Lou, K, Yang, Y, Wang, E, Liu, Z, Baker, T and Bashir, AK

Reinforcement Learning Based Advertising Strategy Using Crowdsensing Vehicular Data

Reinforcement Learning Based Advertising Strategy Using Crowdsensing Vehicular Data

Kaihao Lou, Yongjian Yang, En Wang*, Zheli Liu, Thar Baker, Ali Kashif Bashir

***Abstract*—As an effective tool, roadside digital billboard advertising is widely used to attract potential customers (e.g., drivers and passengers passing by the billboards) to obtain commercial profit for the advertiser, i.e., the attracted customers' payment. The commercial profit depends on the number of attracted customers, hence the advertiser needs to adopt an effective advertising strategy to determine the advertisement switching policy for each digital billboard to attract as many potential customers as possible. Whether a customer could be attracted is influenced by numerous factors, such as the probability that the customer could see the billboard and the degree of his/her interests in the advertisement. Besides, cooperation and competition among all digital billboards will also affect the commercial profit. Taking the above factors into consideration, we formulate the dynamic advertising problem to maximize the commercial profit for the advertiser. To address the problem, we first extract potential customers' implicit information by using the vehicular data collected by Mobile CrowdSensing (MCS), such as their vehicular trajectories and their preferences. With this information, we then propose an advertising strategy based on multi-agent deep reinforcement learning. By using the proposed advertising strategy, the advertiser could determine the advertising policy for each digital billboard and maximize the commercial profit. Extensive experiments on three real-world datasets have been conducted to verify that our proposed advertising strategy could achieve the superior commercial profit compared with the state-of-the-art strategies.**

***Index Terms*—Digital Billboard Advertising, Multi-agent Deep Reinforcement Learning, Crowdsensing Vehicular Data**

I. INTRODUCTION

Digital roadside billboard is one of the most effective tools for advertising. According to PQ Media [1], global digital roadside billboard advertising industry has grown by a large margin in 2017. Specially, digital roadside billboard advertising sales have increased by 10% to a total amount of 3.2 billion dollars in US. Compared with traditional static roadside billboard, the digital roadside billboard can easily make a deeper impression on potential customers (e.g., the drivers and passengers), since it can dynamically deliver graphic advertising content (e.g., images and videos). In this way, digital roadside billboards can deepen potential customers' impression of products, and achieve the purpose of increasing sales volume.

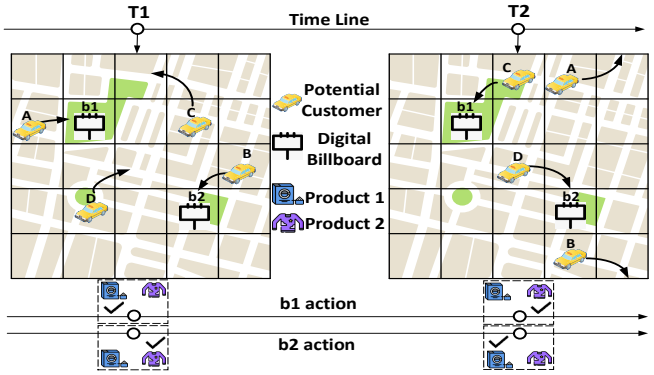


Fig. 1: An example of the dynamic advertising problem.

By advertising on roadside digital billboards, an advertiser could attract potential customers driving the vehicles or the passengers for his/her products. Once a potential customer is attracted by the advertisement on the digital billboard, he would purchase the relative product and the advertiser would obtain the commercial profit. Hence, the commercial profit depends on the number of attracted customers. To maximize the commercial profit, the advertiser needs to attract the potential customers as many as possible. However, whether a potential customer could be attracted by the digital billboards is determined by many factors, such as the customer's mobility (whether he can see the billboard) and the customer's preferences (whether he is interested in the product). For example, students are more likely to see billboards near the school and less likely to see billboards on the highway. Besides, students could be more interested in sports clothing and less interested in business clothing. Hence, the advertiser should adopt appropriate advertising strategies to decide which advertisement should be delivered on each digital billboard to attract as many potential customers as possible.

Most of the existing advertising strategies focus on what advertising content should be delivered and how to select locations for the static roadside billboards. By using the data collected by RFID, S. Nigam *et al.* in [2], decide the locations of billboards. In [3] and [4], the billboard locations are determined by using GPS and phone data. Besides, advertising

Manuscript received January 28, 2020; revised March 24, 2020; accepted April 10, 2020. This work is supported by the National Natural Science Foundations of China under Grant No. 61772230 and No. 61972450, Natural Science Foundation of China for Young Scholars No. 61702215, China Postdoctoral Science Foundation No. 2017M611322 and No. 2018T110247, and Changchun Science and Technology Development Project No.18DY005. (*Corresponding author: En Wang)

Kaihao Lou, Yongjian Yang and En Wang, Department of Computer Science and Technology, Jilin University, Changchun, Jilin, 130012, China. (e-mail: loukh17@mails.jlu.edu.cn; yyj@jlu.edu.cn; wangen@jlu.edu.cn)

Zheli Liu, College of Cyber Science, College of Computer Science and Tianjin Key Laboratory of Network and Data Security Technology, Nankai University, Tianjin, 300071, China. (e-mail: liuzheli@nankai.edu.cn)

Thar Baker, Department of Computer Science, Liverpool John Moores University, UK. (e-mail: t.baker@ljmu.ac.uk)

Ali Kashif Bashir, Department of Computing and Mathematics, Manchester Metropolitan University, UK. (e-mail: dr.alikashif.b@ieee.org)

content on the billboard is determined by the preferences of potential customers and the detour distance in [5] and [6]. In the real world, the potential customers passing the same billboard location change over time, and hence the traditional static roadside billboards do not perform well. For instance, during lunch or dinner time, there are many hungry customers, the billboards in shopping malls should place advertisements for food. However, at other time advertisements for clothes may be a good choice.

In order to maximize the commercial profit for the advertiser, we decide to use the digital roadside billboards in this paper. And according to the preferences of passing potential customers, the digital billboards should switch their advertising content to attract as many potential customers as possible. How to switch advertisement dynamically according to the situation is the problem of this paper, which is called dynamic advertising problem. For example, consider a dynamic advertising scene, which is shown in Fig. 1. There are two available digital billboards (b1 and b2) and four different potential customers (A, B, C and D) driving their vehicles in this area. The advertiser wants to do advertising for his/her two different products (product 1 and 2) to obtain commercial profit. The commercial profit is quantified by the payments of the attracted customers, hence the advertiser needs to attract as many potential customers as possible. As shown in Fig. 1, at time T1, two potential customers A and B are about to pass through the digital billboards b1 and b2. Suppose that customer A prefers the product 1 and customer B is interested in product 2. Thus, digital billboards b1 and b2 should deliver the advertising content about product 1 and product 2, respectively. Then, at time T2, the potential customers C and D are about to pass through the digital billboards b1 and b2. Suppose that customer C is interested in the product 2 and customer D is interested in the product 1. Obviously, the two digital billboards b1 and b2 should deliver the advertising content of product 2 and product 1 respectively when the customers C and D are passing these digital billboards.

To solve the above dynamic advertising problem, firstly, we need to know the preferences of the passing potential customers and their mobility patterns, which are privacy-sensitive. Besides, cooperation and competition among all digital billboards will also affect the commercial profit. For example, the two digital billboards of A and B are close to each other, the commercial profit may not increase much when they advertise for the same product. Because customers passing these two billboards may be the same and the commercial profit will be improved more if the two billboards advertise for different products. We may solve this challenge by using a central server to control the digital billboards, but it costs considerable resource. Traditional random strategy does not perform well. These two challenges greatly limit the research about digital billboard advertising and reinforcement learning [7, 8] may be a good method to solve this problem.

In this paper, we adopt Mobile CrowdSensing (MCS) [9–11] to gather the privacy-sensitive customer profiles [5, 12] such as their vehicular trajectories and preferences. For example, a MCS application may record a user's vehicular trajectories when the user finishes some sensing tasks. Moreover, the user's completed task history can also be used to infer the user's preferences [12]. Suppose there is a user, who has driven to or has finished tasks near the food market for many times, then we can infer that this user may be attracted by the food advertisements, in other words, the food market could be considered as a preference of this user. And we use a semi-markov method to predict customers' mobility pattern. With this information, we propose an advertising strategy based on a multi-agent approach called multi-agent deep deterministic policy gradient (MADDPG) [13] to derive advertisement switching policy for each digital billboard.

The main contributions of this paper are summarized as follows:

- We formulate a dynamic advertising problem to determine how to switch the advertising content for each digital billboard at different time slice so that the advertiser could achieve the maximal commercial profit.
- We propose an effective advertising strategy based on the multi-agent deep deterministic policy gradient to solve the dynamic advertising problem by using the vehicular data collected by mobile crowdsensing.
- We conduct extensive simulations based on three real-world trajectories: *roma/taxi*[14], *epfl*[15], and *geo-life*[16]. The results show that our advertising strategies could achieve superior commercial profit for the advertiser compared with other strategies.

The remainder of this paper is organized as follows. We review the related work in Section II. We describe the system models and define the dynamic advertising problem in Section III. We describe the general technologies in Section IV. The detailed advertising strategy based on multi-agent deep reinforcement learning is proposed in Section V. In Section VI, we compare the performances of our adverting strategy with other advertising strategies by conducting simulations. We conclude this paper in Section VII.

II. RELATED WORK

In this section, we review various related work on advertising strategy on billboard, multi-agent deep reinforcement learning and mobile crowdsensing.

Advertising Strategy. There have been many works on advertising strategy. Most of them are about how to select the locations of the roadside billboards or how to select the advertisements on the roadside billboards. In [2], S. Nigam $et\ al.$ propose an intelligent advertising system for multi retail stores, malls and shopping complexes. It analyses data collected by RFID tags in order to provide better offers and deals to customers, which are attached to products. This system also helps to select the locations for the billboards. In [3], D. Liu $et\ al.$ propose an interactive visual analytics system that combines the state-of-the-art mining and visualization techniques with large-scale GPS trajectory data for billboard placements. M. Huang $et\ al.$ propose a methodology in [4], in order to solve the problem of displaying the relevant advertisements (ads) and maximizing their coverage. By using the data collected by phones, they can identify the interests and mobility patterns of

individuals. These works are about how to select the locations for the roadside billboards by analyzing the data collected by different ways. In [17], T. T. An $et\ al.$ design an advertisement system by using Wi-Fi union mechanism in order to enhance the efficiency of advertisement. In [5], L. Wang $et\ al.$ design a model to quantify advertisement influence spread and propose a utility evaluation-based optimal searching approach so that the total expected advertisement influence spread could be maximized. In [6], H. Zheng $et\ al.$ investigate a promising application for Vehicular Cyber-Physical Systems (VCPS). They propose bounded RAP placement algorithms to maximally attract potential customers for the shopkeeper. These works aim at improving the influence of advertisements by selecting the advertisement on the billboards. In [18], Zhang, Yipeng $et\ al.$ optimize the influence of outdoor advertising (ad) with the consideration of impression counts. They propose a tangent line based algorithm to select roadside billboards for maximizing the influence of outdoor advertising.

Multi-Agent Deep Reinforcement Learning. There have also been many works about multi-agent deep reinforcement learning [19, 20]. The multi-agent deep deterministic policy gradient (MADDPG) is proposed in [13], Lowe $et\ al.$ present this method for cooperative or competitive scenarios which takes the action policies of other agents into consideration. In [21], T. Chu $et\ al.$ propose a fully scalable and decentralized MARL algorithm for large-scale traffic signal control which could achieve the best performance in simulations compared with other MARL algorithms. In [22], S. Zheng $et\ al.$ propose the improved Multi-Agent Deep Deterministic Policy Gradient (MADDPG) algorithm for large-scale crowd path planning. In [23], Y. Pan $et\ al.$ propose a novel method by using the centralized training and distributed execution with parameter sharing among homogeneous agents, so that the partial calculation of network parameters in policy evolution can be substituted.

Mobile Crowdsensing. Mobile crowdsensing has been extensively studied in recent years. For the task allocation, in [24], J. Wang $et\ al.$ proposes 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. J. Wang $et\ al.$ [25] study multi-task allocation problem and propose a novel multi-task allocation framework named MTasker to maximize the overall system utility. In [26],J. Wang $et\ al.$ propose a two-phased hybrid framework called HyTasker, which jointly optimizes two phases with a total incentive budget constraint. For the worker recruitment problem, J. Wang $et\ al.$ [27] study the worker recruitment problem and propose two algorithms to leverage the influence propagation on the social network and assist the MCS worker recruitment.

Compared with these existing research works, in this paper, we focus the problem of advertising on the digital billboards. The digital billboards need to decide what the advertising content to deliver for achieving maximal commercial profit. Hence, we propose an effective advertising strategy based on the multi-agent deep reinforcement learning to maximize the commercial profit for the advertiser using crowdsensing vehicular data.

III. SYSTEM OVERVIEW

This section first discusses the system model of the dynamic advertising problem. Then, we define the dynamic advertising problem.

A. System Model

The system model of the dynamic advertising is firstly discussed in this section. Considering there are n potential customers $C = \{c_1, c_2, ..., c_n\}$ moving in the area $L = \{l_1, l_2, ..., l_h\}$. An advertiser owns m digital billboards $B = \{b_1, b_2, ..., b_m\}$ and these billboards are located at different locations $L_B = \{l_{b_1}, l_{b_2}, ..., l_{b_m}\} \subseteq L$. The advertiser also has a series of products for advertising, which can be denoted as $A = \{a_1, a_2, ..., a_k\}$. The attribute types of products and customers' preferences are denoted by the set $Attr = \{attr_1, attr_2, ..., attr_j\}$. Without loss of generality, we denote the preferences of a potential customer c_i as $Attr_{c_i}$, and $Attr_{c_i} \subseteq Attr$. Similarly, the attributes of a product a_d can be denoted as $Attr_{a_d}$, and $Attr_{a_d} \subseteq Attr$.

During the whole advertising life cycle, each potential customer moves in the area. A potential customer would see the advertising content on the digital billboard once he enters an area, which contains a digital billboard. The potential customer could be attracted by the advertisement and purchase the corresponding product so that the advertiser would obtain the commercial profit. Hence, at the beginning of each time slice, each digital billboard needs to decide what advertising content to deliver from the product set to attract potential customers as many as possible. We suppose that there is no communication among these digital billboards and the communication cost. In this paper, if a potential customer is attracted by an advertisement, then he would buy the corresponding product a_d and the commercial profit for the advertiser is f_{a_d}. And we suppose that if different customers buy the same product a_d, the advertiser will get the same profit f_{a_d} from each attracted customer. In other words, the advertiser needs to attract as many potential customers as possible for different products so that the commercial profit could be maximized.

B. Problem Definition

Based on the above system model, we can define the problem as follows:

Problem (Dynamic Advertising Problem): Given a potential customer set C with their preference set $Attr_C$, a digital billboard set B and a product set A with their attribute set $Attr_A$. At the beginning of each time slice, each billboard should decide what advertising content to deliver from the product set to attract potential customers as many as possible so that the commercial profit for the advertiser could be maximized:

$$\begin{aligned} Maximize\ F = \sum_{i=1}^{C}\sum_{d=1}^{A} \phi_{c_i} f_{a_d}, \\ s.t.\ \phi \in \{0, 1\}, \end{aligned} \tag{1}$$

where F denotes the total commercial profit for the advertiser. And ϕ_{c_i} represents whether a potential customer c_i is attracted.

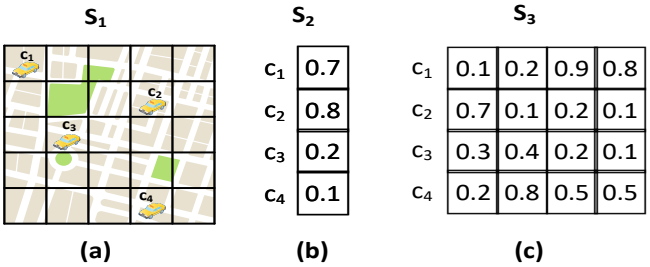


Fig. 2: Observation of each digital billboard.

If the customer c_i is attracted, then $\phi_{c_i} = 1$, otherwise $\phi_{c_i} = 0$. f_{a_d} denotes the commercial profit that the advertiser can get for selling product a_d to a potential customer. Obviously, the advertiser needs to make sure that each potential customer should be attracted by as many advertisements as possible in order to maximize the commercial profit.

IV. PROBLEM FORMULATION

The dynamic advertising problem can be considered as a Markov Decision Process (MDP) [28] [29], which is a common model of reinforcement learning. Briefly speaking, in markov decision process, an agent will repeatedly observe the current state s_t of the environment and take an action a from all available actions in this state. Then, the state of the environment will transfer to s_{t+1} and the agent will get a reward r_t from the environment for its action. In our digital billboard advertising scenario, we suppose that each digital billboard observes the state of the environment, such as the locations of the potential customers and the preferences of the potential customers. Each digital billboard could decide what advertisement to deliver, hence each digital billboard could be considered as an agent in reinforcement learning. Compared with general reinforcement, the digital billboard advertising scene has multiple digital billboards, which can be considered as multiple agents. Hence, the digital billboard advertising scene is a multi-agent reinforcement learning scene. And the MDP is defined as $\mathcal{M} = \{\mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{F}, \gamma\}$, where $\mathcal{S}$ represents the state space, $\mathcal{A}$ represents the action space and $\mathcal{R}$ represents the reward space. γ represents the discount factor and $0 < \gamma < 1$.

A. State Space

First of all, we will discuss the state space in the digital billboard advertising scene. In the digital billboard advertising scenario, the state space is denoted as $\mathcal{S} \triangleq \{s_t = \{\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3\}\}$, which consists of three components (customers' current locations, mobility prediction and customers' preference level). Each billboard b as an agent could observe part of $\mathcal{S}$ as its observation $s_t^i \in s_t$. Next, we will describe the detailed composition of $\mathcal{S}$.

1) Customer Current Location: The first part of $\mathcal{S}$ is the current locations of each potential customers at the beginning of the time slice, which is denoted by $\mathcal{S}_1$. By using the data collected by the GPS on phone, it is not difficult for each digital billboard to collect the current locations of each potential customers. This feature is important for the training process. For example, if a potential customer's current location is in the area, which includes a digital billboard, he would see the advertisement on the digital billboard immediately. Hence, the impact of customers in the billboard area on the commercial profit cannot be ignored. In this paper, we suppose that each digital billboard collects each customer's current location at the beginning of each time slice, which are denoted as $\mathcal{S}_1 = \{l_{c_1}, l_{c_2}, ..., l_{c_n}\}$. For example, as we can see from Fig. 2 (a), there are four potential customers, and their locations can be denoted as $\mathcal{S}_1 = \{l_{c_1}, l_{c_2}, l_{c_3}, l_{c_4}\}$. Each digital billboard would have the same observation about the current locations of each potential customer.

2) Mobility Prediction: The next part of $\mathcal{S}$ is $\mathcal{S}_2$, which denotes the mobility prediction of each potential customer. This is because the advertisement on each digital billboard will last for a period of time (a time slice may include many time units), so it will not only affect the customers who are currently in the billboard area, but also affect the customers who would come the billboard area within the time slice. The commercial profit depends on the number of the attracted customers. Hence, each digital billboard needs to predict the probability of each customer arriving at its located area so that each billboard could estimate the reward more accurately.

We can predict the mobility of each potential customer by using their historical trajectories. First, based on the potential customers' historical vehicular trajectories, we can map their trajectories into a square area in a plane region, especially when the area is small [30]. For example, we can divide the area in the map into a grid-shape like Fig. 1 and convert each potential customer's trajectories to a series of grid coordinates, so that we can reduce the difficulty of calculation. Then we need to find a method to predict each customer's mobility. In this paper, we consider the movement of each potential customer as a markov process, where the next location for each potential customer is only related to the current location. Hence we adopt the semi-markov model [31–33] to predict the customers' mobility. One of the most important equations of semi-markov, $Z(\cdot)$ is defined by Eq. (2).

$$\begin{aligned} Z_u(l_i, l_j, T) =& P(S_u^{n+1} = l_j, t_u^{n+1} - t_u^n \leq T | S_u^0, ..., S_u^n; \\ & t_u^0, ..., t_u^n) \\ =& P(S_u^{n+1} = l_j, t_u^{n+1} - t_u^n \leq T | S_u^n = l_i) \end{aligned} \quad (2)$$

where $Z_u(l_i, l_j, T)$ is the probability that the potential customer u will move from his/her current location l_i to the location l_j at or before time T in the next move. S_u^k denotes the potential customer u's k-th location and the corresponding arrival time for the movement is denoted as t_u^k. As we have discussed that the next location for each potential customer is only related to the current location, we can obtain the transitions from each potential customer's historical trajectories. Then, we can define another key equation $G(\cdot)$, denoted by Eq. (3).

$$G_u(l_i,l_j,T)=\begin{cases}\Sigma_{k=1}^{L}\Sigma_{t=1}^{T}(Z_u(l_i,l_k,t)-Z_u(l_i,l_k,t-1))\cdot \\ G_u(l_k,l_j,T-t), \quad i\neq j \\ 1-\Sigma_{k=1,k\neq i}^{L}Z_u(l_i,l_k,T)+ \\ \Sigma_{k=1,k\neq i}^{L}\Sigma_{t=1}^{T}(Z_u(l_i,l_k,t)-Z_u(l_i,l_k,t-1))\cdot \\ G_u(l_k,l_i,T-t), \quad i=j\end{cases} \tag{3}$$

It is easy to find that the potential customers cannot move from one grid to another when $T = 0$, so we can get $G_u(l_i, l_i, 0) = 1$ and $G_u(l_i, l_j, 0) = 0$ $(i \neq j)$. Next, we calculate the probability of a customer passing any grid l_j before deadline X, as follows:

$$P_{go}^{l_j}(u) = 1 - \prod_{T=t_s}^{X}(1 - G_u(l_i, l_j, T)) \tag{4}$$

where t_s is the time when the customer starts moving for this prediction. By now, we can calculate the probability of a customer passing a billboard area before a certain time. At the beginning of each time slice, each digital billboard b_g would predict the probability of each customer entering its area denoted as $\mathcal{S}_2^{b_g} \in \mathcal{S}_2$, which is a matrix of $n \times 1$. An example is shown in Fig. 2 (b), and we can see that potential customers c_1 and c_2 are more likely to pass the billboard area, hence, the preferences of potential customers c_1 and c_2 may be more important when the digital billboard is making decision.

3) Customer Preference Level Prediction: The last part of $\mathcal{S}$ is $\mathcal{S}_3$, which are the probabilities that each potential customer will purchase different products at the current time. It is a matrix of $n \times k$, which depends on how well product attributes match customers' preferences and how many times customers have seen the product advertising content. Suppose there are four potential customers and four different products, the probabilities of each potential customer purchasing these products are shown in Fig. 2 (c).

In order to quantify the preference level for each potential customer, firstly, we need to collect or infer the preferences of each potential customer. There are many ways to get relevant data. For example, based on the data collected by mobile crowdsensing (MCS), we can adopt the method in [5, 12] to infer the preferences of each potential customer. In this paper, the specific collection and inference process is not the focus, hence we use the generated preference data for the simulations so that we can reduce the difficulty of the calculation. The preferences of each potential customer c_i can be denoted as $Attr_{c_i}$. Then the preference level of the potential customer c_i about a product a_d can be defined as follows:

$$P_{prefer}^{a_d}(c_i) = \frac{Attr_{c_i} \cap Attr_{a_d}}{Attr_{c_i}}. \tag{5}$$

And the probability of a customer c_i buying a product a_d is defined as the following equation:

$$P_{buy}^{a_d}(c_i)=\begin{cases}0, & if\ c_i\ hasn't\ seen\ the\ product\ a_d \\ P_{prefer}^{a_d}(c_i), & if\ c_i\ first\ sees\ the\ product\ a_d \\ 1-(1-P_{buy}^{'a_d}(c_i))\times(1-P_{prefer}^{a_d}(c_i)), & otherwise\end{cases} \tag{6}$$

where $P_{buy}^{'a_d}(c_i)$ is the probability of customer c_i buying the product a_d before he sees the next advertisement about product a_d. It is obvious that the greater the potential customer's interest in the product a_d, the greater the value of $P_{buy}^{a_d}(c_i)$ is, which is reasonable. And we can also find that the more time a potential customer sees the same advertisement, the more likely he is to buy the product. This is reasonable because if a potential customer is interested in the product, the more he sees the advertisement of this product, the stronger his/her willingness to purchase this product will be. And if a potential customer is likely to buy a certain product, the impact of seeing the same advertisement many times on his/her final purchase will be small.

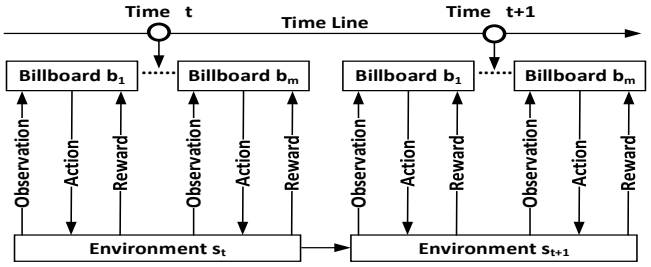


Fig. 3: Overall system flow of digital billboard advertising.

For now, we have defined the observation of each digital billboard. Generally speaking, each digital billboard b_g would observe the environment and obtain the observation $o_t^{b_g} = \{\mathcal{S}_1, \mathcal{S}_2^{b_g}, \mathcal{S}_3\} \in s_t$ at the beginning of each time slice t. Next, after each digital billboard has observed the environment, they need to decide what advertisement to deliver for the next time slice. Hence, we will discuss the action space of each digital billboard in next part.

B. Action Space

As we have discussed, each digital billboard can be regarded as an agent in reinforcement learning, hence, it is reasonable to regard a digital billboard switching every kind of advertisement as its actions. In this digital billboard advertising scene, the action space is denoted as $\mathcal{A} = \{a_1, a_2, ..., a_k \mid a_d \in A, d \in [1, k]\}$, where each action a_d is advertising for a relative product a_d. At the beginning of each time slice, each digital billboard should decide which action to take, in other words, the digital billboard b_g should determine what advertising content to deliver for this time slice to maximize its reward. The reward function is defined in next section.

C. Reward Function

After observing the current state of the environment, each digital billboard will take an action and obtain its reward from the environment, this process can be denoted by $\mathcal{S} \times \mathcal{A} \to \mathcal{R}$. In this paper, the reward function of digital billboard b_g is defined as follows:

$$r_{b_g}^{o_t^{b_g}}(a_d) = \sum_{i=1}^{n} f_{a_d} \times (P_{buy}^{a_d}(c_i) - P_{buy}^{'a_d}(c_i)), \tag{7}$$

where $P_{buy}^{'a_d}(c_i)$ is the probability of customer c_i buying the product a_d before he sees the billboard b_g, and $P_{buy}^{a_d}(c_i)$ is the probability of customer c_i buying the product a_d after he sees

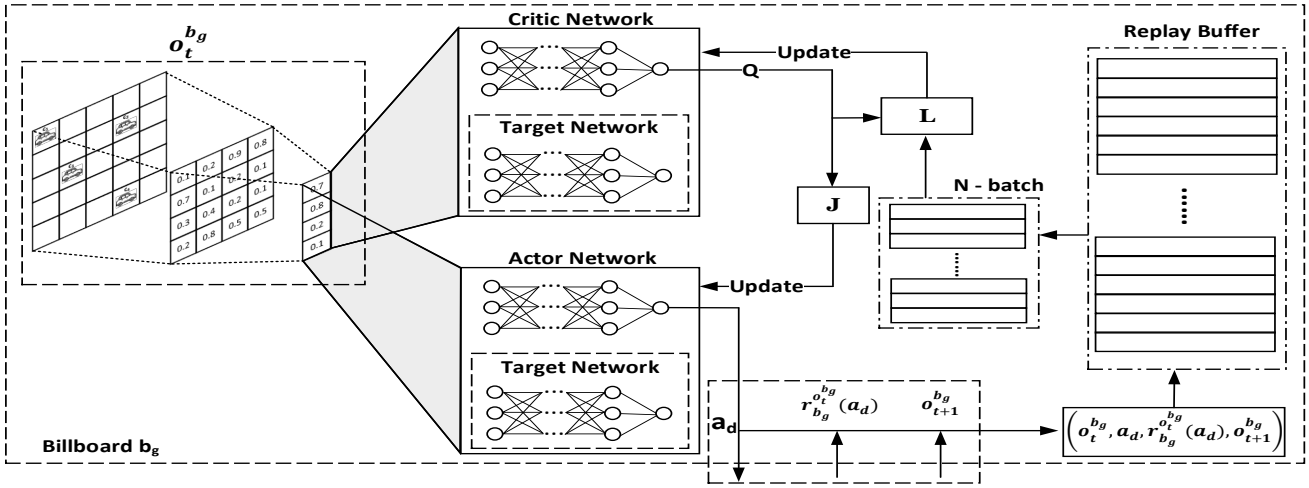


Fig. 4: Neural network of each digital billboard.

the billboard b_g. In other words, the reward function of each billboard is defined as the gain of the expected commercial profit. In the real world, the cost of digital billboard switching advertisement is very low, so in this paper, we ignore the cost of switching different advertisements for each digital billboard. And we suppose that there is no communication among digital billboards, hence the cost of communication can be also ignored. If a digital billboard increases the purchase probabilities of many potential customers after delivering an advertisement, then it could receive more reward, which is reasonable. At the beginning of each time slice, each billboard hopes to choose the action (advertisement) that can bring the greatest reward to itself.

D. State Transition and Probability Distribution

After a digital billboard performs an action and obtains its reward, the state s_t of the environment will transit to a new state s_{t+1}, which can be denoted as $\mathcal{F}: \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$. As we have defined the state space of the environment, we can find that the digital billboards only affect the probabilities that potential customers purchasing different products.

E. Problem Formulation

For now, we have defined the entire digital billboard advertising scene. The problem for the advertiser is how to determine the advertisement switching policies for the digital billboards so that the advertiser could achieve the maximal commercial profit, which is the dynamic advertising problem. Obviously, it is unreasonable for the advertiser to switch advertisements manually for each digital billboard. And a lot of resources could be consumed if we use a central server to control these digital billboards. Hence, we can use another way to solve this problem. As we have defined the MDP in our digital billboard advertising scene, for each digital billboard b_g, we can formulate our problem as follows:

$$\max V_{b_g}(o_t^{b_g}) = [r_{b_g}^{o_t^{b_g}}(a_d) + \gamma V_{b_g}(o_{t+1}^{b_g})], \tag{8}$$

where γ is the discount factor to measure the importance between future reward and current reward. For each digital billboard b_g, its optimal advertising strategy can be defined as follows:

$$\pi_{b_g} = \arg\max[r_{b_g}^{o_t^{b_g}}(a_d) + \gamma V_{b_g}(o_{t+1}^{b_g})]. \tag{9}$$

We can find that the goal of each digital billboard is to maximize its expected commercial profit. Hence, in this paper, we adopt Deep Reinforcement Learning (DRL) to find the optimal advertising strategy for the advertiser. However, in our digital billboard advertising scene, each billboard is considered as an independent agent and each agent makes decisions independently. In this way, we cannot directly use the traditional deep reinforcement learning method, (e.g., policy gradient), hence we propose our advertising strategy based on the Multi-Agent Deep Deterministic Policy Gradient (MADDPG).

V. ADVERTISING STRATEGY

In this section, we describe the structure of the neural networks for each digital billboard. And, we also discuss the detailed adverting strategy for the advertiser in order to maximize the commercial profit.

A. Advertising System Overview

First of all, we discuss the system flow of the digital billboard advertising scene. As we can see from Fig. 3, there are m digital billboards, which can be considered as different agents. These billboards are interacting with the environment at the same time. The current state of environment is s_t. Each billboard b_g would independently observe the state of the environment and take an action a_d from action space. After all the billboards have performed actions, the environment would change its state to next state s_{t+1} and each billboard would obtain its reward. This process will be repeated until the deadline and the advertiser could confirm the commercial profit.

B. Billboard Modeling

1) Taking action after observing: For a digital billboard b_g, first, the digital billboard b_g would observe the environment at

Algorithm 1 Advertising Strategy For Digital Billboard Advertising (ASFDBA)

Input: a set of billboards B, a set of potential users C, a set of advertisements (products) A

Output: target actor and critic networks

Initialize discount factor γ, update rate τ;
for billboard $b_g \subseteq B$ **do**
  Randomly initialize critic network $Q^{b_g}(o_t^{b_g}, a_d)$ and actor network $\pi^{b_g}(o_t^{b_g})$;
  Initialize target critic network $Q'^{b_g}(o_t^{b_g}, a_d) = Q^{b_g}(o_t^{b_g}, a_d)$ and target actor network $\pi'^{b_g}(o_t^{b_g}) = \pi^{b_g}(o_t^{b_g})$;
  Initialize the replay buffer RB_{b_g};
for episode = $1, 2, ..., E$ **do**
  Initialize environment;
  for epoch t = $1, 2, ..., T$ **do**
    for billboard $b_g \subseteq B$ **do**
      Get observation $o_t^{b_g}$, and select action $a_{b_g} = \pi^{b_g}(o_t^{b_g})$
      Execute each billboard's action a_{b_g}, get observation $o_{t+1}^{b_g}$ and reward $r_{b_g}^{o_t^{b_g}}(a_{b_g})$
    for billboard $b_g \subseteq B$ **do**
      Store transition $(o_t^{b_g}, a_{b_g}, r_{b_g}^{o_t^{b_g}}(a_{b_g}), o_{t+1}^{b_g})$ into RB_{b_g}
      Randomly sample N batches from RB_{b_g}
      Update critic network and actor network by algorithm 2

the beginning of a time slice t and obtain its observation $o_t^{b_g}$, which is shown on the left side of Fig. 4. Then the observation would be transformed into a matrix and fed into the neural network called actor network $\pi^{b_g}(o_t^{b_g})$. After calculation, the actor network will give the probabilities of taking different actions, which could maximize the expected reward. Then the digital billboard could decide its action for this time slice. This process happens on all digital billboards at the same time. By now, each digital billboard should have decided the action for the current time slice. Next, we will discuss the transition of environment state.

2) State Transition and Storage: After all the digital billboards have performed actions, potential customers would see the advertising content when they enter billboard areas. When the potential customers see the advertising content, the probabilities of potential customers purchasing different products would change, and each digital billboard could obtain the reward from the environment. We have defined observation of digital billboards, hence, after the probabilities of potential customers buying different products change, the state of the environment would also transit to another state. Each billboard could obtain a tuple, which consists of old state, the action, reward and new state, and the digital billboard would store this tuple into its replay buffer for training.

3) Training The Neural Networks: In this part, we will discuss how to train the neural networks of each digital billboard. First of all, we should see the detailed structure

Algorithm 2 Update The Critic and Actor Networks

$y = r_{b_g}^{o_t^{b_g}}(a_d) + \gamma Q'^{b_g}(o_{t+1}^{b_g}, a'_{b_1}, a'_{b_2}, ..., a'_{b_m})$
$L(Q^{b_g}) = \mathbb{E}[(Q^{b_g}(o_t^{b_g}, a_{b_1}, a_{b_2}, ..., a_{b_m}) - y)^2]$
Update the critic network by minimizing $L(Q^{b_g})$
Update the actor network by using gradients:
  $\nabla J = \mathbb{E}[\nabla \pi^{b_g}(a)|_{a=a_{b_g}}$
    $\nabla Q^{b_g}(o_t^{b_g}, a_{b_1}, a_{b_2}, ..., a_{b_m})]|_{a_{b_g}=\pi^{b_g}(o_t^{b_g})}$
Update the weights of corresponding target networks by:
$w_{Q'^{b_g}} = \tau w_{Q^{b_g}} + (1-\tau) w_{Q'^{b_g}}$
$w_{\pi'^{b_g}} = \tau w_{\pi^{b_g}} + (1-\tau) w_{\pi'^{b_g}}$

of the neural networks, which is shown in Fig. 4. We can see that there are two neural networks: actor network $\pi^{b_g}(o_t^{b_g})$ and critic network $Q^{b_g}(o_t^{b_g}, a_d)$. The actor network is used to identify the action, which the digital billboard should take. And the critic network is used to estimate the reward of different observations and actions. For example, after a digital billboard b_g has observed the environment and performed its action, the next time this digital billboard encounters the same state, if this digital billboard takes a different action and gets a bigger reward, then the actor network should increase the probability of taking the new action. Hence, we can propose the method to update two neural networks. There are three fully connected layers in each neural networks, where each node of the full connection layer is connected to all nodes of the previous layer.

For the critic network, during the training process, each digital billboard should sample N-batches from the replay buffer, and update the critic network with minimizing the loss function:

$$\begin{aligned} L(Q^{b_g}) &= \mathbb{E}[(Q^{b_g}(o_t^{b_g}, a_{b_1}, a_{b_2}, ..., a_{b_m}) - y)^2] \\ y &= r_{b_g}^{o_t^{b_g}}(a_d) + \gamma Q'^{b_g}(o_{t+1}^{b_g}, a'_{b_1}, a'_{b_2}, ..., a'_{b_m}) \end{aligned} \tag{10}$$

where a'_{b_m} is the next action that the digital billboard b_g would take for the next time slice and the value of critic network $Q^{b_g}(o_t^{b_g}, a_d)$ is the digital billboard b_g's Q function with observations and actions from the replay buffer. Obviously, the method of updating the critic network is to reduce the difference between the predicted reward and the real reward, so as to improve the accuracy of prediction. Note that when predicting the reward, both the immediate reward in the current state and the possible future reward are taken into consideration, we use the parameter γ to represent the importance of future rewards.

After we have trained the critic network, we need to train the actor network by using gradients, which is shown in Eq. 11:

$$\begin{aligned} \nabla J = \mathbb{E}[&\nabla \pi^{b_g}(a)|_{a=a_{b_g}} \\ &\nabla Q^{b_g}(o_t^{b_g}, a_{b_1}, a_{b_2}, ..., a_{b_m})]|_{a_{b_g}=\pi^{b_g}(o_t^{b_g})} \end{aligned} \tag{11}$$

The goal of updating the actor network is to increase the expected reward of the output action probabilities.

4) Inferring Policies of Other Digital Billboards: As we have discussed that the cooperation and competition among all digital billboards will also affect the commercial profit, it

is necessary for each digital billboard to consider the possible actions of other digital billboards when deciding the action for the current time slice. However, in our digital advertising scene, there is no communication among digital billboards. Besides, all digital billboards determine the next action at the same time, so there is no real-time communication. Hence, for a digital billboard b_g, it should infer the advertisement switching policies of other digital billboards so that it may achieve more commercial profit.

This advertisement switching policy is learned by maximizing the log probability of digital billboard b_j's actions [13], which is shown as follows:

$$\mathcal{L}(\theta_{b_g}^{b_j}) = -\mathbb{E}[\log \hat{\pi}_{b_g}^{b_j}(a_{b_j}|o_t^{b_j}) + \lambda H(\hat{\pi}_{b_g}^{b_j})] \quad (12)$$

where H is the entropy of the policy distribution. $\theta_{b_g}^{b_j}$ is the approximate parameter of π^{b_j} by b_g. With the approximate advertisement switching policies, we can replace the y in Eq.10 by the approximate value $\hat{y}$, which is shown as follows:

$$\hat{y} = r_{b_g}^{o_t^{b_g}}(a_d) + \gamma Q'^{b_g}(o_{t+1}^{b_g}, \hat{\pi'}_{b_g}^{b_1}(o^{b_1}), \hat{\pi'}_{b_g}^{b_2}(o^{b_2}), ..., \hat{\pi'}_{b_g}^{b_m}(o^{b_m})). \quad (13)$$

Note that Eq.12 can be optimized by a online approach. For example, before updating Q^{b_g}, the digital billboard b_g could take the latest sample from the replay buffer of each digital billboard and update the approximate advertisement switching policy of each digital billboard.

C. Testing The Neural Network

After we have trained the neural networks for each digital billboard, we only need the target actor network of each billboard for testing. Each digital billboard could observe the environment and obtain $o_t^{b_g}$ at the beginning of time slice t. Then the actor network could determine the probabilities of taking different actions for the digital billboard b_g by giving the observation. After all digital billboards have performed their actions, the environment would give a reward to each digital billboard and change its state to $o_{t+1}^{b_g}$. It is obvious that there is no communication between each billboard during the test phase. And each billboard could make its decision by its own observation.

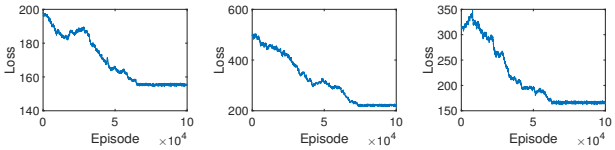
D. Advertising Strategy

The detailed advertising strategy is shown in algorithm 1. First of all, we need to train the neural networks for each digital billboard. At the beginning of training process, we initialize the parameters, environment and neural networks for the training (line 1-5). We consider each digital billboard b_g as an agent, and each agent could obtain its observation at the beginning of each epoch (time slice). Then, the agents could decide their actions by feeding their actor networks with their observations (line 6-11). Next, all the digital billboards will switch the advertisements by their actions. After all the digital billboards have executed their actions, potential customers in the environment start moving and the digital billboards will obtain the rewards from the environment at the end of the time slice (epoch). These above information will be stored into their replay buffers, and next we can update their neural networks by using samples from replay buffer (line 12-17).

After we have trained the neural networks for each digital billboard, we could use the target actor network of each digital billboard to do the advertising for the advertiser. During the testing process, all the digital billboards will observe the environment and feed their observations to their target actor networks. Then they can get the outputs of their actor networks and decide what actions they should take according to the outputs. After all the digital billboards have executed their actions, the environment will change its state. This process will be repeated until the deadline. The advertiser will obtain the commercial profit after the entire life cycle.

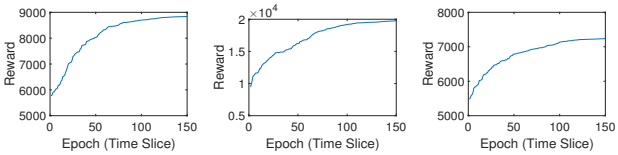
TABLE I: Parameters Settings.

Method	Datasets		
	roma/taxi	*epfl*	*geolife*
Deadline	80,90,100,110,120		
Billboard Number	1,2,3,4,5		
Advertisement Number	3,4,5,6,7		
Customer Preference Number	8,9,10,11,12		
Total Preference Type	20		
Time Unit(s)	15	30	5
Grid Number	13×15	12×14	10×16
User Number	49	57	59
Repeat Time	10000		
Training Episode	80000		
γ	0.9		
N - batch	1000		



(a) *roma/taxi* (b) *epfl* (c) *geolife*

Fig. 5: Average loss on three datasets.



(a) *roma/taxi* (b) *epfl* (c) *geolife*

Fig. 6: Average reward on three datasets.

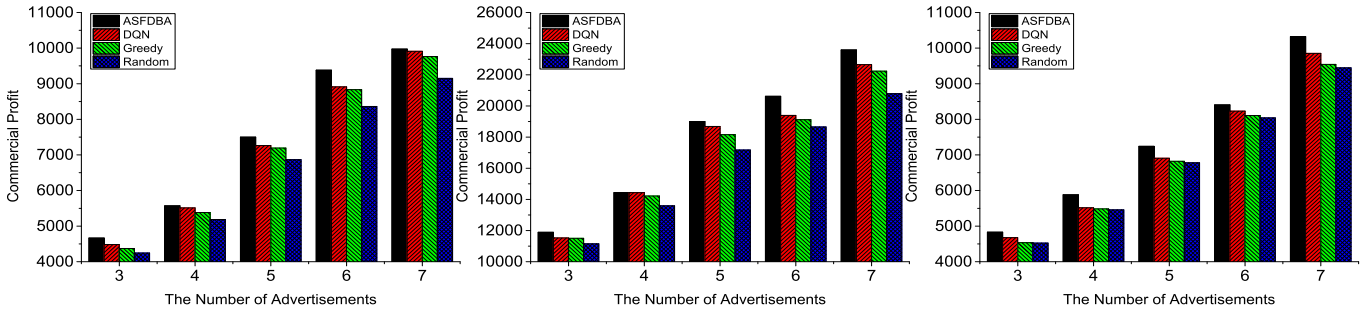
VI. PERFORMANCE EVALUATION

In this section, simulations are conducted to evaluate the performances of different advertising strategies by using three widely-used real-world data sets. We compare the performances of our proposed advertising strategy with other advertising strategies under different conditions.

A. The Simulation Traces And Settings

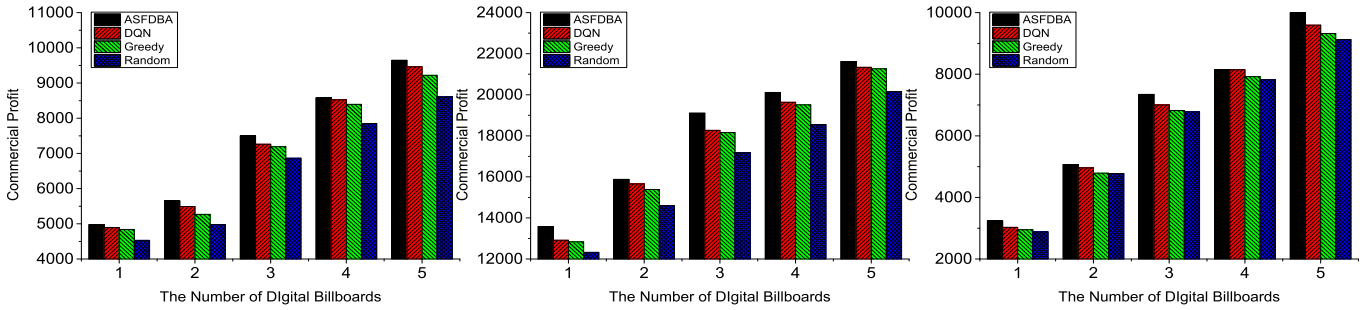
In this paper, three real-world datasets are adopted:

- *roma/taxi* trace set [14]: In *roma/taxi* trace set, 320 taxi drivers that work in the center of Rome city are included.



(a) *roma/taxi* (b) *epfl* (c) *geolife*

Fig. 7: Performances on different number of advertisements.



(a) *roma/taxi* (b) *epfl* (c) *geolife*

Fig. 8: Performances on different number of digital billboards.

The trajectories in this dataset are collected every 7 seconds and sent to a central server, which represent the positions of those taxi drivers.

- *epfl* trace set [15]: In the *epfl* trace set, there are about 500 taxis' GPS coordinates, 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 a central server. The records are fine-grained so that we can accurately collect user positions by these location-updates.
- *geolife* trace set [16]: In *geolife* trace set, there are about 17621 trajectories whose total distance is about 1.2 million kilometers. The total duration of this dataset is about 48000 hours, which are collected by different GPS loggers and phones.

These datasets are preprocessed by filtering out some abnormal users including those with discontinuous trajectories or remote locations. Then, we adopt the Baidu Map API to match these traces into a map area and divide this area into a grid-shape. The trajectories of each user can be converted to a series of grid coordinates. We regard the users in these datasets as the potential customers for the simulations. Hence, during the simulations, the trajectories in these datasets can be regarded as the movement of potential customers in different time slices.

The detailed parameter settings of simulations are shown in Table.I. We have discussed that the specific collection and inference process of potential customers preferences is not the focus of this paper, hence, the preferences of each potential customer are randomly generated in the simulations in order to reduce the difficulty of calculation.

B. Evaluation Results

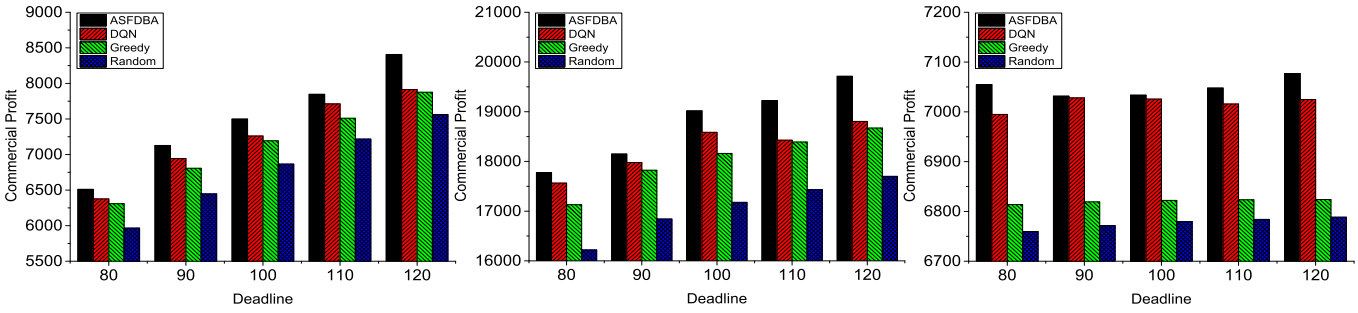
C. Comparison Advertising Strategies And Metric

The dynamic advertising problem is quite different from the existing works, hence we compare our advertising strategy with the following advertising strategies:

- *Deep Q-Learing (DQN)*: Each digital billboard chooses the advertisement by using Deep Q-Learning (DQN)[34], where each digital billboard decides its action by feeding its observation to a neural network.
- *Greedy*: Each digital billboard chooses the advertisement which has the maximal commercial profit at the beginning of each time slice.
- *Random*: Each digital billboard randomly chooses the advertisement at the beginning of each time slice.

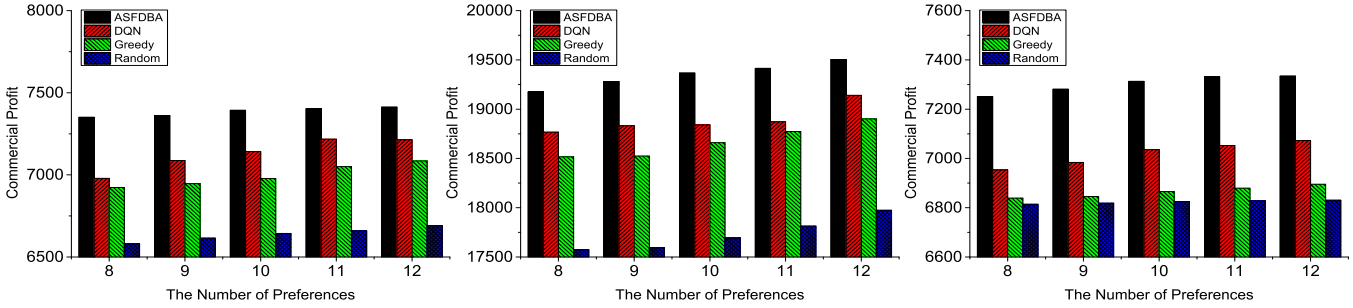
We use the commercial profit as the metric to measure the performances of different advertising strategies. When an advertising strategy performs better, it would have higher commercial profit, which is reasonable. In order to calculate the commercial profit, we need to judge whether a customer is attracted to purchase a product. When a potential customer sees an advertisement on a billboard, he has a chance to buy the product in the advertisement. After the deadline of the whole experiment, each potential customer has different purchase probability 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.

Besides, each potential customer's trajectories and mobility pattern will not be affected by the advertisements they see. In



(a) *roma/taxi* (b) *epfl* (c) *geolife*

Fig. 9: Performances on different deadline.



(a) *roma/taxi* (b) *epfl* (c) *geolife*

Fig. 10: Performances on different number of customers' preferences.

our scenario, each potential customer will have the probability of purchasing a product after seeing the advertisement. At the end of each round of experiments, if a potential customer decides to purchase a product, we consider the advertiser will obtain the commercial profit from the potential customer and ignore the customer's journey to the store.

In this section, we aim to evaluate the performances of our proposed advertising strategy and compare it with other advertising strategies. Specifically, we test the commercial profit along with the changing of the number of advertisements, the number of digital billboards and deadline. The simulation results on three different real-world data sets are illustrated in Fig. 7-Fig. 9.

1) Average Loss and Reward on different datasets: First of all, we show the average loss and reward on three different datasets, and the results are shown in Fig. 5 and Fig. 6. The number of digital billboards in the simulations is set to 3 and the number of advertisements is set to 5. From Fig. 5, we can find that the loss values of simulations on three datasets converge after about 80000 training episodes and hence we set the training episodes to 80000 for the rest simulations. From Fig. 6, we can find that the reward increases with time and converges in about 100 time slices. Hence, we set the deadline to 100 time slices as the baseline. Next, we show the detailed performances of different advertising strategies on three datasets.

2) Performances on different number of advertisements: In this part, we illustrate the results when we change the number of advertisements and keep the others fixed. The number of billboards is set to 3, and the deadline is set to 100 time slices. Each billboard could have 3 to 7 advertisements for advertising, which is also its action space. The results are shown in Fig. 7. We can rank the performances of different strategies as follows: $ASFDBA > DQN > Greedy > Random$, and *ASFDBA* outperforms 2% to 5% than *greedy* on these three datasets. Specifically, the commercial profit increases along with the increase of the number of advertisements. It is reasonable because when there are more different advertisements, they are more likely to match the preferences of different potential customers. Hence, the probabilities of potential customers purchasing products may increase. Besides, the same potential customer could be attracted by different advertisements, which could also enlarge the commercial profit.

3) Performances on different number of digital billboards: In this paper, each digital billboard is considered as an agent and it could decide its action at the beginning of each time slice. Hence, it is necessary to evaluate the performances when we change the number of digital billboards. During the simulation, the number of advertisements is set to 5, and the deadline is set to 100. The number of digital billboards is set from 1 to 5. As we can see from Fig. 8, *ASFDBA* could always obtain the maximal commercial profit. This is because each digital billboard would consider the actions of the other digital billboards when it needs to determine the action for the current time slice by using *ASFDBA* and hence each digital billboard could maximize its expected commercial profit. However, when digital billboards decide their actions by using DQN, their policies may be constantly changed, because the policies of other agents are not considered in the training process.

4) Performances on different deadlines: Next, we conduct the simulations to test performances of different advertising

strategies when we change the deadline and the results are shown in Fig. 9. We can find that our proposed advertising strategy *ASFDBA* could always achieve the maximal commercial profit for the advertiser. The results of *ASFDBA* are improved about 3% -7% compared with *Greedy*. The strategy *Random* performs worst, which is reasonable. Because when the digital billboards randomly take actions, the commercial profit is not stable. Compared with other advertising strategies, it is difficult for the strategy *Random* to choose the optimal solution for each time slice. Hence, the performances of *Random* are the worst. We can also find that the commercial profit increases along with the growth of the deadline. Because potential customers have more time and opportunities to see these digital billboards. Besides, the same potential customer could see the same advertisement multiple times, hence the digital billboards could attract more potential customers and the commercial profit increases.

5) Performances on different number of customer preferences: Finally, we illustrate the results when we change the number of customers' preferences and the results are shown in Fig. 10. We can find that as the number of customers' preferences increases, the performance of different strategies also improves and the differences in the results of these strategies decrease. This is because when the number of customers' preferences increases, so will the customers' interests in different products. In other words, products are more attractive to potential customers. Hence, the commercial profit for the advertiser will increase. Besides, when the number of customers' preferences is the same as the total number of preferences (each potential customer has all kinds of preferences), the potential customers will be interested in all products, so in this case, the impact of different strategies on the results will be smaller. That is the reason why he differences in the results of these strategies decrease.

VII. CONCLUSION

In this paper, a dynamic advertising problem is formulated to determine the advertisement switching policies for different digital billboards so that the advertiser could attract as many potential customers as possible and the commercial profit for the advertiser could be maximized. To address the dynamic advertising problem, first of all, we use the vehicular data of potential customers collected by mobile crowdsensing (MCS) to extract potential customers' implicit information, such as their historical vehicular trajectories and their preferences. Based on the above vehicular data, we adopt a semi-markov model to predict the customers' mobility patterns to show their chances to see the digital billboards. Then, we propose an advertising strategy called *ASFDBA* by using multi-agent deep deterministic policy gradient (MADDPG), where each digital billboard can decide its current advertisement independently without communicating with the other digital billboards. Finally, We conduct extensive simulations based on three widely-used real-world trajectories: *roma/taxi*, *epfl*, and *geolife*. The results show that our advertising strategy could achieve the best commercial profit for the advertiser compared with other advertising strategies.

REFERENCES

[1] https://digitalsignagepulse.com/news/global-digital-ooh-media-revenues-pacing-up-13-in-2017-us-dooh-advertising-expands-10-pq-media.

[2] S. Nigam, S. Asthana, and P. Gupta, "Iot based intelligent billboard using data mining," in *2016 International Conference on Innovation and Challenges in Cyber Security (ICICCS-INBUSH)*, Feb 2016, pp. 107–110.

[3] D. Liu, D. Weng, Y. Li, J. Bao, Y. Zheng, H. Qu, and Y. Wu, "Smartadp: Visual analytics of large-scale taxi trajectories for selecting billboard locations," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 1–10, Jan 2017.

[4] M. Huang, Z. Fang, S. Xiong, and T. Zhang, "Interest-driven outdoor advertising display location selection using mobile phone data," *IEEE Access*, vol. 7, pp. 30 878–30 889, 2019.

[5] L. Wang, Z. Yu, D. Yang, H. Ma, and H. Sheng, "Efficiently targeted billboard advertising using crowdsensing vehicle trajectory data," *IEEE Transactions on Industrial Informatics*, pp. 1–1, 2019.

[6] H. Zheng and J. Wu, "Placement optimization for advertisement dissemination in smart city," *IEEE Transactions on Network Science and Engineering*, pp. 1–1, 2018.

[7] W. Liu, Y. Yang, E. Wang, and J. Wu, "User recruitment for enhancing data inference accuracy in sparse mobile crowdsensing," *IEEE Internet of Things Journal*, pp. 1–1, 2019.

[8] W. Liu, L. Wang, E. Wang, Y. Yang, D. Zeghlache, and D. Zhang, "Reinforcement learning-based cell selection in sparse mobile crowdsensing," *Computer Networks*, vol. 161, pp. 102–114, 2019.

[9] R. K. Ganti, F. Ye, and H. Lei, "Mobile crowdsensing: current state and future challenges," *IEEE Communications Magazine*, vol. 49, no. 11, pp. 32–39, November 2011.

[10] J. Liu, H. Shen, and X. Zhang, "A survey of mobile crowdsensing techniques: A critical component for the internet of things," in *2016 25th International Conference on Computer Communication and Networks (ICCCN)*, Aug 2016, pp. 1–6.

[11] A. Capponi, C. Fiandrino, B. Kantarci, L. Foschini, D. Kliazovich, and P. Bouvry, "A survey on mobile crowdsensing systems: Challenges, solutions, and opportunities," *IEEE Communications Surveys Tutorials*, vol. 21, no. 3, pp. 2419–2465, thirdquarter 2019.

[12] M. Karaliopoulos, I. Koutsopoulos, and M. Titsias, "First learn then earn: Optimizing mobile crowdsensing campaigns through data-driven user profiling," in *Proceedings of the 17th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, ser. MobiHoc '16. New York, NY, USA: ACM, 2016, pp. 271–280. [Online]. Available: http://doi.acm.org/10.1145/2942358.2942369

[13] R. Lowe, Y. WU, A. Tamar, J. Harb, O. Pieter Abbeel, and I. Mordatch, "Multi-agent actor-critic for mixed cooperative-competitive environments," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 6379–6390. [Online]. Available: http://papers.nips.cc/paper/7217-multi-agent-actor-critic-for-mixed-cooperative-competitive-environments.pdf

[14] L. Bracciale, M. Bonola, P. Loreti, G. Bianchi, R. Amici, and A. Rabuffi, "CRAWDAD dataset roma/taxi (v. 2014-07-17)," Downloaded from https://crawdad.org/roma/taxi/20140717, Jul. 2014.

[15] M. Piorkowski, N. Sarafijanovic-Djukic, and M. Grossglauser, "CRAWDAD dataset epfl/mobility (v. 2009-02-24)," Downloaded from https://crawdad.org/epfl/mobility/20090224, Feb. 2009.

[16] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma, "Mining interesting locations and travel sequences from gps trajectories," in *Proceedings of the 18th International Conference on World Wide Web*, ser. WWW '09. New York, NY, USA: ACM, 2009, pp. 791–800. [Online]. Available: http://doi.acm.org/10.1145/1526709.1526816

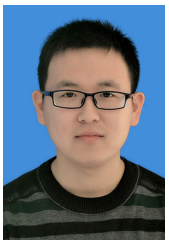
[17] T. T. An, C. Chang, Y. Li, and S. Yuan, "Fog computing architecture-based wi-fi union mechanism for internet advertising system," in *2017 International Conference on Applied System Innovation (ICASI)*, May 2017, pp. 1024–1027.

[18] Y. Zhang, Y. Li, Z. Bao, S. Mo, and P. Zhang, "Optimizing impression counts for outdoor advertising," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ser. KDD '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1205–1215. [Online]. Available: https://doi.org/10.1145/3292500.3330829

[19] L. Busoniu, R. Babuska, and B. De Schutter, "Multi-agent reinforcement learning: A survey," in *2006 9th International Conference on Control, Automation, Robotics and Vision*, Dec 2006, pp. 1–6.

[20] I. Althamary, C. Huang, and P. Lin, "A survey on multi-agent re-

inforcement learning methods for vehicular networks," in *2019 15th International Wireless Communications Mobile Computing Conference (IWCMC)*, June 2019, pp. 1154–1159.
[21] T. Chu, J. Wang, L. Codecà, and Z. Li, "Multi-agent deep reinforcement learning for large-scale traffic signal control," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–10, 2019.
[22] S. Zheng and H. Liu, "Improved multi-agent deep deterministic policy gradient for path planning-based crowd simulation," *IEEE Access*, vol. 7, pp. 147 755–147 770, 2019.
[23] Y. Pan, H. Jiang, H. Yang, and J. Zhang, "A novel method for improving the training efficiency of deep multi-agent reinforcement learning," *IEEE Access*, vol. 7, pp. 137 992–137 999, 2019.
[24] J. Wang, Y. Wang, D. Zhang, F. Wang, Y. He, and L. Ma, "Psallocator: Multi-task allocation for participatory sensing with sensing capability constraints," in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, ser. CSCW '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 1139–1151. [Online]. Available: https://doi.org/10.1145/2998181.2998193
[25] J. Wang, Y. Wang, D. Zhang, F. Wang, H. Xiong, C. Chen, Q. Lv, and Z. Qiu, "Multi-task allocation in mobile crowd sensing with individual task quality assurance," *IEEE Transactions on Mobile Computing*, vol. 17, no. 9, pp. 2101–2113, Sep. 2018.
[26] J. Wang, F. Wang, Y. Wang, L. Wang, Z. Qiu, D. Zhang, B. Guo, and Q. Lv, "Hytasker: Hybrid task allocation in mobile crowd sensing," *IEEE Transactions on Mobile Computing*, vol. 19, no. 3, pp. 598–611, March 2020.
[27] J. Wang, F. Wang, Y. Wang, D. Zhang, L. Wang, and Z. Qiu, "Social-network-assisted worker recruitment in mobile crowd sensing," *IEEE Transactions on Mobile Computing*, vol. 18, no. 7, pp. 1661–1673, July 2019.
[28] D. P. Bertsekas, D. P. Bertsekas, D. P. Bertsekas, and D. P. Bertsekas, *Dynamic programming and optimal control*, 1995, vol. 1.
[29] C. H. Liu, T. He, K. Lee, K. K. Leung, and A. Swami, "Dynamic control of data ferries under partial observations," in *2010 IEEE Wireless Communication and Networking Conference*, April 2010, pp. 1–6.
[30] M. Lin, W.-J. Hsu, and Z. Q. Lee, "Predictability of individuals' mobility with high-resolution positioning data," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, ser. UbiComp '12. New York, NY, USA: ACM, 2012, pp. 381–390. [Online]. Available: http://doi.acm.org/10.1145/2370216.2370274
[31] E. Wang, Y. Yang, J. Wu, W. Liu, and X. Wang, "An efficient prediction-based user recruitment for mobile crowdsensing," *IEEE Transactions on Mobile Computing*, vol. 17, no. 1, pp. 16–28, Jan 2018.
[32] Q. Yuan, I. Cardei, and J. Wu, "An efficient prediction-based routing in disruption-tolerant networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 1, pp. 19–31, Jan 2012.
[33] Y. Yang, W. Liu, E. Wang, and J. Wu, "A prediction-based user selection framework for heterogeneous mobile crowdsensing," *IEEE Transactions on Mobile Computing*, pp. 1–1, 2018.
[34] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," *arXiv preprint arXiv:1312.5602*, 2013.



Kaihao Lou received the B.E. degree in software engineering from Jilin University, Changchun, Jilin, China, in 2017. He is currently pursuing the M.S. degree in computer system architecture from Jilin University, Changchun, Jilin, China. His current research focuses on mobile crowdsensing.

Yongjian Yang received his B.E. degree in automatization from Jilin University of Technology, Changchun, Jilin, China in 1983; his M.E. degree in computer communication from Beijing University of Post and Telecommunications, Beijing, China in 1991; and his Ph.D. in software and theory of computer from Jilin University, Changchun, Jilin, China in 2005. He is currently a professor and a PhD 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 communication.

En Wang the corresponding author, received his B.E. degree in software engineering from Jilin University, Changchun in 2011, his M.E. degree in computer science and technology from Jilin University, Changchun in 2013, and his Ph.D. in computer science and technology from Jilin University, Changchun in 2016. He is currently a lecturer in the Department of Computer Science and Technology at Jilin University, Changchun. He is also a visiting scholar in the Department of Computer and Information Sciences at Temple University in Philadelphia. His current research focuses on the efficient utilization of network resources, scheduling and drop strategy in terms of buffer-management, energy-efficient communication between human-carried devices, and mobile crowdsensing.

Zheli Liu received the BSc and MSc degrees in computer science from Jilin University, China, in 2002 and 2005, respectively. He received the PhD degree in computer application from Jilin University in 2009. After a postdoctoral fellowship in Nankai University, he joined the College of Computer and Control Engineering of Nankai University in 2011. Currently, he works at Nankai University as an Associate Professor. His current research interests include applied cryptography and data privacy protection.

Thar Baker Dr Thar Baker is a Senior Lecturer in Software Systems in the Department of Computer Science at the Faculty of Engineering and Technology. He has received his PhD in Autonomic Cloud Applications from LJMU in 2010. Dr Baker has published numerous refereed research papers in multidisciplinary research areas including: Cloud Computing, Distributed Software Systems, Big Data, Algorithm Design, Green and Sustainable Computing, and Autonomic Web Science. He has been actively involved as member of editorial board and review committee for a number peer reviewed international journals, and is on programme committee for a number of international conferences. Dr. Baker was appointed as Expert Evaluater in the European FP7 Connected Communities CONFINE project (2012-2015). He worked as Lecturer in the Department of Computer Science at Manchester Metropolitan University (MMU) in 2011. Prior to this, he was working as a Post-Doctoral Research Associate in the area of Autonomic Cloud Computing at LJMU, where he built the first private cloud computing research platform for the Department of Computer Science. He has successfully completed the Strategic Executive Development for Diverse Leasers in Higher Education (StellarHE) Course in 2016.

Ali Kashif Bashir Ali Kashif Bashir is a Senior Lecturer at the Department of Computing and Mathematics, Manchester Metropolitan University, United Kingdom. He is also holding Adjunct Professor position at National University of Science and Technology, Pakistan. He is a senior member of IEEE, invited member of IEEE Industrial Electronic Society, member of ACM, and Distinguished Speaker of ACM. His past assignments include Associate Professor of ICT, University of the Faroe Islands, Denmark; Osaka University, Japan; Nara National College of Technology, Japan; the National Fusion Research Institute, South Korea; Southern Power Company Ltd., South Korea, and the Seoul Metropolitan Government, South Korea. He has worked on several research and industrial projects of South Korean, Japanese and European agencies and Government Ministries. He is also advising several start-ups in the field of STEM based education, block chain, robotics, and smart homes. He received his Ph.D. in computer science and engineering from Korea University South Korea. He has authored over 100 research articles and is supervising/co-supervising several graduate (MS and PhD) students. His research interests include internet of things, wireless networks, distributed systems, network/cyber security, cloud/network function virtualization, etc. He is serving as the Editor-in-chief of the IEEE FUTURE DIRECTIONS NEWSLETTER.