Trans. Indust. Inform. 14 (9) (2018) 3801–3811.
[12] X. Liu, W. Liu, Y. Liu, H. Song, A. Liu, X. Liu, A trust and priority based code updated approach to guarantee security for vehicles network, IEEE Access 6 (2018) 55780–55796.
[13] H. Ma, MR. Lyu, I. King, Learning to recommend with trust and distrust relationships, in: Proceedings of the third ACM conference on Recommender systems, ACM, 2009, pp. 189–196.
[14] GS. Njoo, KW. Hsu, WC. Peng, Distinguishing friends from strangers in location-based social networks using co-location, Pervas. Mob. Comp. 50 (2018) 114–123.
[15] A. Noulas, S. Scellato, C. Mascolo, M. Pontil, An empirical study of geographic user activity patterns in foursquare, ICWSM 11 11 (2011) 70–573.
[16] K. Ota, MS. Dao, V. Mezaris, FGB. De. Natale, Deep learning for mobile multimedia: a survey, ACM Trans. Multimedia Comp. Commun. Appl. (TOMM) 13 (3 s) (2017) 34.
[17] TJ. Ogundele, CY. Chow, JD. Zhang, SoCaST*: personalized event recommendations for event-based social networks: a multi-criteria decision making approach, IEEE Access 6 (2018) 27579–27592.
[18] K. Pelechrinis, P. Krishnamurthy, Socio-spatial affiliation networks, Comp. Commun. 73 (2016) 251–262.
[19] S. Peng, G. Wang, D. Xie, Social influence analysis in social networking big data: opportunities and challenges, IEEE Netw. 33 (1) (2017) 11–17.
[20] B. Schmitt, Experience marketing: Concepts, Frameworks and Consumer Insights, Now Publishers Inc, 2011.
[21] C. Shim, W. Kim, W. Heo, S. Yi, YD. Chung, Relationship recommender system in a business and employment-oriented social network, Inform. Sci. 433 (2018) 204–220.
[22] C. Shim, W. Kim, W. Heo, Nearest close friend search in geo-social networks, Inform. Sci. 423 (2018) 235–256.
[23] C. Song, Z. Qu, N. Blumm, AL. Barabási, Limits of predictability in human mobility, Science 327 (5968) (2010) 1018–1021.
[24] T. Wang, J. Zhou, A. Liu, M. Bhuiyan, G. Wang, W. Jia, Fog-based computing and storage offloading for data synchronization in IoT, IEEE Internet Things J. (2018), doi:10.1109/JIOT.2018.2875915.
[25] Z. Wang, J. Liao, Q. Cao, H. Qi, Z. Wang, Friendbook: a semantic-based friend recommendation system for social networks, IEEE Trans. Mob. Comp. 14 (3) (2015) 538–551.
[26] J. Xu, A. Liu, N. Xiong, T. Wang, Z. Zuo, Integrated collaborative filtering recommendation in social cyber-physical systems, Intern. J. Distribut. Sensor Netw. 13 (12) (2017) 1550147717749745, doi:10.1177/1550147717749745.
[27] W. Yang, G. Wang, M. Bhuiyan, K. Choo, Hypergraph partitioning for social networks based on information entropy modularity,, J. Netw. Comp. Appl. 86 (2017) 59–71.
[28] M. Ye, X. Liu, WC. Lee, Exploring social influence for recommendation: a generative model approach, in: Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval, ACM, 2012, pp. 671–680.
[29] M. Ye, P. Yin, W.C Lee, Location recommendation for location-based social networks, in: Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems, ACM, 2010, pp. 458–461.
[30] Z. Yu, D. Zhang, Z. Yu, D. Yang, Participant selection for offline event marketing leveraging location-based social networks, IEEE Trans. Syst. Man Cybernet. Syst. 45 (6) (2015) 853–864.
[31] Q. Yuan, G. Cong, Z. Ma, A. Sun, Time-aware point-of-interest recommendation, in: Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval, ACM, 2013, pp. 363–372.
[32] C. Zhang, R. Chen, L. Zhu, A. Liu, Y. Lin, F. Huang, Hierarchical information quadtree: efficient spatial temporal image search for multimedia stream, Multimedia Tools Appl. (2018), doi:10.1007/s11042-018-6284-y.
[33] Y. Zhuang, S. Fong, M. Yuan, Y. Sung, K. Cho, R. Wong, Location-based big data analytics for guessing the next foursquare check-ins, J. Supercomp. 73 (7) (2017) 3112–3127.