

# Maximum-Profit Advertising Strategy Using Crowdsensing Trajectory Data



LOU Kaihao<sup>1</sup>, YANG Yongjian<sup>1</sup>, YANG Funing<sup>1</sup>, ZHANG Xingliang<sup>2</sup>

(1. Jilin University, Changchun 130012, China;

2. China Mobile Group Jilin Co., Ltd., Changchun 130021, China)

**Abstract:** Out-door billboard advertising plays an important role in attracting potential customers. However, whether a customer can be attracted is influenced by many factors, such as the probability that he/she sees the billboard, the degree of his/her interest, and the detour distance for buying the product. Taking the above factors into account, we propose advertising strategies for selecting an effective set of billboards under the advertising budget to maximize commercial profit. By using the data collected by Mobile Crowdsensing (MCS), we extract potential customers' implicit information, such as their trajectories and preferences. We then study the billboard selection problem under two situations, where the advertiser may have only one or multiple products. When only one kind of product needs advertising, the billboard selection problem is formulated as the probabilistic set coverage problem. We propose two heuristic advertising strategies to greedily select advertising billboards, which achieves the expected maximum commercial profit with the lowest cost. When the advertiser has multiple products, we formulate the problem as searching for an optimal solution and adopt the simulated annealing algorithm to search for global optimum instead of local optimum. Extensive experiments based on three real-world data sets verify that our proposed advertising strategies can achieve the superior commercial profit compared with the state-of-the-art strategies.

DOI: 10.12142/ZTECOM.202102005

<https://kns.cnki.net/kcms/detail/34.1294.TN.20210629.1639.002.html>, published online June 30, 2021

Manuscript received: 2021-03-11

**Keywords:** billboard advertising; mobile Crowdsensing; probabilistic set coverage problem; simulated annealing; optimization problem

**Citation** (IEEE Format): K. H. Lou, Y. J. Yang, F. N. Yang, et al., "Maximum-profit advertising strategy using crowdsensing trajectory data," *ZTE Communications*, vol. 19, no. 2, pp. 29 - 43, Jun. 2021. doi: 10.12142/ZTECOM.202102005.

## 1 Introduction

Out-door billboards are one of the most effective tools for advertising. According to PQ Media<sup>[1]</sup>, global digital roadside billboard advertising industry grew by a large margin in 2017; specifically, digital roadside

billboard advertising sales have increased by 10% to a total amount of 3.2 billion dollars in the US. Compared with other advertising methods, the out-door billboard can easily make a deeper impression on potential customers, since it provides the strong visual impact, long placement duration and rich information content.

By advertising on out-door billboards, an advertiser can attract potential customers for his/her products. For some products or activities such as temporary promotion, potential customers may immediately decide whether to go to the shop to purchase products after seeing the advertisement. In this situation, once a potential customer is attracted by the advertisement on the billboard, he/she will purchase the relative prod-

This work is supported by Jilin Science and Technology Department Key Technology Project (20190304127YY), the National Natural Science Foundations of China (1772230, 61972450 and 62072209), Natural Science Foundations of Jilin Province (20190201022JC), National Science Key Lab Fund Project (61421010418), Innovation Capacity Building Project of Jilin Province Development and Reform Commission (2020C017-2), Changchun Science and Technology Development Project (18DY005), Key Laboratory of Defense Science and Technology Foundations (61421010418), and Jilin Province Young Talents Lifting Project (3D4196993421).

A conference version of the paper appeared in *Proceedings of SocialSec*<sup>[2]</sup>.

uct and the advertiser will obtain the commercial profit. However, due to the advertising budget constraint, an advertiser cannot advertise on all the available billboards. Hence, an advertiser should decide which billboards to do advertising to attract as many potential customers as possible, in order to maximize the commercial profit. Note that whether a billboard can attract a customer is determined by many factors, such as customers' mobility (whether they can see the billboard), customers' preferences (whether they are interested in the product), and the location to buy the product (the detour distance). More importantly, all the above information is privacy-sensitive, especially for the customers' trajectories and preferences, which greatly limits the billboard advertising research.

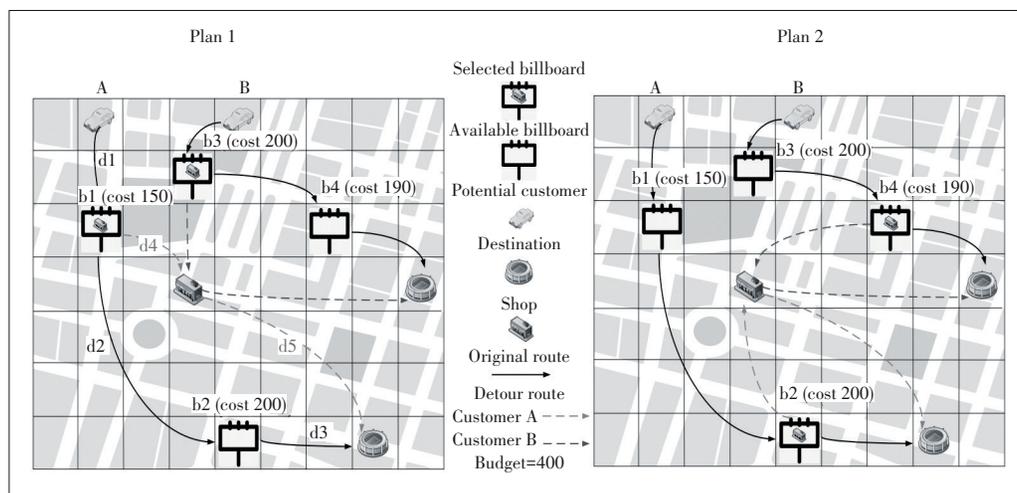
Most of the existing advertising strategies focus on what advertising content should be delivered and how to select locations for the static roadside billboards, but they do not jointly take potential customers' mobility, preferences and detour distance into consideration. In Ref. [3], NIGAM et al. decide the locations of billboards by using the data collected by radio frequency identification (RFID). In Refs. [4] and [5], the advertisers can select the billboard locations by using GPS and phone data. Besides, in Refs. [6] and [7], the advertising content on the billboard is determined by the potential customers' preferences or their detour distance. All these existing works do not jointly take potential customers' mobility, preferences and detour distance into consideration, which results in the advertiser failing to accurately quantify the commercial profit of the available billboards and thus prevents the corresponding strategies from achieving the maximal commercial profit.

In this paper, we focus on a billboard advertising scenario, which is shown in Fig. 1. There are four available billboards placed at different locations. Two potential customers unconsciously move among the billboard locations. An advertiser wants to do advertising for a product with a limited budget (400 in Fig. 1) and that is not enough to place advertisements

on all billboards. Each billboard has an advertising cost (e.g.,  $b_1=150$ , etc.) and also a potential profit. The profit is measured quantitatively by the expected number of attracted customers. Obviously, whether a potential customer is attracted by the billboard is influenced by many factors, such as the probability that the customer can see the billboard, the degree of his/her interest in billboard advertising, and the detour distance that he/she goes to buy the product. We use the detour distance as an example to describe our problem (Fig. 1). Under a total budget of 400, the advertiser can at most cover the costs of two billboards. Two example plans show the different advertising strategies: Plan 1 advertises at billboards  $b_1$  and  $b_3$ , while plan 2 advertises at billboards  $b_2$  and  $b_4$ . If the purpose is to minimize the detour distance, obviously plan 1 is better than plan 2. But actually, only using detour distance is not comprehensive, as described above, and many factors should be taken into consideration when we measure a billboard's potential profit. For example, if customers A and B cannot pass by the billboards  $b_1$  and  $b_3$ , but they can pass by the billboards  $b_2$  and  $b_4$ , plan 2 is better for this situation. Hence, in order to maximize the commercial profit under the limited budget, this paper focuses on the problem of selecting billboards with the consideration of multiple factors, which can affect the probabilities of customers being attracted.

To solve the above billboard selection problem, we need to collect the potential customers' preferences, mobility patterns and their detour distance, which are privacy-sensitive. Hence, how to collect the potential customers' preferences, mobility patterns and their detour distance is the first challenge. Moreover, the advertiser needs to accurately quantify the expected commercial profit of each billboard to decide which billboards to do advertising, which is the second challenge. When the advertiser has multiple products to advertise, it is essential to make rational use of the budget and a reasonable distribution of advertisements, which is the third challenge.

In this paper, we normalize the influences of the above three factors, formulate the billboard selection problem as a probabilistic set coverage problem, and propose some effective advertising strategies to address this problem. First, we adopt Mobile Crowdsensing (MCS)<sup>[8-10]</sup> to collect the privacy-sensitive customer profiles<sup>[6, 11]</sup> including their vehicular trajectories and preferences. For example, if a user has performed some sensing tasks for an MCS application, the application may record his/her GPS trajectories



▲ Figure 1. Problem description of billboard advertising, where the advertiser has only one kind of product

during his/her working time. Also, the user's historical information can be used to infer his/her preferences for the advertising<sup>[11]</sup>. For example, if a user has completed many tasks near the food market, the food market may be considered as a preference of this user. Based on their vehicular trajectories and the location of the advertiser's shop, we can estimate the potential customers' detour distance for purchasing the product. With the information, we then study the advertising problem under two situations, where the advertiser may have only one or multiple products. For the first situation where the advertiser has only one kind of product to advertise, we propose two heuristic advertising strategies to greedily choose advertising billboards, which can maximize the total expected commercial profit for the advertiser. For the second situation where the advertiser has multiple products, we adopt the simulated annealing algorithm to search the global optimum instead of local optimum.

The main contributions of this paper are summarized as follows:

- We formulate this billboard advertising problem as a non-deterministic polynomial (NP)-hard problem to select appropriate billboards for the advertiser to achieve the maximum commercial profit with the constraint of budget. We design our advertising strategies with consideration of customers' mobility, preferences and detour distance.
- For the situation where the advertiser has only one kind of product to advertise, we propose two bounded heuristic advertising strategies, whose approximation ratios are  $(1 - \frac{1}{e})$ .
- For the other situation where the advertiser has multiple products to advertise, we propose an advertising strategy by using the simulated annealing algorithm to search the global optimum.
- We conduct extensive simulations based on three real-world trajectories: roma/taxi<sup>[12]</sup>, epfl<sup>[13]</sup>, and geolife<sup>[14]</sup>. The results show that compared with other strategies, our advertising strategies can achieve superior commercial profit for the advertiser.

The remainder of this paper is organized as follows. We review the related work in Section 2. We describe the system models and formulate this billboard selection problem in Section 3. We describe the general technologies we used in Section 4. The detailed advertising strategies are proposed in Section 5. In Section 6, we conduct the simulations to determine the performances of our advertising strategies. We conclude this paper in Section 7.

## 2 Related Work

### 2.1 Advertising Strategy

There have been many works on advertising strategy. In Ref. [6], WANG et al. propose a utility-evaluation-based optimal searching approach to empower audience targeted bill-

board advertising by using vehicle trajectory data with consideration of audiences' interests. In Ref. [7], ZHENG et al. study a promising application in Vehicular Cyber-Physical Systems (VCPS) to attract potential customers for the shopkeeper by using Roadside Access Points (RAPs). In Ref. [3], NIGAM et al. present the design and implementation of an intelligent advertising system which is integrated in a network and can be used in many retail stores, shopping malls and shopping centers. In Ref. [4], LIU et al. propose a system which uses taxi trajectories to help select the locations of billboards. In Ref. [5], HUANG et al. propose a strategy to maximize the coverage of advertisements with consideration of individuals' interests and mobility patterns. In Ref. [15], KRISHNA et al. develop a new application for detecting significant billboards for adults and older people in street-laying areas. In Ref. [16], AN et al. propose an advertisement system for enhancing the efficiency of advertisement by using the Wi-Fi union mechanism. In Ref. [17], ZHANG et al. optimize the influence of outdoor advertising with the consideration of impression counts. They propose a tangent line based algorithm to select roadside billboards for maximizing the influence of outdoor advertising.

Different from the research works mentioned above where the authors do not jointly take potential customers' mobility, preferences and detour distance into consideration, in this paper, we focus on the problem of selecting billboards with the consideration of the above factors, in order to maximize the commercial profit for an advertiser.

### 2.2 Mobile Crowdsensing

There have also been some works focusing on mobile crowdsensing. In Ref. [11], KARALIOPOULOS et al. draw on logistic-regression techniques from machine learning to learn users' individual preferences from past data in Mobile Crowdsensing. In Ref. [18], ARIYA SANJAYA et al. present an application program, which provides data for SOROT (Citizen Reporting MCS Application) platform by analyzing citizen participation, and speeds up the solution of urban problems. In Ref. [19], CHEUNG et al. propose an algorithm for calculating the optimal user decision-making under general conditions by using the dynamic programming method. Based on the game-theory approach, CAO et al. in Ref. [20] propose an incentive mechanism in order to encourage mobile devices to share their own resource to perform sensing tasks cooperatively. In Ref. [21], GONG et al. focus on the path planning and task assignment problem in Mobile Crowdsensing, so that the total task quality can be maximized with the constraints of user travel distance and budget. In Ref. [22], MARJANOVIĆ et al. propose an edge computing architecture, which is suitable for large-scale MCS services by putting the main MCS functions in the reference of MEC architecture. For truth discovery in mobile crowdsensing, ZHENG et al. in Ref. [23] propose two novel privacy-aware crowdsensing designs with truth discov-

ery so that the bandwidth and computation performance on individual users can be significantly improved. In Ref. [24], WANG et al. propose a two-stage solution to the heterogeneous multi-task assignment (HMTA) problem, which utilizes the implicit spatiotemporal correlation between heterogeneous tasks to effectively handle multiple concurrent tasks in shared resource pools. In Ref. [25], WANG et al. propose a new framework of participatory perceptual multi-task allocation, which coordinates the allocation of multiple tasks on the multi-task PS platform to maximize the overall effectiveness of the system.

These works we mention above focus on different areas of mobile crowdsensing, while we attempt to extract potential customers' implicit information, such as their trajectories and preferences, by using the MCS data, in order to select appropriate billboards for the advertiser.

### 3 System Model and Problem Formulation

#### 3.1 System Model

Consider that there is an advertising system which is composed of a crowd of potential customers, denoted by the set  $U = \{u_1, u_2, \dots, u_n\}$  and a set of available billboards  $V = \{v_1, v_2, \dots, v_m\}$ . The costs of available billboards are denoted by  $C = \{c_1, c_2, \dots, c_m\}$ . The different areas in the map can be represented as  $L = \{l_1, l_2, \dots, l_h\}$ . All billboards are located at different areas and each area has only one billboard. Moreover, the preference types are denoted by the set  $A = \{a_1, a_2, \dots, a_j\}$ . Hence, without the loss of generality, we denote the preferences of potential customer  $u_i$  as  $A_{u_i} \subseteq A$ . We suppose that each kind of products has an advertisement which can be denoted by  $T = \{t_1, t_2, \dots, t_r\}$ , and the attributes of product  $t_x$  can be denoted by  $A_{t_x} \subseteq A$ . Meanwhile, each billboard is available for only one advertisement.

In our scenario, at the beginning, each potential customer  $u_i$  starts moving from his/her initial location, and goes to his/her destination. Every time a potential customer sees an advertisement for a product, he/she will decide whether to buy the product. The detour distance is an important factor affecting the customer's decision, and then we use  $d(u_i)$  to denote the detour distance for  $u_i$ , which represents how much more distance  $u_i$  needs to go than his/her original route for buying the product. If a potential customer has been attracted to buy the product, this customer cannot be attracted by the same advertisement again. In other words, each potential customer can be attracted by a product at most once. The attracted customers are denoted by  $U_{attr}$ .

#### 3.2 Problem Description

With the limit of budget, which is denoted by  $B$ , we attempt to choose a set of billboards denoted by  $S = \{s_1, s_2, \dots, s_k\}$  from  $V$  for advertising. When the advertiser has multiple products,

we also need to choose the advertisements for the selected billboards to maximize the commercial profit. In this paper, we suppose that if a billboard  $v_i$  is selected in  $S$ , then it will deliver advertisement content to those potential customers whose locations are in its area until the deadline. The deadline is how long the advertiser can use a billboard, and we suppose that all billboards have the same deadline. Our purpose is to find the best advertising strategy for the following billboard selection problem:

$$\begin{aligned} & \text{Maximize } F = \sum_{i=1}^{U_{attr}} \sum_{x=1}^r \phi f^{t_x} - B \\ & \text{s.t. } \sum_{j=1}^k c_{s_j} \leq B, \forall s_j \in S, S \subseteq V, \phi \in \{0, 1\}, \end{aligned} \quad (1)$$

where  $F$  is the total commercial profit for the advertiser from billboard advertising. The problem is that an advertiser should attract potential customers as many as possible with limited budget  $B$ . In this paper, we assume that if a customer is attracted after he/she sees the advertisement, then the advertiser will obtain a profit from the customer. If a potential customer is attracted by an advertisement  $t_x$ , the profit that the advertiser can get is denoted by  $f^{t_x}$  and  $\phi = 1$ . If the customer is not attracted,  $\phi = 0$ . In order to reduce the complexity of the calculation, we suppose that the customers attracted by the same advertisements can create the same profit for the advertiser, and the customers attracted by the different advertisements may create different profits for the advertiser. In other words, it can be denoted by  $f^{t_x} = f^{t_y}, \forall t_x \in T, \forall t_y \in T$  or  $f^{t_x} \neq f^{t_y}, \forall t_x \in T, \forall t_y \in T$ . Our advertising strategies aim at finding the best set of billboards to deliver the advertisements, so that the commercial profit for the advertiser is maximized, with the constraint that the total costs of selected billboards should be less or equal to the budget.

#### 3.3 NP-Hard Proof

Before solving the above optimization problem, we first attempt to prove that the billboard selection problem is NP-hard, which is shown as follows:

First of all, we formulate this problem in Eq. (1) as the probabilistic set coverage problem, which includes a collection of element sets  $X = \{X_1, X_2, \dots, X_m\}$  with the corresponding costs  $c = \{c_1, c_2, \dots, c_m\}$ .  $X_i$  consists of a lot of elements, which is denoted by  $O = \{O_1, O_2, \dots, O_n\}$ . The associated possibilities that the elements can be covered are denoted by  $p = \{p_1, p_2, \dots, p_n\}$  and the associated weights of elements are denoted by  $W = \{w_1, w_2, \dots, w_n\}$ . The objective is to select a subcollection of  $X$  under the budget constraint  $B$  to maximize the weights of covered elements.

Then, we reconsider the billboard selection problem in this paper. First of all, we consider the situation where the advertiser has only one type of product. Since the commercial profit depends on the number of attracted customers, we can regard

the potential customers as the element set  $O$ . The probabilities that the potential customers decide to buy the product when they see the advertisement can be regarded as  $p$ . The profit that the advertiser gets from each customer can be considered as  $W$ . We can regard the billboard set we need to choose as  $X$ , and the total costs of selected billboards can be regarded as  $c$ . We need to select the billboards to attract as many potential customers as possible, hence the billboard selection problem can be regarded as the probabilistic set coverage problem. Since the probabilistic set coverage problem is NP-hard, the billboard selection problem when the advertiser has only one type of product is NP-hard. Moreover, when the advertiser has multiple products to do advertising, the selection problem becomes more complicated, since we should not only consider how to select billboards but also the advertisement placed on the billboard. Hence, under this situation where the advertiser has multiple products, the billboard selection problem is also NP-hard.

## 4 General Technologies

### 4.1 Single-Product

In this section, we consider the situation where the advertiser needs to advertise for only one kind of product and potential customers need to decide whether to go to buy the product every time they see the advertisement. We use  $T$  to denote the advertisement of the product for ease of calculation and each attracted customer can create the same profit  $f$  for the advertiser in this section. We first discuss how to predict the potential customers' mobility patterns. Then, we quantify the influences of customers' preferences and detour distance on customers' attraction probabilities, respectively. Finally, we quantify the utilities of different billboards.

#### 4.1.1 Mobility Prediction

First of all, we attempt to predict each potential customer's location so that we can select the appropriate billboards to improve the effectiveness of advertising. It is not difficult to map each customer's trajectories into a square area in a plane region, especially when the area is small<sup>[26]</sup>. Thus, we can grid the area in the map like Fig. 1 and convert each customer's trace into a sequence of grids, in order to reduce the difficulty of calculation. After we grid the map, the billboards' locations can be converted into fixed grids. We assume that, if a potential customer enters a grid which has a selected billboard, he/she will see the advertisement and decide whether to buy the product. In this paper, we adopt the semi-markov model<sup>[27-29]</sup> to predict the customers' mobility. One of the most important equations of semi-markov,  $Z(\cdot)$  is defined by Eq. (2).

$$Z_u(l_i, l_j, X) = P(S_u^{n+1} = l_j, x_u^{n+1} - x_u^n \leq X | S_u^0, \dots, S_u^n; x_u^0, \dots, x_u^n) =$$

$$P(S_u^{n+1} = l_j, x_u^{n+1} - x_u^n \leq X | S_u^n = l_i), \quad (2)$$

where  $Z_u(l_i, l_j, X)$  is the probability that the customer  $u$  will move from his/her current grid  $l_i$  to the grid  $l_j$  at or before time  $X$  when he/she moves next time.  $S_u^k$  represents the customer  $u$ 's  $k$ -th location during his/her moving and its corresponding arrival time is denoted as  $x_u^k$ . The grid that the customer will enter in the next time unit is related to his/her current grid, which can be obtained from the customer's historical trace records. Then, we can define another key equation  $Q(\cdot)$ , denoted by Eq. (3).

$$Q_u(l_i, l_j, X) = \begin{cases} \sum_{k=1}^h \sum_{x=1}^X (Z_u(l_i, l_k, x) - Z_u(l_i, l_k, x-1)) \cdot \\ Q_u(l_k, l_j, X-x), \quad i \neq j \\ 1 - \sum_{k=1, k \neq i}^h Z_u(l_i, l_k, X) + \\ \sum_{k=1, k \neq i}^h \sum_{x=1}^X (Z_u(l_i, l_k, x) - Z_u(l_i, l_k, x-1)) \cdot \\ Q_u(l_k, l_i, X-x), \quad i = j \end{cases} \quad (3)$$

$Q(\cdot)$  denotes the probability of a potential customer  $u$  moving across  $l_j$  from  $l_i$  before time slot  $X$ . It is easy to find that the potential customers cannot move from one grid to another when  $X = 0$ , which is reasonable, so we can get  $Q_u(l_i, l_i, 0) = 1$  and  $Q_u(l_i, l_j, 0) = 0, (i \neq j)$ . Next, we calculate the probability of a customer passing any grid  $l_j$  before deadline  $X$  as follows:

$$P^j(u) = 1 - \prod_{x=0}^X (1 - Q_u(l_i, l_j, x)). \quad (4)$$

#### 4.1.2 Customer Preference Level

After considering customers' mobility, in order to determine the expected commercial profit of each billboard, we need to measure a potential customer  $u_i$ 's preference level for the product  $T$ , which is denoted as  $P_{prefer}(u_i)$ . First, we collect the customer  $u_i$ 's preferences  $A_{u_i}$  and the product  $T$ 's attributes  $A_T$  where  $A_{u_i} \subseteq A$  and  $A_T \subseteq A$ . Then the preference level  $P_{prefer}$  can be calculated by the following equation:

$$P_{prefer}(u_i) = \frac{A_{u_i} \cap A_T}{A_{u_i}} \quad (5)$$

Obviously, if the product's attributes  $A_T$  can match all the customer's preferences  $A_{u_i}$ , then  $P_{prefer} = 1$ , which means the potential customer  $u_i$  would be likely to buy the product by the factor of preferences.

#### 4.1.3 Detour Distance

In the single-product scenario, the potential customer may change his/her trajectory if he/she is attracted by the advertisement. Hence, as mentioned above, another factor that the customer  $u_i$  will consider when he/she decides whether to buy a

product is the detour distance, which is denoted by  $d(u_i)$ . We first use Euclidean distance to measure potential customer's path length. As we can see from the Fig. 1, the original path for the customer A is  $d1 + d2 + d3$ . When the customer A sees the advertisement from the billboard b1, which is selected for advertising, he/she will decide whether to go to the shop for buying the product. If he/she is attracted to buy the product, the path to the shop is  $d4$ , and the path from the shop to his/her original destination is  $d5$ . The detour distance can be calculated as follows:

$$d(u_i) = \begin{cases} \min_{v_j \in V, v_j \in S} d4 + d5 - (d2 + d3), & \text{if } \exists v_j, v_j \in S \\ \infty, & \text{otherwise} \end{cases} \quad (6)$$

During the path, the customer may see a lot of billboards that have been chosen for advertising, so he/she will decide whether to buy the product after he/she sees a selected billboard. Hence, the detour distance should be calculated as the distance from the current billboard to the customer's destination. Then we need to measure how the detour distance affects the probability that the customer would go to the shop. The equation is shown as follows:

$$P_{detour}(u_i) = \begin{cases} 1 - \frac{d(u_i)}{D_{max}}, & \text{if } d(u_i) \leq D_{max} \\ 0, & \text{otherwise} \end{cases}, \quad (7)$$

where  $D_{max}$  is a predefined threshold and  $P_{detour}(u_i)$  represents the detour distance level which affects the probability that the customer will be attracted to buy the product. In this paper, we set  $D_{max}$  as the maximum diagonal length in the selected area. It is not difficult to find that the less the detour distance is, the higher chance that the customer will go to the shop for buying the product, which is reasonable.

#### 4.1.4 Billboard Utility

In this part, we use utility to represent the expected commercial profit of a billboard  $v_j$ , which is denoted by  $F(v_j)$ .  $F(v_j)$  is the expected commercial profit that the billboard  $v_j$  can bring to the advertiser. First of all, we need to calculate the probability that the customer will be attracted to buy the product after he/she sees the advertisement, the equation of which is shown as follows:

$$P_{attract}^{v_j}(u_i) = \alpha P^{l_j}(u_i) \times \beta P_{prefer}(u_i) \times \gamma P_{detour}(u_i), \quad (8)$$

where  $l_j$  is the grid that the billboard  $v_j$  is located.  $\alpha$ ,  $\beta$  and  $\gamma$  are relative weights where  $\alpha + \beta + \gamma = 1$ . By now, the probability that the customer will be attracted to buy the product after he/she sees the advertisement is obtained. Besides, the probability that the customer will be attracted to buy the product can be affected by the different billboards that the customer

sees. Therefore, it is necessary to determine the influences of different billboards on the same potential customer, which can be denoted as follows:

$$P_{attract}(u_i) = 1 - \prod (1 - P_{attract}^{v_j}(u_i)), \quad \forall v_j \in S, \quad (9)$$

where  $P_{attract}(u_i)$  is the probability that the customer  $u_i$  will be attracted to the shop after he/she sees the current billboard with consideration of different billboards' impacts. In other words,  $P_{attract}(u_i)$  is the probability that the potential customer will decide to buy the product at least once when he/she sees the same advertisement many times. Then the utility of a billboard for a specific advertisement can be calculated as follows:

$$F(v_j) = [1 - \prod_{i=1}^n (1 - P_{attract}^{v_j}(u_i))] \times f, \quad v_j \in V, u_i \in U. \quad (10)$$

Then we can get the total utility of the billboard set, which is shown in Eq. (11):

$$F = f \times \sum P_{attract}(u_i) - B, \quad \forall u_i \in U. \quad (11)$$

## 4.2 Multi-Product

In this subsection, we consider the situation where the advertiser may have multiple products and potential customers do not need to go to the shop immediately to buy products. In other words, the potential customer will not change his/her trajectory if he/she is attracted by the advertisement. We suppose that each product has a corresponding advertisement  $T = \{t_1, t_2, \dots, t_r\}$  and each advertisement of a product attracts a customer with different commercial profit which can be denoted as  $\eta = \{f_1, f_2, \dots, f_r\}$ . Each billboard can only be selected for one advertisement. For example, as shown in Fig. 2, there are four available billboards (b1 - b4) placed at different locations. Two potential customers are unconsciously moving among the billboard locations. The advertiser wants to do advertising for three different products (Type 1 - 3) while he/she has a limited budget (500 in Fig. 2), and that is not enough to place the advertisements on all billboards. Each billboard has an advertising cost (e.g., b1=150, etc.) and also an expected commercial profit. Two example plans show the different advertising strategies under the budget constraint. Plan 1 advertises at billboards b1, b2 and b4 for products 1, 2 and 3, while plan 2 advertises at billboards b1, b2 and b3 for products 3, 1 and 2. In order to maximize the profit within the limited budget, the advertiser needs to determine which plan is better. Next, we will discuss how to address this problem.

### 4.2.1 Mobility Prediction

First of all, we still consider how to predict customers' mobility patterns. Due to the reason that potential customers' mobility patterns wouldn't be affected by the number of products, the mobility prediction method we proposed earlier can

be applied for this situation.

#### 4.2.2 Customer Preference Level

Next, we will discuss how to determine the customers' preference level in this part. Since the advertiser has multiple products which have different attributes, it is not difficult for us to find that we can reformulate Eq. (5) as follows:

$$P_{prefer}^{t_x}(u_i) = \frac{A_{u_i} \cap A_{t_x}}{A_{u_i}}, \quad (12)$$

where  $A_{u_i}$  denotes the preferences of a customer  $u_i$  and  $A_{t_x}$  denotes the attributes of product  $t_x$ . It is obvious that the more the product's attributes match the customer's preferences, the more likely that the customer will decide to buy the corresponding product.

#### 4.2.3 Billboard Utility

In this situation, potential customers do not need to go to the shop immediately to buy products, hence, we can ignore the influence of detour distance on customers' decisions. As a result, the probability that the customer will be attracted can be reformulated as follows:

$$P_{attract}^{v_j, t_x}(u_i) = \alpha P^{l_j}(u_i) \times \beta P_{prefer}^{t_x}(u_i), \quad (13)$$

where  $l_j$  is the grid where the billboard  $v_j$  is located.  $\alpha$  and  $\beta$  are relative weights where  $\alpha + \beta = 1$ . Since different products have different attributes, we consider there is no competition among different products and the probabilities of a potential customer buying different products are independent. In other words, a customer's purchase of one product does not affect the possibility of buying another different product. Hence, each customer can be attracted by different products and bring more commercial profit to the advertiser.

The utility of a billboard for a specific advertisement  $t_x$  can

be calculated as follows:

$$F^{t_x}(v_j) = [1 - \prod_{i=1}^n (1 - P_{attract}^{v_j, t_x}(u_i))] \times f^{t_x},$$

$$v_j \in V, u_i \in U. \quad (14)$$

Then we can get the total utility of the billboard set, which is shown as follows:

$$F = \sum_{x=1}^r (f_{t_x} \times P_{attract}^{v_j, t_x}(u_i)) - B, \forall u_i \in U, \forall v_j \in S. \quad (15)$$

## 5 Advertising Strategies

### 5.1 Single-Product

In this section, we propose two heuristic advertising strategies to address the billboard selection problem for the situation where the advertiser has only one kind of product to advertise.

#### 5.1.1 Same Cost for Each Billboard

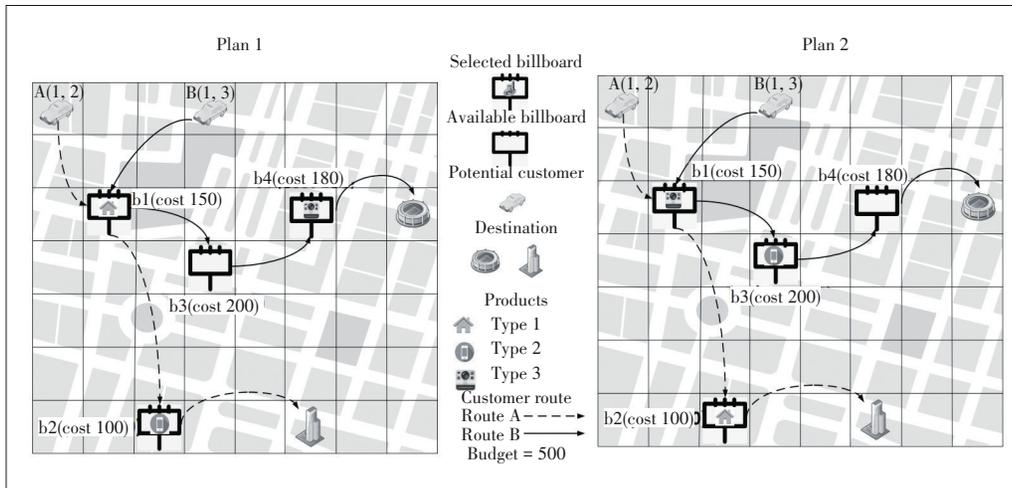
First of all, we consider the situation where all billboards have the same cost. In this situation, we can convert the budget restriction into the billboard's quantity restriction where we need to select a billboard set to maximize the profit for the advertiser with the constraint of billboard number  $k$ . The detailed greedy algorithm Advertising Strategy for Constant Cost (ASFCC) is shown in Algorithm 1.

#### Algorithm 1. Advertising Strategy for Constant Cost (ASFCC)

**Input:** number of billboard  $k$ , a set of billboards  $V$

**Output:** the selected billboard set  $S$

- 1:  $S \leftarrow \emptyset$ ;
- 2:  $F \leftarrow 0$ ;
- 3: **for**  $i=1$  to  $k$  **do**
- 4:  $v_h \leftarrow \arg \max_{v_h \in V \setminus S} F_{S \cup v_h}$
- 5:  $S = S \cup v_h$ ; update  $F$
- 6:  $V = V \setminus v_h$
- 7: **return** the selected billboard set  $S$ .



▲ Figure 2. Problem description of billboard advertising, when the advertiser has multiple products

In Algorithm 1, we select the billboard which has the largest utility at the beginning of the algorithm. Then, we continue selecting the billboard, which can maximize the value of the total utility  $F$  to be the second billboard among the other billboards and add it into  $S$ . This process will be executed for  $k$  times until the number of billboards which have been selected meets the requirements or all the billboards have been selected.

The reason why we do not select the billboard with the current largest utility is that the local optimal solution is not necessarily the global optimal solution. For example, consider that there are three billboards  $a, b, c$  and two customers  $u_1, u_2$ . Billboards  $a$  and  $b$  may attract customer  $u_1$  with probabilities 1 and 0.8. Billboard  $c$  may attract customer  $u_2$  with probability 0.5. Since each attracted customer can create the same profit for the advertiser, it is obvious that we should select the billboards  $a$  and  $c$  to achieve the maximum commercial profit.

By now, we have proposed a greedy advertising strategy to address the above NP-hard problem. According to Ref. [30], we can confirm that  $F$  is a submodular function, which can be summarized as follows: consider that there are two arbitrary node sets  $S_1$  and  $S_2$ ,  $S_1 \subset S_2$ , and  $\forall v_k \in V \setminus S_1$ , the submodular property holds, i. e.,  $F_{S_1 \cup v_k} - F_{S_1} \geq F_{S_2 \cup v_k} - F_{S_2}$ . The bound can also be derived from Ref. [30], which is  $(1 - \frac{1}{e})$ .

### 5.1.2 Different Costs for Each Billboard

Now, we attempt to propose another heuristic advertising strategy Advertising Strategy for Different Costs (ASFDC) for the situation where all billboards have different costs. As we can see from Algorithm 2, the exhaustive method is used to determine the billboard set  $S_1$  which has the largest expected commercial profit where  $k$  is the quantity restriction. Then we execute the greedy process until the budget is lower than the lowest cost of available billboards or all the billboards have been selected to the  $S_2$ . At last, we compare the utility of billboard set  $S_1$  with the utility of billboard set  $S_2$  to determine which is larger to be the final result.

**Algorithm 2.** Advertising Strategy for Different Costs (ASFDC)

**Input:** number of billboard  $k$ , a set of billboards  $V$ , cost set for each billboard  $C$ , total budget  $B$

**Output:** the selected billboard set  $S$

```

1:  $S_1 \leftarrow \arg \max \{F(S_{temp}) \mid |S_{temp}| < k, S_{temp} \subseteq V, \text{ and } c(S_{temp}) \leq B\}$ ;
2:  $S_2 \leftarrow \emptyset$ ;
3: for  $S_{temp} \subseteq V, |S_{temp}| = k$  and  $c(S_{temp}) \leq B$  do
4:   while  $\forall S_{temp} \neq \emptyset$  do
5:      $v_h \leftarrow \arg \max_{v_h \in V \setminus S_2} \frac{F(v_h)}{c(v_h)}$ 
6:     if  $c(S_{temp}) + c_{v_h} \leq B$  then
7:        $S_{temp} = S_{temp} \cup v_h$ 
8:     if  $F(S_{temp}) > F(S_2)$  then
9:        $S_2 \leftarrow S_{temp}$ 
10: if  $F(S_1) > F(S_2)$  then
11:   return the selected billboard set  $S_1$ .
12: else
13:   return the selected billboard set  $S_2$ .
```

By now, we have proposed another advertising strategy to address the billboard selection problem when all billboards

have different costs. According to Ref. [31], we can get  $F(S_i) \geq (1 - \frac{1}{e})F(S_{opt}), k \geq 3$ , which represents that when  $k \geq 3$ , the approximation ratio for this algorithm is  $(1 - \frac{1}{e})$ .

### 5.2 Multi-Product

We have described the scenario of the situation where the advertiser has multiple products to advertise, which is more difficult than the situation in Section 4. In order to address this problem, we attempt to adopt the simulated annealing algorithm. In this situation, we consider that all billboards have different costs and the same billboard would cost the same for different products.

First of all, we modify the algorithm ASFDC so that it can be applied in this situation, which is shown in Algorithm 3. The difference is that we need to calculate the utilities of each billboard for different products in each iteration, and then we select the billboard with the advertisement which can maximize the total expected commercial profit. This process will be repeated until we run out of the budget. Then we take the result obtained in the previous step as the initial solution of the simulated annealing algorithm Advertising Strategy for Multi-advertisement with Different Costs (ASFMD), which is shown in Algorithm 4.

**Algorithm 3.** Advertising Strategy for Different Advertisements with Different Costs (ASFADDC)

**Input:** number of billboard  $k$ , a set of billboards  $V$ , cost set for each billboard  $C$ , total budget  $B$ , advertisement set  $T$

**Output:** the selected billboard set  $S$

```

1:  $S_1 \leftarrow \arg \max \{F(S_{temp}) \mid |S_{temp}| < k, S_{temp} \subseteq V, \text{ and } c(S_{temp}) \leq B, \forall t_x \in T\}$ ;
2:  $S_2 \leftarrow \emptyset$ ;
3: for  $S_{temp} \subseteq V, |S_{temp}| = k$  and  $c(S_{temp}) \leq B$  do
4:   while  $\forall S_{temp} \neq \emptyset$  do
5:      $v_h \leftarrow \arg \max_{v_h \in V \setminus S_2} \frac{F^{t_x}(v_h)}{c(v_h)} \quad \forall t_x \in T$ 
6:     if  $c(S_{temp}) + c_{v_h}^{t_x} \leq B$  then
7:        $S_{temp} = S_{temp} \cup v_h^{t_x}$ 
8:     if  $F(S_{temp}) > F(S_2)$  then
9:        $S_2 \leftarrow S_{temp}$ 
10: if  $F(S_1) > F(S_2)$  then
11:   return the selected billboard set  $S_1$ .
12: else
13:   return the selected billboard set  $S_2$ .
```

**Algorithm 4.** Advertising Strategy for Multi-Advertisement with Different Costs (ASFMD)

**Input:** a set of billboards  $V$ , cost set for each billboard  $C$ , total budget  $B$ , a set of potential users  $U$ , an initial set



$S, \theta, Temp, Temp_{max}$

**Output:** the final selected billboard set  $S^*$

```

1:  $S^* = S$ 
2: while stopping condition not met do
3:   choose a billboard from  $S$  to delete
4:   choose a billboard from available billboard in  $V$  to add
5:   generate a new billboard set  $S'$ 
6:   if  $F(S') > F(S)$  then
7:      $S = S'$ 
8:   else
9:      $S = S'$  with a probability  $p = \exp(- (F(S) - F(S'))/Temp)$ 
10:  if  $F(S) > F(S^*)$  then
11:     $S^* = S, Temp_b = Temp$ 
12:   $Temp = \theta \times Temp$ 
13:  if  $Temp < 0.01$  then
14:     $Temp_b = 2 \times Temp_b$ 
15:   $Temp = \min\{Temp_b, Temp_{max}\}$ 
16: return the selected billboard set  $S^*$ .

```

In Algorithm 4, we first select a billboard to delete from the selected billboard set  $S$  and then select a billboard from the available billboards in  $V$  to generate a new solution  $S'$  (lines 3-5). Next, we compare the expected commercial profit of  $S'$  with that of  $S$  to determine which solution should be accepted in the next iteration (lines 6 - 9). Then, we compare the expected commercial profit of the current solution with the current best solution to update the current best solution (lines 10 - 11). The temperature  $Temp$  is updated at each iteration (lines 11 - 15). Next, we will discuss the method of removal and insertion we used in Algorithm 4, as well as the stopping conditions.

### 5.2.1 Billboard Removal Method

We adopt two methods to decide which billboards to delete, which are described as follows:

- Random removal: we randomly select a billboard from the selected billboard set  $S$ , and remove it.
- Probability removal: we calculate the expected commercial profit of each selected billboard and remove one of them with probability. The higher the expected profit is, the lower the probability of removal would be.

### 5.2.2 Billboard Insertion Method

We also design two methods to determine which billboards to insert and the advertisements on the billboards. The detailed description is as follows:

- Probability Insertion: we calculate the expected commercial profit of each available billboard, and select one of them with probability to add to the selected billboard set  $S$ . The higher the expected profit is, the higher the probability of insertion would be.
- Expectation Maximization Insertion: we calculate the expected commercial profit of each available billboard, and select

one of them with the max expected profit to insert to the billboard set  $S$ .

For a better understanding, we provide an example: as shown in Table. 1, there are two available billboards and three different advertisements. The expected profit has been calculated and shown in the table. First of all, we consider the advertisement that brings the maximum expected profit for each billboard as the final advertisement for that corresponding billboard. Obviously, the final advertisement for billboard  $A$  should be advertisement 3, and the final advertisement for billboard  $B$  should be advertisement 1. If Probability Insertion is adopted, then we should normalize the expected profit. Thus, the probability for selecting billboard  $A$  is  $4/7$ , and the probability for selecting billboard  $B$  is  $3/7$ . On the other hand, if Expectation Maximization Insertion is adopted, then we will select the billboard with the maximum expected profit, i. e., the billboard  $A$  in this example. The process of removal is similar to that of insertion.

We use a multi-thread approach to improve the experimental performance which combines the above removal and insertion methods. As a result, we can create four threads with different combinations of removal and insertion.

### 5.2.3 Stopping Conditions

We determine two stopping conditions in our simulations, which are listed as follows:

- The maximum number of iterations is set to 1 000 000, and when the iteration number exceeds the limited number, the algorithm would stop.
- The maximum number of iterations without improving is set to 100 000, which means after 100 000 iterations, if the result is not improved, the algorithm would stop.

## 6 Performance Evaluation

### 6.1 Simulation Traces and Settings

In this paper, three real-world datasets, the roma/taxi trace set<sup>[12]</sup>, epfl trace set<sup>[13]</sup> and geolife trace set<sup>[14]</sup>, are adopted to verify the performances of our proposed strategies. In the roma/taxi trace set, 320 taxi drivers that work in the center of Rome are included. The traces in this dataset represent the positions of those taxi drivers, which are collected every 7 seconds and sent to a central server. In the epfl trace set, about 500 taxis' GPS coordinates are included which are collected over 30 days in the San Francisco Bay Area. Each taxi is equipped with a GPS receiver and sends a location-update to

▼Table 1. An example for removing and inserting billboards

	Billboard A	Billboard B
Advertisement 1	200	300
Advertisement 2	300	200
Advertisement 3	400	100

a central server. The records are fine-grained so that we can accurately interpolate user positions between location-updates. In the geolife trace set, there are 17 621 trajectories whose total distance is about  $1.2 \times 10^6$  km. The total duration of this dataset is about 48 000 h which are collected by different GPS loggers and phones.

First of all, we process the dataset by filtering out some abnormal users including those with discontinuous traces or remote locations. Next, we match these traces into a map area and convert it into a gridded map, which are processed by Baidu Map application programming interface (API). We randomly select a supermarket or a mall in the city as the advertiser's shop and we also select 35 billboards located in different areas to be the candidate billboards.

For the first situation where the advertiser has only one kind of product, we set the  $\alpha$ ,  $\beta$  and  $\gamma$  in Eq. (8) to  $1/3$ . The preferences of each customer are randomly generated to reduce the difficulty, but they can also be inferred from each customer's historical information, which is not the focus of this paper. The total number of preference types  $|A| = 20$ . The deadline is set from 500 to 600. The costs of billboards are set to 10 when the costs are constant and the costs of billboards are set from 10 to 20 when the costs are different. The budget is set from 50 to 170 when the costs are constant and it is set from 100 to 200 when the costs are different. The number of each customer's preferences is set from 5 to 15 in our simulations. Each attracted customer can bring a profit of 10 to the advertiser. We repeat our simulation 10 000 times, taking the average result as the final result.

For the situation where the advertiser has multiple products, we set the deadline in the simulation from 150 to 250. We set the  $\alpha$  and  $\beta$  in Eq. (13) to  $1/2$ . The budget in this situation is set from 100 to 200 and the profit per attracted customer for each product is set from 10 to 30. We set  $\theta = 0.999$ ,  $Temp = 10\ 000$  and  $Temp_{max} = 10\ 000$ . The other experimental parameters are the same as those when the costs are different for the single-product scene. In this paper, we take the total commercial profit as the evaluation metric to measure the performances of different strategies.

## 6.2 Strategies for Comparison and Metric

For the first situation where the advertiser has only one kind of product, in order to determine the performances of our advertising strategies, we compare the ASFCC with ASFCC-Basic, Random and Capped Greedy (CG)<sup>[7]</sup> when all billboards have the same cost. ASFCC-Basic would select the billboards which have the largest commercial profit, and Random would randomly select the billboards for advertising. The CG would select the billboards, which can maximize the total commercial profit of selected billboards without consideration of customers' preferences.

When all billboards have different costs, we compare ASFDC with ASFDC-Basic and Random, where ASFDC-Basic

would choose the billboards which have the largest commercial profit for advertising. Random would randomly select the billboards with the limit of budget.

For the other situation where the advertiser has multiple products, we compare the ASFMDCC and ASFDDADC with ASFDDADC-Basic and Random, where ASFDDADC-Basic would choose the billboards with the largest commercial profit. The other strategy Random would randomly select the billboards and their advertisements.

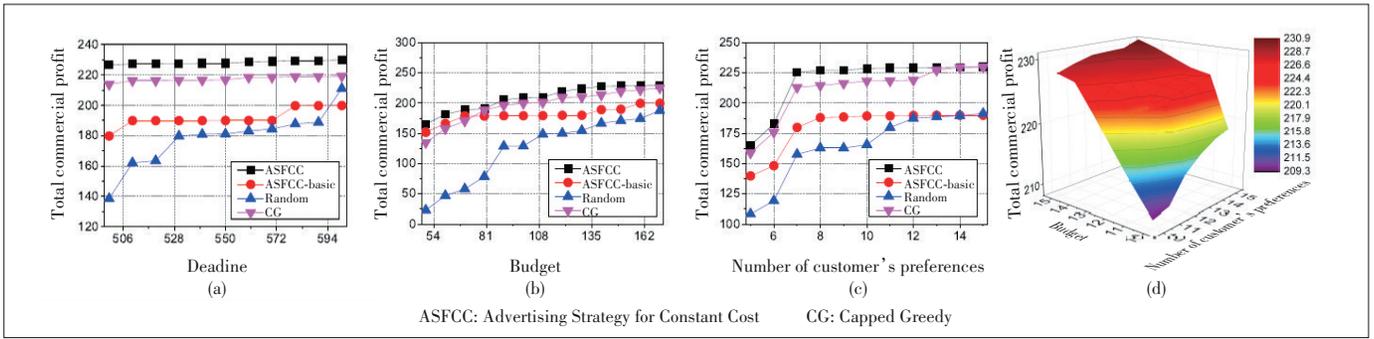
We use the commercial profit as the metric to measure the performances of different advertising strategies. When an advertising strategy performs better, it would have higher commercial profit, which is reasonable. In order to calculate the commercial profit, we need to judge whether a customer is attracted to purchase a product. When a potential customer sees an advertisement on a billboard, he/she has a chance to buy the product in the advertisement. After the deadline of the whole experiment, each potential customer has different purchase probabilities for each product. Through these probabilities, we can use a random number generator to repeatedly test whether a potential customer has bought products, and finally we can get the commercial profit by averaging the payment of potential customers.

## 6.3 Simulation Results for Single-Product with Constant Billboard Cost

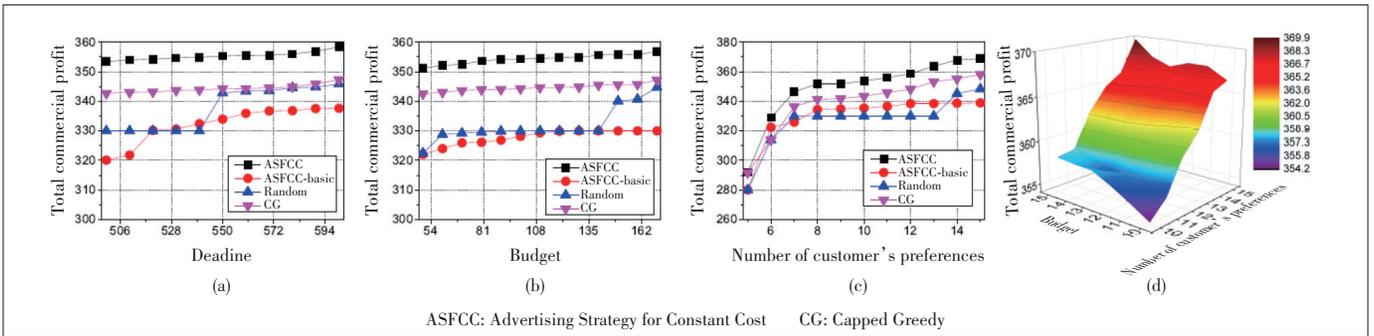
In this section, we aim to test the performance of the proposed strategy ASFCC, when all available billboards have the same cost. In this situation, the advertiser only needs to determine which  $k$  billboards to select for advertising. Hence, we compare our advertising strategy ASFCC with ASFCC-Basic, Random and CG on three datasets. The results of the simulation are shown in Figs. 3, 4 and 5.

First of all, we evaluate the performances of the above four strategies on the roma/taxi trace set. As illustrated in Fig. 3, we investigate the influence of the four variables on the total commercial profit of different strategies. Obviously, ASFCC can achieve the maximum commercial profit for the advertiser, while the performance of Random is the worst. We can find that the total commercial profit increases with the growth of deadline, which represents the duration of billboard advertising. It is reasonable for this phenomenon, because when the deadline increases, the selected billboards may have more chances to attract the potential customers so that the commercial profit may increase. We can also find that the total commercial profit shows an upward trend with the increase of budget. The reason is that with the budget increasing, the advertiser can select more billboards for advertising so that more potential customers may be attracted. By changing the number of customers' preferences, we can see that the performances of ASFCC and CG are very close but far better than those of ASFCC-Basic and Random.

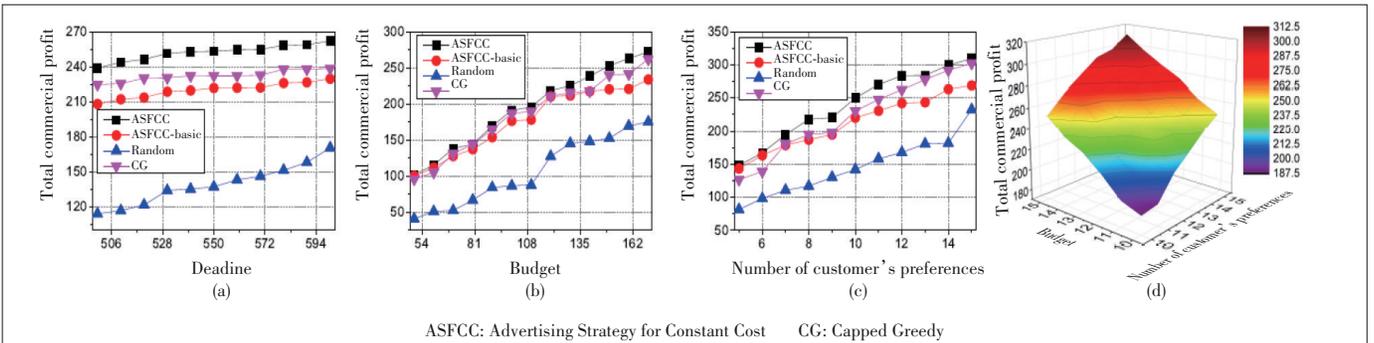
Next, we compare the performances of different advertising



▲ Figure 3. Performances on the roma/taxi trace set, when all billboards have the same cost



▲ Figure 4. Performances on the epfl trace set, when all billboards have the same cost



▲ Figure 5. Performances on the geolife trace set, when all billboards have the same cost

strategies on the epfl trace set, which are shown in Fig. 4. As we can see from Fig. 4a, the performance of ASFCC is far better than that of other strategies. In particular, the strategy Random outperforms the strategy ASFCC-Basic when the deadline is set to 550 - 600, which can prove the limitation of the local optimum of ASFCC-Basic. The similar phenomenon can also be seen in Fig. 4b, where the performance of Random is better than that of ASFCC-Basic. We also test the influence of budget and the number of customers' preferences on ASFCC's performance at the same time and the results are shown in Fig. 4d. It is obvious that the performance of ASFCC shows an upward trend when two variables increase, which is also consistent with the performances in Figs. 4b and 4c.

In Fig. 5, we show the performances of different strategies on geolife traces. We rank the performances of different strategies as follows: ASFCC > CG > ASFCC-Basic > Random. It is reasonable because the instability of the Random leads to poor

performance. At the same time, CG and ASFCC-Basic have their limitations, resulting in the final results not as good as ASFCC. Based on the experimental results of the three datasets, we can find that our strategy ASFCC can always achieve the optimal results under different conditions.

Finally, we conduct the simulations to verify the correctness of the approximation ratio for ASFCC. As shown in Table 2, the results of ASFCC when the deadline is from 500 to 550 are obviously larger than  $(1 - \frac{1}{e})$ Optimal, which are consistent with our theoretical analysis.

#### 6.4 Simulation Results for Single-Product with Different Billboard Costs

In this part, we conduct the simulations to compare the performance of ASFDC with other two strategies when all billboards have different costs. The results are shown in Figs. 6, 7 and 8.

First of all, we compare the performances of different advertising strategies on the roma/taxi trace set, and the results are shown in Fig. 6. By changing the deadline, we can find that the strategy ASFDC outperforms the other two strategies. Because ASFDC selects the billboards with the consideration of potential customers' preferences, detour distance and the probabilities of seeing the billboards, thus the billboards selected by ASFDC can achieve better commercial profit. Note that it is reasonable that the commercial profit may increase when the budget grows, because the advertiser may select more billboards for advertising so that more potential customers can be attracted. The results in Fig. 6b can match our analysis.

Next, we conduct the simulations on the epfl trace set and the results are shown in Fig. 7. Obviously, we can find that the ASFDC performs much better than ASFDC-Basic and Random. Because ASFDC-Basic selects the billboards with maximum commercial profit which is a local optimum, the billboards selected by ASFDC-Basic may be seen by a small group of potential customers so that the commercial profit cannot be maximized. We also test the performance difference between ASFDC and Random when the costs of billboards are under different distributions (uniform distribution, poisson distribution and normal distribution), which are shown in Fig. 7d. The difference values in Fig. 7d are calculated by subtracting the profit of strategy Random from that of strategy ASFDC. The larger the difference value is, the greater the performance gap between these two strategies would be. As we can see from Fig. 7d, it is clear that the difference value decreases as the budget increases. When the costs of billboards are under uniform distribution, the difference between ASFDC and Random is the greatest, and when the costs of billboards are under normal distribution, the difference is the smallest. However, this difference is small among the three distributions when the budget is same. Hence, our proposed strategy ASFDC can be adopted when the costs of billboards are under these three distributions.

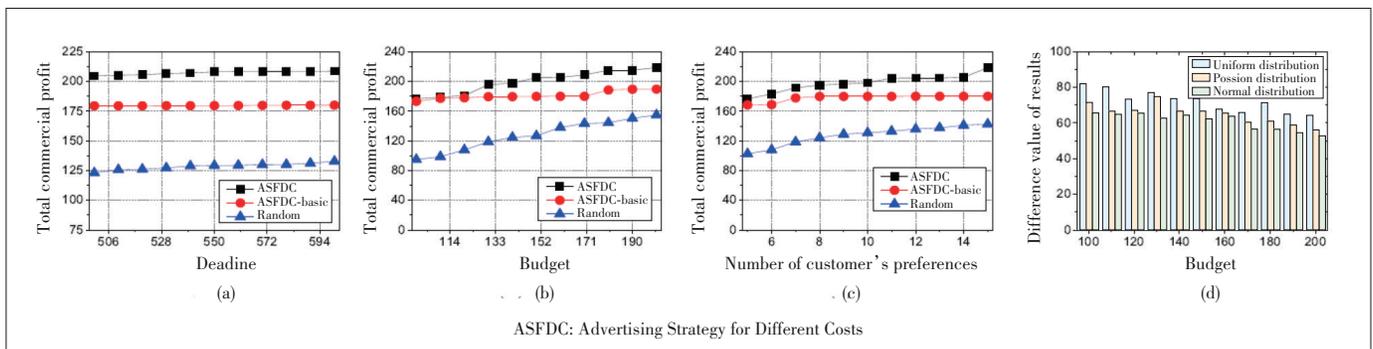
We then evaluate the performances of three strategies: ASFDC, ASFDC-Basic and Random on the geolife traces, which are shown in Fig. 8. We can find that similar phenomenon as that in Figs. 6 and 7 also appears in Fig. 8, where the performance of ASFDC is much better than other two strategies. The reason for this phenomenon is similar to that in Figs. 6 and 7, so it is omitted here.

Finally, we conduct simulations to verify the correctness of the approximation ratio for ASFDC on epfl trace set where the deadline is set from 500 to 550, and the results are shown in Table 3. Compared with the results of  $(1 - \frac{1}{e})optimal$ , we can easily see that the results of ASFDC are larger, which matches

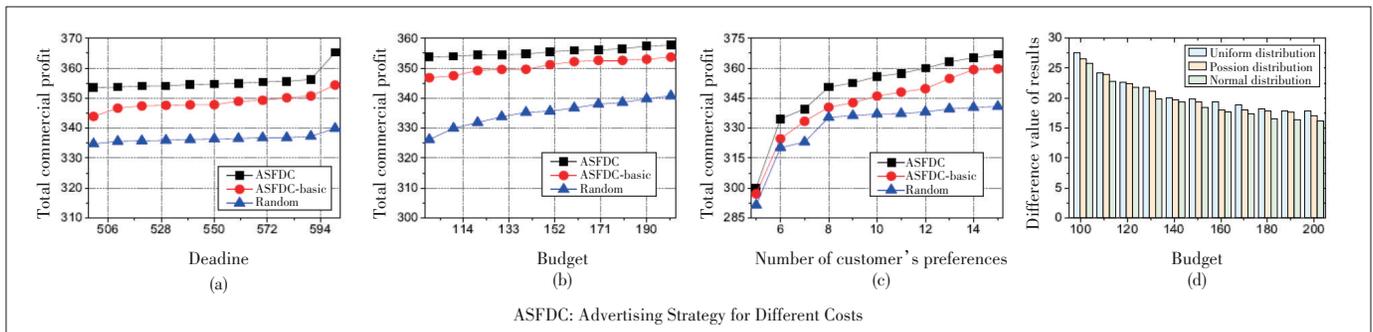
▼ Table 2. Simulation results on epfl, when all billboards have the same cost

Algorithm	Deadline					
	500	510	520	530	540	550
Optimal	37.06	37.27	37.71	37.85	37.87	37.91
$(1 - \frac{1}{e})Optimal$	23.42	23.56	23.84	23.92	23.94	23.96
ASFCC	35.34	35.39	35.42	35.48	35.52	35.54

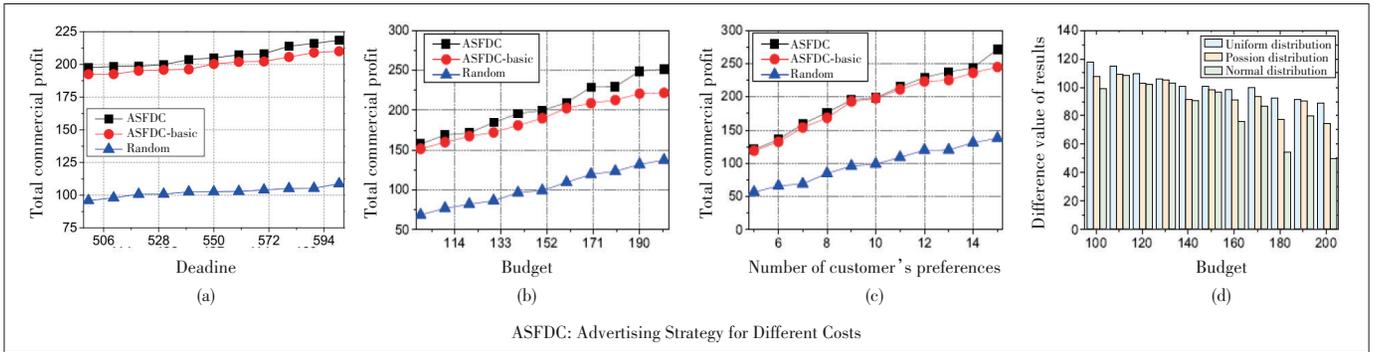
ASFCC: Advertising Strategy for Constant Cost



▲ Figure 6. Performances on the roma/taxi trace set, when all billboards have different costs



▲ Figure 7. Performances on the epfl trace set, when all billboards have different costs



▲ Figure 8. Performances on the geolife trace set, when all billboards have different costs

▼ Table 3. Simulation results on epfl, when all billboards have different costs

Algorithm	Deadline					
	500	510	520	530	540	550
Optimal	37.38	37.41	37.59	37.62	38.02	38.71
$(1 - \frac{1}{e})$ Optimal	23.63	23.65	23.76	23.78	24.03	24.47
ASFCC	35.37	35.40	35.44	35.45	35.47	35.54

ASFCC: Advertising Strategy for Constant Cost

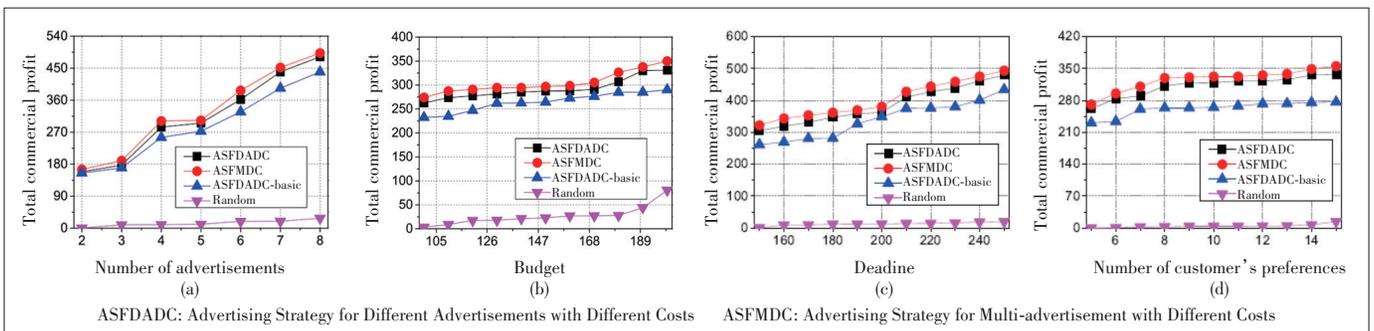
the theoretical analysis.

### 6.5 Simulation Results for Multi-Product

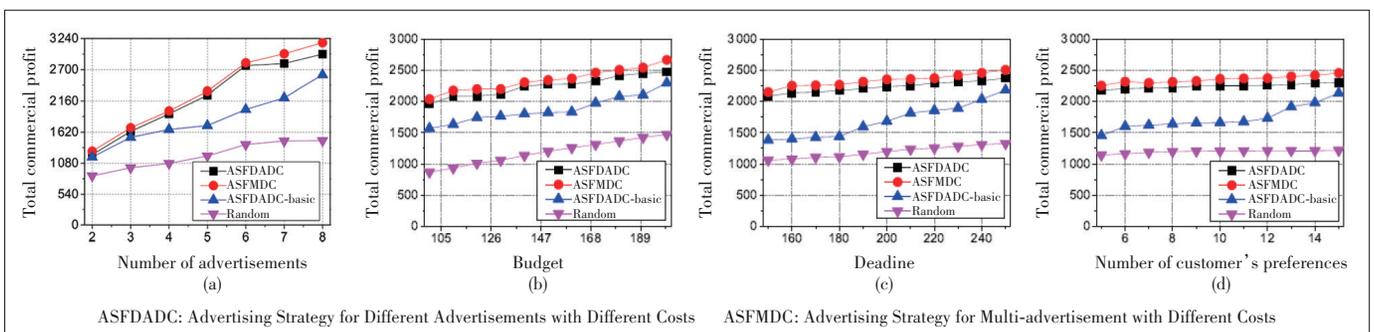
In this part, we conduct the simulations to compare the performance of ASFMDC with other three advertising strategies: ASFDADC, ASFDADC-Basic and Random. The detailed results are shown in Figs. 9, 10 and 11.

Firstly, as shown in Fig. 9, we conduct the simulations on

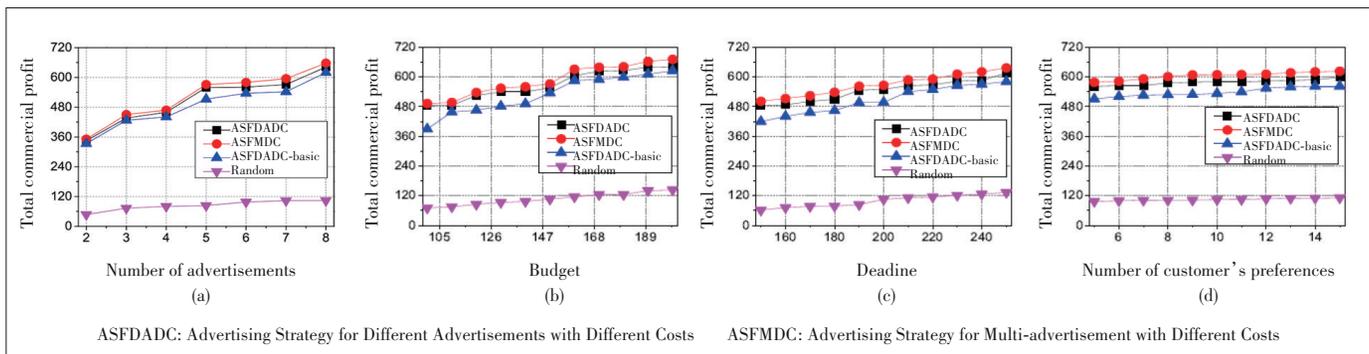
the roma/taxi trace set to compare the performances of different strategies. We can rank the performances of different strategies as follows: ASFMDC > ASFDADC > ASFDADC-Basic > Random, which can prove the effectiveness of our proposed strategy ASFMDC. We can also find that the result of ASFMDC is improved by about 5% - 10% compared with ASFDADC, where ASFMDC is based on ASFDADC. The reason is that ASFDADC selects the billboards which can maximize the total expected commercial profit, while ASFMDC conducts a search in different directions based on ASFDADC, selecting the optimal neighbor solution as the final result. Thus, our strategy can improve 5% to 10% compared with ASFDADC, which is reasonable. The performance of Random is much worse than that of the previous two experiments. Because the advertiser needs to select billboards and determines their corresponding advertisements, Random introduces a lot of uncertainties, which makes the result much worse than the other



▲ Figure 9. Performances on the roma/taxi trace set, when the advertiser has multiple products



▲ Figure 10. Performances on the epfl trace set, when the advertiser has multiple products



▲ Figure 11. Performances on the geolife trace set, when the advertiser has multiple products

strategies.

Then, we show the results of the simulations on the epfl trace set in Fig. 10. ASFMDC can achieve the best commercial profit for the advertiser and the performance of ASFDADC is close to ASFMDC. However, the performances of the above two strategies are far better than ASFDADC-Basic, which is different from the phenomenon in Fig. 9. Because the billboards selected by ASFDADC-Basic may achieve a local optimum, when the global optimal solution is very different from the local optimal solution, the result of ASFDADC-Basic would be very bad. In addition, we can find that changing the number of advertisements has a greater impact on the results. This is because a potential customer can be attracted by different products, when the number of advertisements increases, potential customers can be attracted to buy new products and the commercial profit is improved significantly.

Finally, Fig. 11 shows the performances of different strategies on the geolife traces. We can find that the phenomenon in Fig. 11 is similar to that in Fig. 9, where ASFMDC performs better than other three strategies. The performances of ASFDADC and ASFDADC-Basic are close to ASFMDC and much better than that of Random. It is not difficult to find that the results of different strategies increase with the growth of the four variables in Fig. 11, which is reasonable. Because the growth of budget and deadline would increase the number of the selected billboards and the probabilities that selected billboards would be seen by the customers. When the number of products and customers' preferences grow, the potential customers who see the advertisements would be more likely to be attracted. From the above figures, we can prove that our advertising strategy ASFMDC can bring more commercial profit to the advertiser, compared with other common strategies.

## 7 Conclusions

In this paper, a billboard selection problem is formulated in order to maximize the commercial profit for the advertiser under the limited budget. In order to address this problem, first of all, we adopt MCS to collect potential customers' traces and preferences. Then, we use the semi-markov model to pre-

dict the customers' mobility patterns. Next, we quantify the utility of each billboard with the consideration of customers' preferences, detour distance and the probabilities of seeing the billboards. Two heuristic advertising strategies are proposed in this paper to determine which billboards to select for the situation where the advertiser has only one type of product. Then, we adopt the simulated annealing algorithm to address this problem when the advertiser has multiple products. We conduct the extensive simulations based on the widely-used real-world trajectories: roma/taxi, epfl, and geolife. The results show that our advertising strategies can bring the best commercial profit for the advertiser compared with other advertising strategies.

## References

- [1] QUINN P. Global digital OOH media revenues pacing up 13% in 2017, US DOOH advertising expands 10%: PQ media [EB/OL]. (2017-09-06) [2020-01-01]. <https://digitalsignagepulse.com/news/global-digital-oooh-media-revenues-pacing-up-13-in-2017-us-doooh-advertising-expands>
- [2] LOU K H, LI S Q, YANG F N, et al. Advertising strategy for maximizing profit using Crowdsensing trajectory data [C]// Security and Privacy in Social Networks and Big Data. Singapore: Springer Singapore, 2020: 395 - 406. DOI: 10.1007/978-981-15-9031-3\_35
- [3] NIGAM S, ASTHANA S, GUPTA P. IoT based intelligent billboard using data mining [C]//International Conference on Innovation and Challenges in Cyber Security (ICICCS-INBUSH). Greater Noida, India: IEEE, 2016: 107 - 110
- [4] LIU D Y, WENG D, LI Y H, et al. SmartadP: visual analytics of large-scale taxi trajectories for selecting billboard locations [J]. IEEE transactions on visualization and computer graphics, 2017, 23(1): 1 - 10. DOI: 10.1109/TVCG.2016.2598432
- [5] HUANG M, FANG Z X, XIONG S L, et al. Interest-driven outdoor advertising display location selection using mobile phone data [J]. IEEE access, 2019, 7: 30878 - 30889. DOI: 10.1109/ACCESS.2019.2903277
- [6] WANG L, YU Z W, YANG D Q, et al. Efficiently targeted billboard advertising using crowdsensing vehicle trajectory data [J]. IEEE transactions on industrial informatics, 2020, 16(2): 1058 - 1066. DOI: 10.1109/TII.2019.2891258
- [7] ZHENG H Y, WU J. Placement optimization for advertisement dissemination in smart City [J]. IEEE transactions on network science and engineering, 2020, 7(1): 239 - 252. DOI: 10.1109/TNSE.2018.2805768
- [8] GANTI R K, YE F, LEI H. Mobile crowdsensing: current state and future chal-

- lenges [J]. *IEEE communications magazine*, 2011, 49(11): 32 – 39. DOI: 10.1109/MCOM.2011.6069707
- [9] LIU J W, SHEN H Y, ZHANG X. A survey of mobile crowdsensing techniques: a critical component for the Internet of Things [C]//25th International Conference on Computer Communication and Networks (ICCCN). Waikoloa, USA: IEEE, 2016: 1 – 6. DOI: 10.1109/ICCCN.2016.7568484
- [10] CAPPONI A, FIANDRINO C, KANTARCI B, et al. A survey on mobile crowdsensing systems: challenges, solutions, and opportunities [J]. *IEEE communications surveys & tutorials*, 2019, 21(3): 2419 – 2465. DOI: 10.1109/COMST.2019.2914030
- [11] KARALIOPOULOS M, KOUTSOPOULOS I, TITSIAS M. First learn then earn: optimizing mobile crowdsensing campaigns through data-driven user profiling [C]//Proceedings of the 17th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '16). New York, USA: ACM, 2016: 271 – 280
- [12] BRACCIALE L, BONOLA M, LORETI P, et al. CRAWDAD dataset roma/taxi [EB/OL]. (2014-07-17) [2020-01-01]. <https://crawdad.org/roma/taxi/20140717>
- [13] PIORKOWSKI M, SARAFIJANOVIC - DJUKIC N, GROSSGLAUSER M. CRAWDAD dataset epfl/mobility [EB/OL]. (2009-02-24)[2020-01-01]. <https://crawdad.org/epfl/mobility/20090224>
- [14] ZHENG Y, ZHANG L, XIE X, et al. Mining interesting locations and travel sequences from GPS trajectories [C]//Proceedings 13 of the 18th International Conference on World Wide Web (WWW '09). New York, USA: ACM, 2009: 791 – 800
- [15] KRISHNA O, AIZAWA K. Billboard saliency detection in street videos for adults and elderly [C]//25th IEEE International Conference on Image Processing (ICIP). Athens, Greece: IEEE, 2018: 2326 – 2330
- [16] AN T T, CHANG C P, LI Y H, et al. Fog computing architecture-based Wi-Fi union mechanism for Internet advertising system [C]//International Conference on Applied System Innovation (ICASI). Sapporo, Japan: IEEE, 2017: 1024 – 1027. DOI: 10.1109/ICASI.2017.7988110
- [17] ZHANG Y P, LI Y C, BAO Z F, et al. Optimizing impression counts for outdoor advertising [C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York, USA: ACM, 2019: 1205 – 1215. DOI: 10.1145/3292500.3330829
- [18] ARIYA SANJAYA I M, SUPANGKAT S H, SEMBIRING J. Citizen reporting through mobile crowdsensing: a smart city case of bekasi [C]//International Conference on ICT for Smart Society (ICISS). Semarang, Indonesia: IEEE, 2018: 1 – 4. DOI: 10.1109/ICTSS.2018.8549976
- [19] CHEUNG M H, HOU F, HUANG J W. Delay-sensitive mobile crowdsensing: Algorithm design and economics [J]. *IEEE transactions on mobile computing*, 2018, 17(12): 2761 – 2774. DOI: 10.1109/TMC.2018.2815694
- [20] CAO B, XIA S C, HAN J W, et al. A distributed game methodology for crowdsensing in uncertain wireless scenario [J]. *IEEE transactions on mobile computing*, 2020, 19(1): 15 – 28. DOI: 10.1109/TMC.2019.2892953
- [21] GONG W, ZHANG B X, LI C. Location-based online task assignment and path planning for mobile crowdsensing [J]. *IEEE transactions on vehicular technology*, 2019, 68(2): 1772 – 1783. DOI: 10.1109/TVT.2018.2884318
- [22] MARJANOVIĆ M, ANTONIĆ A, ŽARKO I P. Edge computing architecture for mobile crowdsensing [J]. *IEEE access*, 2018, 6: 10662 – 10674. DOI: 10.1109/ACCESS.2018.2799707
- [23] ZHENG Y F, DUAN H Y, YUAN X L, et al. Privacy-aware and efficient mobile crowdsensing with truth discovery [J]. *IEEE transactions on dependable and secure computing*, 2020, 17(1): 121 – 133. DOI: 10.1109/TDSC.2017.2753245
- [24] WANG L, YU Z W, ZHANG D Q, et al. Heterogeneous multi-task assignment in mobile crowdsensing using spatiotemporal correlation [J]. *IEEE transactions on mobile computing*, 2019, 18(1): 84 – 97. DOI: 10.1109/TMC.2018.2827375
- [25] WANG J T, WANG Y S, ZHANG D Q, et al. PSAllocator: multi-task allocation for participatory sensing with sensing capability constraints [C]//Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. New York, USA: ACM, 2017: 1139 – 1151. DOI: 10.1145/2998181.2998193
- [26] LIN M, HSU W -J, LEE Z Q. Predictability of individuals' mobility with high-resolution positioning data [C]//Proceedings of the 2012 ACM Conference on Ubiquitous Computing (UbiComp '12). New York, USA: ACM, 2012: 381 – 390
- [27] WANG E, YANG Y J, WU J, et al. An efficient prediction-based user recruitment for mobile crowdsensing [J]. *IEEE transactions on mobile computing*, 2018, 17(1): 16 – 28. DOI: 10.1109/TMC.2017.2702613
- [28] YUAN Q, CARDEI I, WU J. An efficient prediction-based routing in disruption-tolerant networks [J]. *IEEE transactions on parallel and distributed systems*, 2012, 23(1): 19 – 31. DOI: 10.1109/TPDS.2011.140
- [29] YANG Y J, LIU W B, WANG E, et al. A prediction-based user selection framework for heterogeneous mobile Crowdsensing [J]. *IEEE transactions on mobile computing*, 2019, 18(11): 2460 – 2473. DOI: 10.1109/TMC.2018.2879098
- [30] YANG Y J, XU Y B, WANG E, et al. Exploring influence maximization in on-line and offline double-layer propagation scheme [J]. *Information sciences*, 2018, 450: 182 – 199. DOI: 10.1016/j.ins.2018.03.048
- [31] KHULLER S, MOSS A, NAOR J S. The budgeted maximum coverage problem [J]. *Information processing letters*, 1999, 70(1): 39 – 45. DOI: 10.1016/s0020-0190(99)00031-9

### Biographies

**LOU Kaihao** (loukh20@mails.jlu.edu.cn) received the B.E. degree in software engineering from Jilin University, China in 2017, M.S. degree in computer science and technology from Jilin University in 2020. He is currently pursuing the Ph.D. degree in computer science and technology at Jilin University. His current research focuses on mobile crowdsensing and multi-agent reinforcement learning.

**YANG Yongjian** received his B.E. degree in automatization from Jilin University of Technology, China in 1983, M.E. degree in computer communication from Beijing University of Post and Telecommunications, China 1991, and Ph. D. in software and theory of computer from Jilin University, China in 2005. He is currently a professor and a Ph. D. supervisor at Jilin University, the Vice Dean of the Software College of Jilin University, Director of Key lab under the Ministry of Information Industry, Standing Director of the Communication Academy, and a member of the Computer Science Academy of Jilin Province. His research interests include network intelligence management, wireless mobile communication and services, and wireless mobile communications.

**YANG Funing** received her B.E. degree in software engineering from Jilin University, China in 2010 and master's degree from the school of computer science, Beijing University of Posts and Telecommunications, China in 2013. She is currently a teacher at Jilin University, China. Her current research interests include management, mobile crowdsensing, integration and mining of massive traffic data.

**ZHANG Xingliang** received his B.E. degree in software engineering from Jilin University, China in 2010. He is currently an engineer at the Network Management Center of China Mobile Group Jilin Co., Ltd. His current research interests include mobile communication data analysis and big data analysis of 5G.