PAS: Prediction Based Actuation System for City-scale Ride Sharing Vehicular Mobile Crowdsensing

Xinlei Chen*, Susu Xu*, Jun Han, Haohao Fu, Xidong Pi
Carlee Joe-Wong, Yong Li, Lin Zhang, Hae Young Noh, Pei Zhang

***Abstract*—Vehicular mobile crowdsensing (MCS) enables many smart city applications. Ride sharing vehicle fleets provide promising solutions to MCS due to advantages of low cost, easy maintenance, high mobility and long operational time. However, as non-dedicated mobile sensing platforms, the first priorities of these vehicles are delivering passengers, which may lead to poor *sensing coverage quality*. Therefore, to help MCS derive good (large and balanced) *sensing coverage quality*, an actuation system is required to dispatch vehicles with a limited amount of monetary budget.**

This paper presents *PAS*, a prediction based actuation system for city-wide ride sharing vehicular MCS to achieve optimal *sensing coverage quality* with a limited budget. In *PAS*, two prediction models forecast probabilities of potential near-future vehicle routes and ride requests across the city. Based on prediction results, a prediction based actuation planning algorithm is proposed to decide which vehicles to actuate and the corresponding routes. Experiments on city scale deployments and physical feature based simulations show that our *PAS* achieves up to 40% more improvement in *sensing coverage quality* and up to 20% higher ride request matching rate than baselines. In addition, to achieve a similar level of *sensing coverage quality* as the baseline, our *PAS* only needs 10% budget.

***Index Terms*—Mobile Crowdsensing, Urban Sensing, Mobile Computing, Sensing Optimization, Vehicular Network**

I. INTRODUCTION

IN the past decade, the increasing number of sensing devices and wireless networks promotes the development of mobile crowdsensing (MCS), which enables many smart city applications [1], [2], [3], [4]. With MCS, a group of participants spatially distributed across different parts of the city, collectively sense, share and extract information to measure, map and infer phenomena of common interest, including air pollution, traffic congestion, urban fire [5], [6], [7], [8].

The inference accuracy of MCS is affected by two properties of data collection, as depicted in Fig. 1: (1) the size of sensed areas and (2) the evenness of spatial distribution [9], [10]. The former ensures sufficient information for inference, while the latter prevents inference accuracy bias over different areas [11]. This is because the local inference accuracy is decided by the amount of collected data in adjacent areas. Similar to previous works, we adopt *sensing coverage quality* to represent these two parts and the detailed definition can be found in Section III-B [10], [12].

Corresponding Author: Xinlei Chen, xinlei.chen@sv.cmu.edu. Xinlei Chen, Susu Xu, Xidong Pi, Carlee Joe-Wong, and Pei Zhang are with Carnegie Mellon University. Hae Young Noh is with Stanford University. Yong Li and Lin Zhang are with Tsinghua University. Haohao Fu is with UC Berkeley. Jun Han is with National University of Singapore. * Equal contribution.

Ride sharing vehicle fleets, including taxis, Uber, Lyft, Didi, offer promising solutions to MCS due to advantages of low cost, easy maintenance, high mobility and long operational time [13]. Mobile sensors are pre-installed on these ride sharing vehicles to provide real-time sensing when vehicles move around the city. MCS collects and utilizes data from different participants (i.e. sharing vehicles) to infer and map the phenomena of common interest [14], [15], [16]. With different sensors mounted on vehicles, ride sharing vehicular MCS systems are able to collect various types of information, such as air pollution, noise, videos, vibrations and etc. For example, gas sensors can be mounted on taxis to collect gas concentration including PM2.5 and CO_2 over the city [17], [18]. The collected gas data can be used to infer the air pollution field over the entire city for real-time air quality monitoring [19].

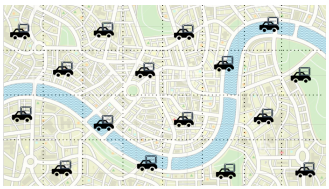
However, in ride sharing vehicular MCS systems, little data is collected at most areas [20]. This is because the first priorities of ride sharing vehicles are delivering passengers as they are non-dedicated mobile sensing platforms. Most of these vehicles tend to gather around busy areas, like central business districts, since they have more opportunities to get new ride orders [13]. As a result, other areas of the city are barely sensed, hence leading to low *sensing coverage quality*. Therefore, to improve *sensing coverage quality*, actuation systems are needed to dispatch vehicles with the given monetary budget.

It is challenging to design an actuation system for ride sharing vehicular MCS to optimize *sensing coverage quality* for the following reasons. *(C_1) Given a limited budget, which vehicles should be selected for actuation?* The limited budget only allows small percent of the vehicle fleet being actuated. Besides the actuated vehicles, the *sensing coverage quality* of data collection is also decided by non-actuated vehicles. Therefore, different vehicle actuation selections lead to different data collections. *(C_2) Given selected vehicles, how should each of them be actuated?* For the same vehicle, multiple actuation routes are available and those who have new ride requests at the destination are more attractive. Therefore, to ensure successful actuation, different routes have different monetary cost. Note that the two challenges are interacted. On the one hand, the vehicle selection process decides the actuation route design. On the other hand, the actuation route design process

(a) 12.5% of entire area is sensed; Sensed data gathers in the center area.

(b) 25% of entire area is sensed; Sensed data gathers in the center area.

(c) 25% of entire area is sensed; Sensed data is evenly distributed.



(d) 50% of entire area is sensed; Sensed data is evenly distributed (ideal situation).

Fig. 1. This figure depicts different data collections with same amount of ride sharing vehicles. (a) shows a undesirable case, where multiple vehicles gather in the same grid to collect redundant information and all vehicles gather in the centered area. (b) shows a larger sensed area than (a) but all vehicles still gather in the center area. (c) shows a better situation than (b): same sensed area with (b) but vehicles are distributed over the entire area. (d) shows an ideal situation, where all vehicles cover largest area and distribute evenly.

decides the actuation cost, which reversely decides how many vehicles can be actuated.

Many methods have been proposed in the last decade, which can be categorized into reverse auction-based and game-theoretical mechanisms [21], [22], [23], [24]. First, participants need to select and bid the actuation task, which makes them distracted from driving. Second, these approaches require many rational participants and incorporate all their preferences, which is extremely challenging in practice. Consequently, the performances of these methods highly rely on participants' involvement and attention. More importantly, they do not consider the future mobility of the rest non-actuated vehicles, which leads to high uncertainty of overall data collection after actuation.

In light of this, we present a prediction based actuation system (*PAS*) for city-wide ride sharing vehicular MCS to achieve optimal *sensing coverage quality* with a limited budget. To address the challenge C_1, *PAS* adopts a vehicle mobility prediction model to forecast probabilities of potential near-future routes. The prediction helps the system to consider overall data collection of the entire vehicle fleet, which guides the system to decide which vehicles to select for actuation to maximize *sensing coverage quality*. Intuitively, the system intends to spend the budget on vehicles, who are predicted to head for busy areas, and actuate them to sparse areas. To address the challenge C_2, a ride request prediction model is incorporated to predict near-future ride requests across the city. According to the prediction, *PAS* intends to select routes where there exist new ride requests at the destination. This not only improves the motivation of vehicles but also lowers the cost of actuation [25]. Given the two prediction models, we propose a prediction based actuation planning algorithm to decide which vehicles to actuate and corresponding routes.

We summarize our contributions as following:

- Propose a prediction based actuation system to optimize the *sensing coverage quality* for ride sharing vehicular MCS.
- Design a practical prediction based actuation planning algorithm, which not only improves the actuation effectiveness, but also lowers the average actuation cost.
- Deploy a taxi-based city scale system and physical feature based simulation for system evaluation within a $15km \times 15km$ area in the city of Beijing, China.

The remainder of the paper is organized as follows: Section II presents related work. Section III introduces problem definition. Section IV presents our system overview, key parts in the system, as well as the key algorithm. Section V introduces how the system is evaluated by city scale deployment and physical feature based simulation. Section VI concludes the paper respectively.

II. RELATED WORK

In this section, we give a literature review of related work from two aspects, *i.e.*, taxi behaviors and, incentives as well as mobile crowdsensing.

A. Taxi Behaviors and Incentives

Taxis are playing an irreplaceable role in a city's transportation system by providing reliable and customized travel services for passengers. Compared with other transportation modes including subway and bus service, taxi has no fixed routes, making it flexible and accessible from almost every corner of a city.

For the safety of passengers, taxis are required to be equipped with GPS trackers by law in many countries [26]. Smart phones also make it easy to record the traces of taxis. These GPS trajectories can be regarded as digital footprints of human mobility. Based on open taxi GPS dataset including Geolife [27], previous studies have addressed various research topics, including road map making [28], [29], urban mobility understanding [30], [31], [32], [33], [34], [35], city region function identification [36], [37], [10], traffic estimation and navigation [38] and location-based social networks [39], [40].

Due to uncertain and time-variable traffic and ride request demand, mobility prediction and ride request prediction are still two challenging tasks for researchers. Human mobility is believed to have limits of predictability [41]. However, with more available data and the usage of state-of-art machine learning and deep learning models, the prediction accuracy of taxi mobility has been improved remarkably in the past few years [42]. Previous study has also shown that ride requests follow the well-known densification power law, which may be used to predict or even synthesize ride request [43].

Driven by the selfishness of taxi drivers, the taxis are unevenly distributed in different areas and the overall efficiency is heavily harmed by the competition in over-supplied areas and supply-demand imbalance in under-supplied areas. To help to improve the performance of taxi drivers and shorten the waiting time of passengers, online taxi-hailing service [44] and

dynamic taxi dispatch system [45] are proposed to improve the scheduling efficiency of taxis. Ride-sharing service [46] is also proposed to increase the delivery capacity.

B. Mobile Crowdsensing

Under different names like participatory sensing or community sensing, crowdsensing has been proposed as a new scheme of collectively sharing data and extracting information to measure and map phenomena of common interest, by individuals with sensing and computing devices [20]. As smart phones become more powerful and equipped with GPS trackers and accelerometers, crowd sensing becomes widely adopted as a flexible and low-cost method of collecting sensing data.

Usually the system's objective is to maximize sensing quality, which might have different metrics in different studies, *e.g.*, k-depth coverage [47], and the system may have different constraints, *e.g.*, budget of rewards for participants. Used in many previous studies including place-centric crowdsensing [48] and people-centric sensing [49], [50], sensing coverage has been a major metric of evaluating sensing quality. While there are some previous studies that aim to maximize sensing quality under budget constraints [51], they design the scheme from a systematic view and do not consider the motivations of users, who may be selfish and strategic.

Auction-based mechanisms and game-theoretical models, *e.g.*, reverse auction [52] and Stackelberg game [21], are used to fix this problem. Furthermore, budget-feasible mechanism [53] and proportional share rule based compensation determination scheme [54] are proposed to guarantee strategy-proofness and budget feasibility. More discussion about auction-based mechanisms, as well as other incentive mechanisms, which may include lotteries, trust and reputation systems, can be found in previous surveys [55], [5], [56].

Auction-based incentive mechanisms can be well designed to possess a bunch of desirable theoretical properties. In real implementations, the strong assumptions of participants being rational and strategic, the obscure theories, and the complex payment rules make them less attractive and practical. The time sensitivity of allocating sensing tasks also make it less likely for the participants to think about every possible situation and give a bid that accurately reflects their utility.

Compared with previous works, the major novelties of this work lie on three aspects. First, our *PAS* optimizes the *sensing coverage quality* for ride sharing based MCS, which considers both the size and the balance of sensed area over spatial and temporal domain. Second, our prediction based actuation planning algorithm considers both the upcoming ride requests over the city and the future mobility of vehicles to make actuation decisions, which not only improves the actuation effectiveness, but also lowers the average actuation cost. Finally, when optimizing the *sensing coverage quality*, our *PAS* tries to match ride requests for ride sharing vehicles, which avoids the conflict of the tasks of mobile crowdsensing and carrying passengers.

III. PROBLEM DEFINITION

This section introduces how the actuation planning problem is formulated to optimize *sensing coverage quality* for ride sharing vehicular MCS. We present preliminary definition and background in Section III-A. Subsequently, we discuss the objective of actuation planning, *sensing coverage quality* in Section III-B. Finally, we formulate the problem of actuation planning in Section III-C.

A. Definitions and Background

Fig. 2 illustrates the architecture of ride sharing vehicular MCS, which consists of four systems. The mobile sensing system is composed of ride sharing vehicles equipped with sensors and GPS, which keeps on collecting data when vehicles move around the city. The sensors are equipped according to the application requirement, such as air pollution, temperature, noise, video, etc. The collected data, as well as location and time, is sent to the actuation system. In order to derive "good" data collection for the learning system, the actuation system decides which vehicles to actuate, the corresponding actuation routes and incentives. The decision is made based on the actuation availability and raw data collection from vehicles, as well as data request from the learning system. The details of actuation system design will be discussed in Section IV. The learning system infers and maps the phenomena of common interest, which can be used by the application system. Our actuation system is independent of any particular applications and suits multiple types of learning and application systems.

Given the spatial resolution r_s, the area of interest is discretized into N_x by N_y congruent *grids* with size of r_s by r_s. Each grid is represented by (x_i, y_j), where i, j are index of longitude and latitude. Similarly, given the temporal resolution r_t, time is discretized into time slice t_k, where k is the index of time. To be noticed, according to average vehicle speed, r_s and r_t are set so that a vehicle covers at most r_s within r_t. The key definitions in the paper are defined as follows.

Participant: A participant is a ride sharing vehicle equipped with sensors, which is denoted as p and belong to the ride sharing vehicle fleet P, i.e. $p \in P$. Each participant moves inside the area of interest and keeps collecting data along the trajectory. The location of p at time slice t_k is represented as $(x^p_{t_k}, y^p_{t_k})$ and derived from GPS.

Actuation Period: The actuation period is decided by the learning system and application system, which is the time length for selected vehicles to finish actuation tasks (routes). One actuation period T is composed of N_a minimum time slice $T = N_a r_t$. Different actuation period is expressed as T_l, where $l \in N$.

Actuation Task: An actuation task refers to a route that the actuation system gives to a participant p to cover within the actuation period T. The route is composed of a sequence of coordinates for each r_t and denoted as

$$\left\{(x^p_\tau, y^p_\tau), (x^p_{\tau+r_t}, y^p_{\tau+r_t}), ..., (x^p_{\tau+N_a r_t}, y^p_{\tau+N_a r_t})\right\},$$

where (x^p_τ, y^p_τ) is the original location of the participant p and τ is starting time of the actuation period.

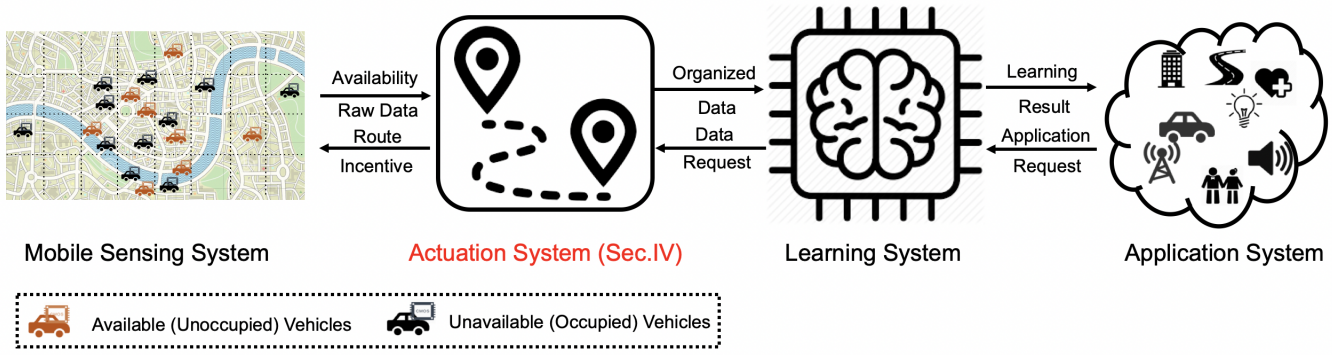


Fig. 2. The figure illustrates the architecture of ride sharing vehicular crowdsensing (MCS), which is composed of four systems. The mobile sensing system keeps on sensing data when vehicles move around the city. The actuation system selects vehicles from available ones (in orange) for actuation and plans their routes to improve *sensing coverage quality* of data collection. The learning system infers and maps the phenomena of common interest, which can be used by the application system.

Actuation Availability: The actuation availability refers to whether a participant p is available for actuation in the next actuation period. An available participant means the vehicle has no delivering task, e.g. no passengers on the vehicles (orange vehicles in Fig. 2), and the participant is willing to follow the assigned trajectory given the monetary incentive. At the beginning of each actuation period, each participant reports its actuation availability. A participant is called an "actuated participant" when it is selected for an actuation task, or "non-actuated participant" otherwise.

Budget & Monetary Incentive: The budget B_l is the total amount of money available to incentivize participants for the actuation period T_l. When a participant p is allocated an actuation task, a monetary incentive b_l^p is also assigned. Within each actuation period T_l, total monetary incentives should not exceed the given budget B_l.

Sensing Coverage: The sensing coverage C_l denotes all the data points collected by all participants during the actuation period T_l, which includes both "actuated" and "non-actuated" participants.

B. Actuation Objective

The objective of the actuation system is to derive data collection of large areas and even spatial distribution by dispatching part of the vehicle fleet with the given budget. In this paper, the objective function *sensing coverage quality* within actuation period T_l is denoted as $\phi(C_l)$, which consists of the total amount of sensed areas and the evenness level of the sensed data distribution. The evenness level represents how uniformly the sensed data is distributed over space. According to [10], information entropy of sensed area spatial distribution is adopted to measure the evenness level. Larger information entropy value means higher evenness level. Therefore, the overall *sensing coverage quality* is defined as the weighted sum of the size of sensed areas and the information entropy of sensed area spatial distribution, and calculated as

$$\phi(C_l) = \beta E(C_l) + (1-\beta)\log Q(C_l), \tag{1}$$

where $E(C_l)$ and $Q(c_l)$ denote the information entropy of sensed area spatial temperol distribution and the size of sensed areas within actuation period T_l. The log function is applied on $Q(C_l)$ to make two parts range in similar order of magnitude. $\beta \in (0,1)$ is a parameter, which tunes the importance of two parts.

$E(C_l)$ gets the maximum value of $log(N_x * N_y * T_l)$ when all vehicles are evenly distributed over the entire space of interest within the actuation period T_l, and gets the minimum value of 0 when all vehicles gather at the same grid. $Q(C_l)$ gets the maximum value of $N_x * N_y * T_l$ when all grids are covered by at least one vehicle, and gets the minimum value of 1 when all vehicles gather at the same grid. Therefore, $E(C_l)$ and $logQ(C_l)$ have the same value range of $[0, log(N_x * N_y * T_l)]$.

C. Problem Formulation

In order to optimize the *sensing coverage quality* $\phi(C_l)$ given the budget B_l within actuation period T_l, the actuation system needs to decide 1) which participants should be selected for actuation and 2) how should each of the selected participant be actuated (actuation route)? Therefore, the mathematical formulation of the actuation problem within actuation period T_l is defined as

$$\max_{I(p_i),\left\{(x^{p_i}_{T_l+r_t},y^{p_i}_{T_l+r_t}),\ldots,(x^{p_i}_{T_l+N_a r_t},y^{p_i}_{T_l+N_a r_t})\right\}} \phi(C_l)$$
$$s.t. \begin{cases} 0 \le x^{p_i}_{\tau} \le N_x r_s \\ 0 \le y^{p_i}_{\tau} \le N_y r_s \\ |x^{p_i}_{\tau+r_t} - x^{p_i}_{\tau}| \le r_s \\ |y^{p_i}_{\tau+r_t} - y^{p_i}_{\tau}| \le r_s \\ \sum_{i=1}^{|P|} b_l^{p_i} \cdot I(p_i) \le B_l \\ T_l + r_t \le \tau \le T_l + N_a r_t, l \in N \end{cases}, \tag{2}$$

where $I(p_i)$ is an indicator function. $I(p_i) = 1$ represents the participant p_i is selected for actuation and vice versa. The first two lines constrain that the actuation system only considers the data collection within the area of interest, while the third and fourth constraints show that each participant cover at most r_s within r_t. The fifth constraint shows that total monetary incentives of all selected participants should not exceed the given budget B_l.

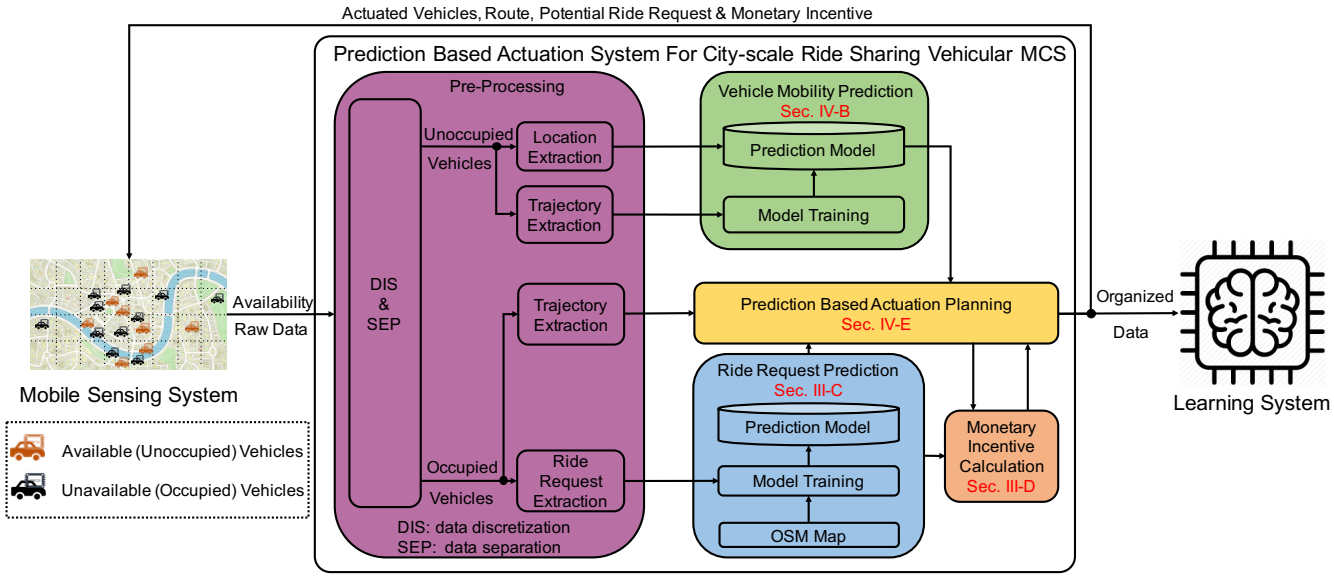


Fig. 3. The figure shows the architecture of our actuation system, which is mainly composed of five modules: *pre-processing*, *vehicle mobility prediction*, *ride request prediction*, *monetary incentive calculation*, and *prediction based actuation planning algorithm*.

IV. SYSTEM DESIGN

This section presents how the actuation system is designed to address two major challenges for optimal *sensing coverage quality*. Section IV-A provides an overview of the system architecture, which integrates a vehicle mobility prediction model and a ride request prediction model. The vehicle mobility prediction model guides the system to select vehicles with higher actuation effectiveness and is presented in Sections IV-B. The ride request prediction model, discussed in IV-C, helps the system select routes with new ride requests at the destination, thus lowering incentive cost and improving motivation. Then, we present how the monetary incentive is calculated based on the ride request prediction model in Sections IV-D. Finally, our prediction based actuation planning algorithm is introduced in IV-E.

A. System Overview

To address two major challenges for *sensing coverage quality* optimization, we design our actuation system based on two key observations. *1) The sensing coverage quality depends on data collection from both actuated and non-actuated vehicles.* Spending the budget actuating vehicles that plan to head for sparse areas is wasteful, since changing their trajectories does not bring *sensing coverage quality* improvement. In contrast, actuating vehicles that plan to head for busy areas and dispatching them towards sparse areas improves *sensing coverage quality* more. Therefore, the information of the vehicles' near future mobility is essential for effective actuation planning. *2) The monetary incentive of actuating one vehicle depends on whether the system can find a new ride request for the vehicle.* Given a new ride request at the destination, a vehicle is willing to accept a lower monetary incentive since it can earn money from the new ride [25]. Therefore, the information of new ride requests helps lowers the incentive cost, thus increasing the number of vehicles to be actuated and improving the *sensing coverage quality*.

Based on the two key observations, two prediction models are integrated in our actuation system. A vehicle mobility prediction model forecasts vehicles' recent trajectories, which guide the system to wisely select vehicles to actuate. To be specific, the actuation system intends to spend the budget on vehicles who are predicted to head for busy areas, and actuate them to sparse areas. A ride request prediction model forecasts near future ride requests over the area of interest, based on which the actuation system designs actuation routes for selected vehicles. By matching the ride request with an actuated vehicle, the individual incentive cost is lowered. As a result, more vehicles can be actuated for better *sensing coverage quality* with same amount of the given budget.

Fig. 3 shows the detailed architecture of our actuation planning system. At the beginning of each actuation period T_l, each vehicle automatically sends the collected information, including vehicle id, current location, and actuation availability for the coming actuation period and sensing data. Based on the information, the actuation system decides 1) which vehicles to be actuated, 2) the corresponding actuation routes, 3) the potential ride requests for actuated vehicles, and 4) monetary incentives for actuated vehicles, which are sent back to vehicles in the mobile sensing system. Unavailability is caused by two reasons: occupied by vehicles, or vehicles' unwillingness to be actuated. We assume vehicles follow the actuation task routes until the end of the actuation period if they set their status as 'available'.

The *Pre-Processing* module discretizes the area of interest and the time according to the spatial resolution r_s and temporal resolution r_t. The data from the mobile sensing system is separated into two parts: 1) data from unoccupied vehicles, who are looking for new passengers; and 2) data from occupied vehicles, who are currently delivering passengers.

Since the upcoming trajectories of unoccupied vehicles are unknown, their current locations are extracted for vehicle mobility prediction. Meanwhile, the historical trajectories of unoccupied vehicles can be used for vehicle mobility prediction model training. In contrast, the upcoming trajectories of occupied vehicles are deterministic, which is usually suggested by navigation Apps according to origins and destinations. Therefore, the ride request information can be extracted from occupied vehicles' trajectories, which can be used for ride request prediction model training. The upcoming trajectories of occupied vehicles are important information for prediction based actuation planning algorithm.

The *Vehicle Mobility Prediction* module, which is trained by historical trajectories of unoccupied vehicles, predicts upcoming trajectories of unoccupied vehicles. The prediction results, together with the deterministic trajectories of occupied vehicles, are fed into the *prediction based actuation planning algorithm* module, to wisely select vehicles for actuation, which brings more *sensing coverage quality* improvement. To be noticed, different mobility prediction models can be adopted in *PAS*. For simplicity but without loss of generality, a Markov based mobility prediction model is adopted in this paper, whose details can be found in Section IV-B.

The *Ride Request Prediction* module forecasts ride requests over the area of interest, based on which the *prediction based actuation planning algorithm* module designs actuation routes for the selected vehicles. The historical ride request data, which is extracted from data of occupied vehicles, is used for model training. Similar to the *vehicle mobility prediction* module, different ride request prediction models can be applied in *PAS*. For simplicity but without loss of generality, in this paper, a graph-based ride request prediction model is adopted, whose details is discussed in Section IV-C.

The *Monetary Incentive Calculation* module calculates the monetary incentive for each actuation route based on the ride request prediction at the destination. The calculation results are sent back to the *Prediction Based Actuation Planning* module for further optimization. The details of *Monetary Incentive Calculation* can be found in Section IV-D.

The *Prediction Based Actuation Planning* module selects the vehicles to actuate and designs actuation routes based on three factors: 1) the deterministic upcoming trajectories of occupied vehicles and the predicted upcoming trajectories of unoccupied vehicles; 2) the upcoming ride requests at the destination of actuation routes; 3) the monetary incentive cost of each selected vehicle and actuation route combination. The details will be discussed in Section IV-E.

B. Vehicle Mobility Prediction

The vehicle mobility prediction model offers information to guide the system to select vehicles for actuation. There are a lot of existing techniques can be integrated into our system for vehicle mobility prediction [57], [58], [59]. As we introduce in the previous section, the trajectories built for vehicles are discrete in both spatial and temporal domains. Therefore, a Markov Chain (MC) model is adopted, which is widely used for modeling transitions within discrete states [60]. In an MC model, each vehicle corresponds to a transition kernel C that describes its mobility pattern. The entry C_{ij} represents the probability that the vehicle moves from location i to location j. Each row C_i of the transition kernel represents the probability distribution of the vehicle moving from location i to its next location.

ALGORITHM 1: Vehicular Mobility Prediction Training

Input: Trajectory $X = x_1, x_2, ...x_n$, grid-to-area Map $f(\cdot)$.
Output: Direction Transition Kernel D.
Initialize:
  Transfer X into transition directions $d = \{d_1, d_2, ..., d_{n-1}\}$
  Set each element of $n \leftarrow \alpha$
for $i \in \{1, ..., n-1\}$ **do**
  Map grid ID into $Area$ ID $a_i \leftarrow f(x_i)$
  $n_{a_i d_i} \leftarrow n_{a_i d_i} + 1$
Calculate D using n based on (3)

As for the training of an MC model, given a trajectory of a vehicle, the maximum posterior estimation of C_{ij} is as follow:

$$\hat{C}_{ij} = \frac{n_{ij} + \alpha}{\sum_{j'} n_{ij'} + \alpha}, \tag{3}$$

where n_{ij} represents the number of times the vehicle moves from location i to location j, and α is a smoothing coefficient to avoid being divided by zero.

Once the estimated transition kernel $\hat{C}$ is acquired, it is used to predict the vehicle's future movements. The process can be formulated as: given a vehicle's current location $\overrightarrow{x_0}$, the estimated transition kernel $\hat{C}$, and a possible trajectory $\overrightarrow{x}_{(1:n)}$, the probability $p(l)$ that the vehicle moves along this trajectory in the future is calculated as:

$$p(l) = \prod_{i=1,2,..,n} \hat{C}_{\overrightarrow{x}_{i-1}\overrightarrow{x}_i} \tag{4}$$

However, we note the size of transition kernel correlates with the number of distinct locations visited by each vehicle, which means the computation complexity drastically increases. Also, we discover that most transitions happen within two connected grids, due to the speed limit in the city. Therefore, in order to reduce the computational work, we assume that each vehicle either moves to adjacent grids or stays in its current grid. This assumption decreases the length of each row C_i of the transition kernel to nine, representing the nine possible directions that a vehicle can move to. We denote the new transition kernel as the direction transition kernel D.

Moreover, we notice that vehicles within adjacent grids tend to follow similar mobility patterns [61]. Thus, to further limit the computation complexity while keeping the same spatial resolution, we first partition the city into several non-overlapping sub-areas that are larger than a grid, and then let grids within the same area share the same transition probability distributions, i.e. the same row of D.

ALGORITHM 2: Vehicular Mobility Prediction

Input: Direction Transition Kernel D, Current location x_0, Fixed trajectory $l = \{x_1, ..., x_m\}$,grid-to-area Map $f(\cdot)$
Output: Probability p.
Initialize:
  Transfer l into transition directions $d = \{d_0, d_1, ..., d_{m-1}\}$ (including current location x_0)
  $p = 1$
for $i \in \{0, ..., m-1\}$ **do**
  Map grid ID into area ID $a_i \leftarrow f(x_i)$
  $p \leftarrow p * D_{a_i d_i}$
Return p

Based on this mobility model, we design the mobility prediction module as having two parts, namely training and predicting. In the training parts, as shown in Alg. 1, a function $f(\cdot)$ that maps the grid ID into $sub-area$ ID is required. In general, the training part takes a vehicle's historical trajectory $X = x_1, x_2, ...x_n$ and the grid-to-sub-area map $f(\cdot)$ as inputs, and generates the direction transition kernel D as output. For each location of the trajectory, it maps the grid ID to $sub-area$ ID, and then counts the transition direction in a count matrix n. Based on n, it uses Eq. 3 to estimate the direction transition matrix D. In the prediction parts, as shown in Alg. 2, a probability is calculated to describe that a vehicle travels along a fixed trajectory l given its transition kernel D and current location x_0. The probability is calculated as product of the probability of each single transition in the fixed trajectory.

It is noticed that in practical situation, vehicle mobility can be affected by many factors, such as time, location, holiday, special events and etc. For example, the average driving speed can be very slow during rush hours. And special events like sports game make the traffic very crowd nearby the stadium. The mobility prediction model we adopt in this paper is robust to common factors like time and location. This model considers the different mobility pattern at different locations and time. However, influences of non-common factors like holidays and special events are not considered in this paper. In our future work, we will try to integrate more influential factors in the mobility prediction model to further improve its accuracy, which will make our actuation system more robust to non-common factors.

C. Ride Request Prediction

Our actuation system requires a model to predict ride request numbers over locations and time in a city. The prediction enables the system to match ride requests with vehicles, which makes vehicles willing to accept lower incentives. As a result, more vehicles can be actuated for better *sensing coverage quality* with the same budget.

The ride requests in the city can be predicted based on the discovery that spatial and temporal ride request patterns tend to repeat on a weekly basis [43]. But even for the same city, ride request numbers vary across different days in a week, at different time periods in a day, and across different areas.

The ride request pattern in a city can be modeled as a time-evolving graph, called a ride request graph (RRG). For each time interval, a directed graph is constructed as follows. Each grid of the city containing the source or destination of a ride request is considered as a node. Each source-destination pair is connected by a directed edge. The weight of the edge represents the ride request frequency between the same source and destination nodes. It is proven that for each time interval t, the number of edges $e(t)$ and the number of nodes $n(t)$ follow the Densification Power Law (DPL):

$$e(t) = Kn(t)^{\gamma}, \tag{5}$$

where K and $\gamma \in [1, 2]$ are constant. The number of edges grows linearly according to the number of nodes if $\gamma = 1.0$, while the RRG becomes fully connected if $\gamma = 2.0$.

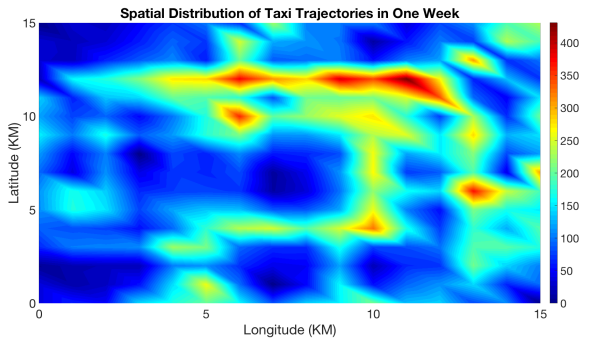
To predict the ride requests over time and location with the RRG, two attributes need to be learned: 1) the DPL factors (K and γ) which represent the temporal evolution property and 2) the spatial distribution of nodes in the RRG. The DPL factor can be calculated the ride request history data used to construct the RRG. The spatial property can be obtained with the help of OSM Points of Interest (PoI) such as traffic signals, businesses, schools, hospitals etc [62], which are used to infer the ride request popularity in different areas. This model has been shown to be accurate in [43] by comparing it to real-world dataset. More details can also be found in [43].

Similar to the vehicle mobility, in practical situation, ride requests can be affected by many factors, such as time, location, holiday, special events and etc. For example, there will be much more ride requests nearby the stadium when a sports game ends. In addition, ride requests also reveal different spatial and temporal patterns during weekdays and weekends. The ride request prediction model we adopt in this paper is robust to common factors like time and location. It is noticed that influences of non-common factors like holidays and special events are not considered in the current model. These factors will be integrated in the ride request prediction model in our future work to improve the prediction accuracy, as well as the robustness of the entire actuation system.

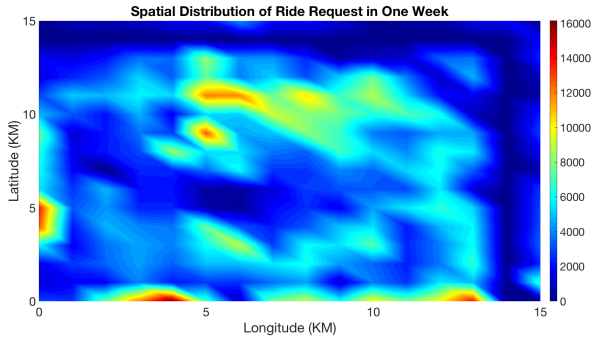
D. Monetary Incentive Calculation

The core idea of our monetary incentive mechanism is to lower the monetary incentive of actuating one vehicle by finding a new ride request for the actuated vehicle. To be specific, the higher chances that a vehicle can get new passengers at the destination of the assigned route, the lower monetary incentive the vehicle is willing to accept. With lower incentive cost, more vehicles can be actuated given the same amount of budget, thus bringing more improvement on *sensing coverage quality*.

As shown in Fig. 4, the spatial distribution of vehicles differs a lot from the spatial distribution of ride requests in the city of Beijing. By comparing Fig. 4(a) and (b), we found that the vehicles' estimation on ride request distribution may be incorrect. For example, in the area of longitude of



(a) Spatial distribution of 3000 taxis in the center of Beijing



(b) Ride request distribution in the center of Beijing

Fig. 4. (a) The distribution of 3000 taxis' trajectories in central area of the city of Beijing during $10min$ time interval; (b) The distribution of historical ride request in central area of Beijing City during the same $10min$ time interval with (a);

$2km \sim 5km$, latitude of $0km$, there are a lot of ride requests, but very few vehicles. At the same time, in the area of longitude of $10km$ and latitude of $11km$, where a lot of vehicles gather, the ride request number is not high. Based on the observation, it is promising to find new ride request for vehicles in sparsely sensed area. If a vehicle can be actuated to a sparsely sensed area with high ride request probabilities, the cost for actuating that vehicle can be decreased and the *sensing coverage quality* will be improved. Therefore, the monetary incentive $b_l^{p_i}$ offered to vehicle p_i for actuation period T_l is calculated as follows:

$$b_l^{p_i} = \max(r_{max} - r_u \cdot R(x^{p_i}_{T_l+N_a r_t}, y^{p_i}_{T_l+N_a r_t}, T_l+N_a r_t), r_{min}) \quad (6)$$

where r_{max} and r_{min} are upper and lower bound of monetary incentive to actuate one vehicle respectively. The maximum incentive r_{max} equals the maximum cost of delivering a passenger with the vehicle for the same route, which includes the gas, time and service cost. The incentive cost is reduced if the system finds a new ride request for the vehicle since it can earn additional money from taking the new ride and compensate the net cost of following our route. $R(i, j, t)$ denotes the predicted ride request distribution at location (i, j) and time t, which is estimated by the prediction model in Section IV-C. r_u represents the unit monetary incentive for one ride request, while r_{min} denotes the minimum monetary incentive for actuating one vehicle, which equals to the gas cost of the route.

ALGORITHM 3: Prediction Based Actuation Planning

Input: Current location X_0, budget B_l, vehicles availability, ride request model R, mobility prediction model D
Output: Actuated vehicles IDs, planned trajectories and monetary incentives for actuated vehicles
Initialize:
  Select vehicles and trajectory randomly until the budget is used up
  Output the initial feasible solution S based on actuated vehicles and D for non-actuated vehicles
while ϕ *not converges* **do**
  Select the grid $\hat{x}_m$ with maximum expected vehicles passing through
  Get the set of vehicles S_{tmp} which expect to pass through $\hat{x}_m$
  Compute and rank contribution to *sensing coverage quality* for vehicles in S_{tmp}
  Select the vehicle with minimum contribution and update its trajectory with monetary incentive
  Keep updating the trajectory until the budget constraint B_l is satisfied
  Update S and calculate the updated *sensing coverage quality* ϕ
Return $S^\star = S$

E. Prediction Based Actuation Planning

Since the optimization problem in Eq. 2 is NP-hard [19], we propose a fast, near-optimal heuristic-based algorithm to find an approximate solution. The core idea of the prediction based actuation planning algorithm is 1) utilizing the vehicle mobility prediction to select vehicles with higher actuation effectiveness and 2) utilizing ride request prediction to select actuation routes with lower incentive cost. To be specific, the algorithm finds the time slices and grids that expect to have many vehicles passing through in the near future. From these vehicles, the algorithm selects some and dispatches them to areas, where few or no vehicles expect to pass through.

Algorithm 3 shows the details of our prediction based actuation planning algorithm, which is based on Complementary Constructive Procedure (CCP). The algorithm first initializes a feasible solution S, which satisfying the constraints. This is easy to implement by randomly selecting vehicles until the budget is used up. Then, the algorithm finds the corresponding grids that expects to collect maximum data points, based on which the set of vehicles S_{tmp} that expects to pass through the grid is extracted. For vehicles in S_{tmp}, their expected contributions to the objective function ϕ is calculated based on the vehicle mobility prediction. The vehicle who expects to contribute least in S_{tmp} is assigned a new actuation route, which usually covers sparse area for *sensing coverage quality* improvement. This process will iteratively continue until the objective function ϕ converages.

V. EVALUATION

This section evaluates our *PAS* on three aspects: 1) optimizing *sensing coverage quality*, which is the objective of the actuation system; 2) matching ride requests with actuated vehicles, which is essential to vehicles' motivation and incentive cost; 3) lowering the average incentive cost, which decides the number of actuated vehicles as well as *sensing coverage quality* given the budget. We first introduce the evaluation setup for experiments on city scale deployment and physical feature based simulation in Section V-A. Then, the details of city scale deployment is presented in Section V-B. Finally, we show and analyze the experiment results on physical feature based simulation in Sections V-C respectively.

The evaluation focuses on the following aspects:

- Validate the effectiveness of our system's integration of the two prediction models. This is done by comparing the performance of our *PAS* with several baselines.
- Evaluate the performance of our prediction based actuation planning algorithm on three aspects: 1) optimizing *sensing coverage quality*, 2) matching ride requests with actuated vehicles, and 3) lowering the average incentive cost.
- Characterize the system performance of three aspects under different system setting up. Two key factors are discussed: the total number of vehicle fleet and the total budget amount. The former one represents the scale of actuation candidates, while the latter one decides the scale of actuated vehicles.

A. Evaluation Setup

Our *PAS* is evaluated by experiments on a city-scale deployment and a physical feature based simulation in the center area of Beijing China. The evaluation area occupies a size of $15km$ by $15km$, as shown in Fig. 5. The city-scale deployment is based on a taxi-based testbed, which checks how our *PAS* performs in real world. The physical feature based simulation is based on a historical Beijing taxi trajectory dataset, which is used to run large scale experiments for system performance characterization. Major evaluation setups are listed as follows.

General System Setup: The spatial resolution r_s and temporal resolution r_t are set as $1km$ and $2min$, since the average taxi speed in Beijing is $\sim 30km/h$ and the vehicle covers one grid within $2mins$. The actuation period is set as $5r_t$ ($10min$). The default number of whole vehicle fleet is set as 500, while the default budget is set as 1000 US dollars (USD). For incentive cost, we set $r_u = 2(USD)$, $r_{min} = 2(USD)$ and $r_{max} = 20(USD)$. This is because 2 USD equals the flag-down fare of Beijing taxis and 20 USD is enough to cover the cost for one trip ($\sim 10km$) in one actuation period ($10min$) even with serious traffic jams.

Performance Metric: Three metrics are introduced to measure three aspects of the system performance: optimizing *sensing coverage quality*, matching ride requests with actuated vehicles, and lowering the average incentive cost. 1) The ability to improve *sensing coverage quality* is measured by ϕ, which is the objective of the actuation system and defined as Eq. 1. The β is set as 0.5, which means equal importance between the size of sensed area and the evenness of spatial distribution, which is the case for most applications. Therefore, the *sensing coverage quality* in the evaluation is calculated as

$$\phi(C) = 0.5E(C) + 0.5\log Q(C), \tag{7}$$

where C is the all data points collected bby all participant vehicles in one actuation period. 2) The ability to match ride requests with actuated vehicles is measured by *ride request matching rate* r_{mat}, which is calculated as

$$r_{mat} = \frac{n_{mat}}{n_{all}}, \tag{8}$$

where n_{all} is the total number of vehicles in one actuation period and n_{mat} is the total number of vehicles who find new passengers at the end of the same actuation period. 3) The ability to lower the average incentive cost is measured by *average inventive cost* $\bar{b}$, which is calculated as

$$\bar{b} = \frac{1}{n_{act}} \sum_{i=1}^{n_{act}} b^{p_i} \tag{9}$$

where b^{p_i} is the incentive cost to actuate vehicle p_i in one actuation period and n_{act} is the total number of actuated vehicles.

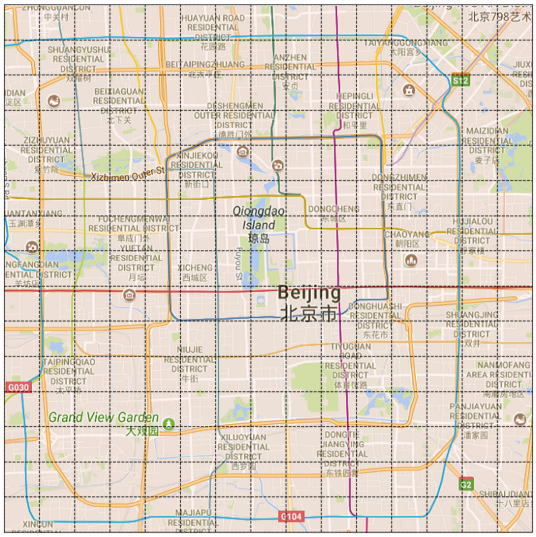


Fig. 5. Our *PAS* is evaluated at the center of Beijing, which occupies an area of $15km$ by $15km$. The evaluation area is discretized into grids of $1km$ by $1km$.

Baselines: Three baselines are adopted to validate different parts of our *PAS* on improving *sensing coverage quality*. These parts include mobility prediction model, ride request prediction model and our prediction based actuation planning algorithm.

- *No Actuation (NA)*: This method does not spend the budget actuating vehicles and all vehicles follow their original trajectories. By comparing *NA* with our *PAS*, we investigate the performance improvement from our entire system, which includes two prediction model and prediction based actuation planning algorithm.
- *Random Actuation (RND)*: This method randomly selects vehicles and corresponding routes for actuation until the budget is used up. The improvement brought from the two

Fig. 6. We evaluate our system with a city scale deployment on taxi-based testbed. An Android App, GPS Logger [63], is used to collect real-time trajectories of taxis.

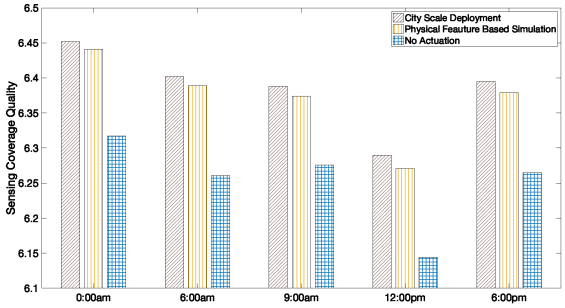
prediction models is checked by comparing *RND* with our *PAS*.

- *Random Actuation with Ride Request Prediction (RNDRQ)*: This method randomly selects vehicles but tries to match ride requests for vehicles when deciding actuation routes. Therefore, the cost to actuate each vehicle is expected to be lower than *RND*. The improvement from the ride request model is checked by comparing *RNDRQ* with *RND*. Meanwhile, the improvement from the mobility prediction model is checked by comparing *RNDRQ* with our *PAS*. This method is similar to state-of-the-art methods used by ride sharing companies like Uber, Lyft, and Didi, which dispatch their ride sharing vehicles according to real time ride requests [64], [65]. Since we have no access to their data and model details, we use this similar model to represent the state-of-the-art method. In our future work, when their data and model details are open to public, these models can be integrated in our system framework, which will further improve the performance of the entire actuation system.

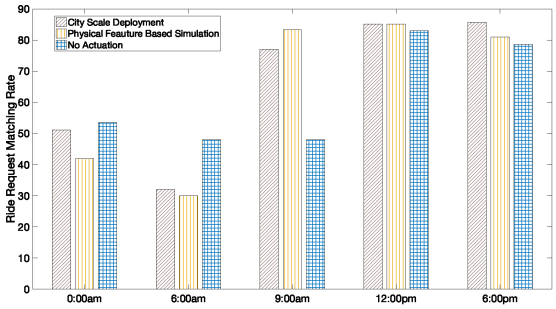
B. City Scale Deployment

To test how our *PAS* work in realistic environment, we deploy a city scale taxi-based testbed in the city of Beijing at five representative time periods. Besides the 0:00am, 6:00am, 12:00pm and 6:00pm, 9:00am is also considered since it is a peak time in a day. Our *PAS* first decides which vehicles to actuate and the corresponding actuation routes. Then the results are sent to selected vehicles, where a researcher stayed inside and suggested routes for the driver. For unactuated vehicles, the drivers are free to decide their routes. As shown in Fig. 6, during the whole actuation period, an Android App GPS Logger was fixed on vehicles for real-time trajectory data collection [63]. The routes from 230 actuated vehicles over 14 days were collected. Our deployment was approved under the university IRB *STUDY2017_00000342*.

Since the city scale deployment is affected by a lot of real world factors, which is not included in the simulation. To check the difference between city scale deployment and physical feature based simulation, we compare their *sensing coverage quality* values in Fig. 7(a). In order to check the *sensing coverage quality* improvement by our *PAS*, the *sensing coverage quality* without actuation is also plotted. First, at all time periods, the trend of *sensing coverage quality* of three methods are same: simulations are similar but slightly larger than deployments and both of them are larger than no actuation. This validates our *PAS*'s consistent ability to improve *sensing coverage quality* at different time. Second, *sensing coverage quality* values from deployments and simulations are similar. This shows that the physical feature based simulation is close to real world situation and can be used for large scale evaluations on system performance characterization. Finally, *sensing coverage quality* values from deployments are a little bit lower than those from simulations. This is because simulations show theoretically near-optimal results, while deployments are influenced by many practical factors, such as traffic jams, temporary road closure, lack of direct routes to follow the designed trajectories, etc., which prevent deployments from achieving exactly same performance as simulations.



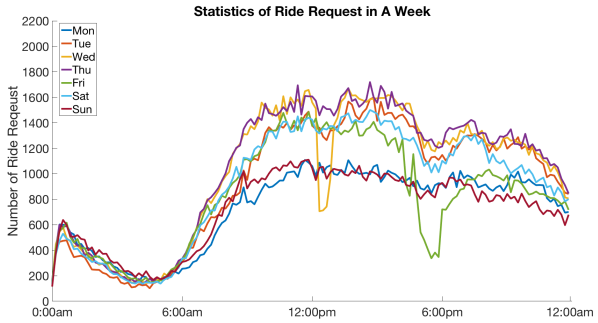
(a) Sensing Coverage Quality of Experiments at Different Time



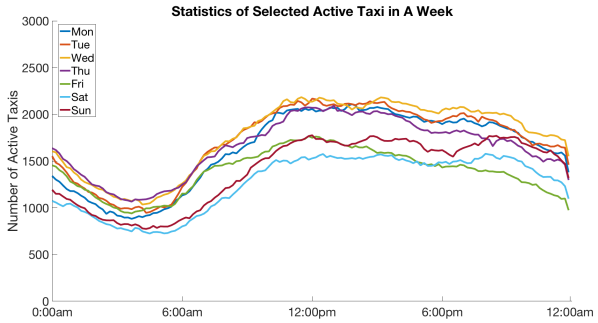
(b) Ride Request Matching Rate of Experiments at Different Time

Fig. 7. (a) The figure shows the *sensing coverage quality* from city scale deployment, physical feature based simulation and no actuation. Deployment results show improvements similar to but slightly lower than simulation results. Both of them show advantages over no actuation at different times; (b) The figure shows the ride request matching rate from real deployments, physical feature based simulation, and no actuation. Deployment results are similar to simulation results. Both figures show that the physical feature based simulations are close to real world deployments and can be used for large scale evaluations on system performance characterization.

Fig. 7(b) shows the ride request matching rate from real deployment, physical feature based simulation and no actuation. The similar ride request matching rate at all representative times in deployment and simulation proves that physical feature based simulation can be used to analyze

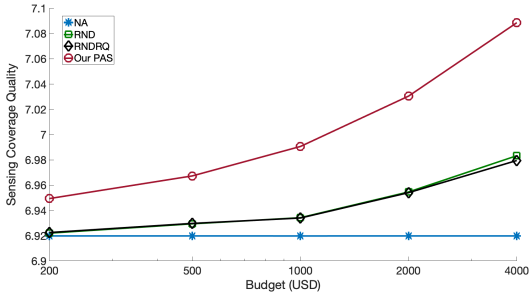


(a) Ride request within 10 minutes time interval on different days in a week.

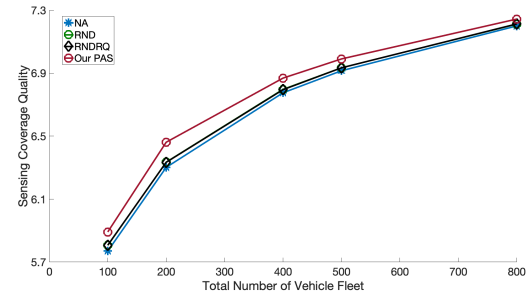


(b) Active taxi number within 10 minutes time interval on different days in a week.

Fig. 8. This figure shows the temporal distribution of ride requests and active taxis in a week. Active taxis have similar trend as ride request, which corresponds to human daily activity pattern in Beijing.



(a) Sensing Coverage Quality with Different Amounts of Budgets



(b) Sensing Coverage Quality with Total Number of Vehicle Fleet

Fig. 9. This figure shows how *sensing coverage quality* values of different methods are affected by the amount of budgets in (a) and total number of vehicle fleet in (b). Our *PAS* consistently show advantages over baselines with variant amounts of budgets and total numbers of vehicle fleet. To achieve similar *sensing coverage quality*, our method needs 200 USD while *RND* and *RNDRQ* need 2000 USD.

system performance in real world. Unlike *sensing coverage quality*, simulation results for ride request matching rate are not always higher than deployment results. This is because optimizing *sensing coverage quality* is the first priority of our *PAS*. The optimal ride request matching rate cannot be achieved at the same time. In addition, at 0:00am and 6:00am, no actuation scheme has higher ride request matching rate than *PAS* simulation and deployment results. When there are not many ride requests in the area (0:00am and 6:00am), the system has to choose the actuation routes that optimize *sensing coverage quality* but reduce ride request matching rate. When there are many ride requests in the area (9:00am, 12:00pm and 6:00pm), the system is able choose the actuation routes that optimize *sensing coverage quality* as well as matching ride requests with vehicles.

C. Experiments on Physical Feature Based Simulation

The physical feature based simulation is based on the historical taxi trajectories (November 2015) in the city of Beijing [66]. The dataset includes the information of taxi id, longitude, latitude, time stamp and occupancy flag. The occupancy flag infers whether a taxi is occupied by passengers. The taxi reported the above information every 60 seconds when they were operating. A ride request can be extracted according to the transformation of the occupancy flag, i.e. when and where the occupancy flag transfers from unoccupied to occupied. Every actuation period, 500 active taxis are randomly selected as the total vehicle fleet. To keep consistency, the same five representative time periods as deployments are selected for simulation. The first 3 weeks' data is used to train the mobility prediction and the ride request prediction model, while data from the rest days of the month is used to test the actuation system performance. The results are obtained from average values of 5 time periods in all testing days. We check the performance variance according to two keys factors: the budget and total number of vehicle fleet. The budget decides the potential amount of vehicles that can be actuated. Larger budget usually means more actuated vehicles for better *sensing coverage quality*. Total number of vehicle fleet decides the size of searching space that system has for actuation. Larger number of vehicle fleet usually brings more candidate choices for the system to actuate, which leads to better *sensing coverage quality*.

Fig. 8 shows the statistics of ride requests and active taxis in one week selected from the dataset. The number of ride requests and active taxis are calculated every 10 minutes. The active taxi number is calculated from a subset of around 3000 taxis. Ride request and active taxi counts show similar daily trends, corresponding to common supply and demand relations. Both show decreasing trends from 0:30am - 5:00am, when most people are asleep. After that, a increasing trend appears until 11:30am, as people go to work and school, do some shopping, etc. Both ride requests and active taxis maintain a high level from noon to midnight, which corresponds to the most busy time in Beijing.

TABLE I
SENSING COVERAGE QUALITY IMPROVEMENT WITH VARIANT AMOUNTS OF BUDGETS

Budget (USD)	200	500	1000	2000	4000
RND	0.7%	3.3%	6.5%	12.1%	22.2%
RNDRQ	0.9%	3.4%	6.4%	12.0%	20.8%
PAS	**10.6%**	**17.1%**	**26.5%**	**40.0%**	**60.7%**

TABLE II
SENSING COVERAGE QUALITY IMPROVEMENT WITH VARIANT TOTAL NUMBER OF VEHICLE FLEET

Total Number of Vehicle Fleet	100	200	00	500	800
RND	5.8%	7.4%	6.8%	6.5%	4.9%
RNDRQ	6.0%	7.1%	6.2%	6.4%	5.4%
PAS	**19.8%**	**36.6%**	**30.0%**	**26.5%**	**18.7%**

1) Sensing Coverage Quality: In order to evaluate how *sensing coverage quality* values are affected by two influential factors, we plot *sensing coverage quality* of four methods under variant amounts of budget and total numbers of vehicle fleet in Fig. 9. In addition, since it is difficult to understand the improvement on *sensing coverage quality* ϕ from different methods, the metric *sensing coverage quality improvement (*$SCQI$*)* is adopted, as shown in Table I and II. This metric evaluates how close the *sensing coverage quality* approaches the ideal-maximum value compared to the *sensing coverage quality* before actuation, and is calculated as

$$SCQI = \frac{\phi^{\star} - \phi_0}{\phi_{ideal} - \phi_0} * 100\%, \tag{10}$$

where $\phi^{\star}$ is the *sensing coverage quality* of the evaluated method and ϕ_0 is the *sensing coverage quality* before actuation. ϕ_{ideal} denotes the ideal-maximum *sensing coverage quality* and is calculated as

$$\begin{aligned} &\phi_{ideal} = \beta \log(N_x N_y N_a r_t) + (1-\beta) \log\left(N_a r * A\right) \\ &A = \min\left(|P|, N_x N_y\right), \end{aligned} \tag{11}$$

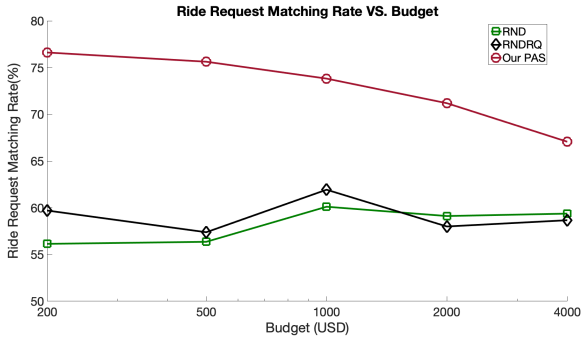
where N_x and N_y are the number of grids in longitude and latitude directions, and β is set as 0.5 as mentioned before. $N_a r_t$ is the actuation period and $|P|$ is the total number of vehicle fleet. It is noticed that $SCQIP$ cannot reach 100% since ϕ_{ideal} is the *sensing coverage quality* in ideal scenario. In the ideal case, all vehicles uniformly distributed over space at all time slices. However, since the limitation of initial spatial distribution of vehicles, this ideal case is usually impossible to achieve in real practices.

First, our *PAS* shows consistent advantages over three baseline with variant amounts of budgets. Especially when the budget is 4000 USD, our *PAS* brings 60.7% improvement while *RND* and *RNDRQ* only has 22.2% and 20.8% improvement respectively. Second, for all methods except *NA*, *sensing coverage quality* improves with increasing of the budget. This is because the higher budget allows more vehicles being actuated, which leads to higher *sensing coverage quality*. The $\sim 40\%$ advantage of our *PAS* comes from two parts. 1) The mobility prediction model guides our prediction based actuation planning algorithm to select vehicles which bring more *sensing coverage quality* improvement for actuation. 2) The ride request prediction model helps our prediction based actuation planning algorithm select routes that have lower incentive cost by matching the ride requests with vehicles. Finally, to achieve similar *sensing coverage quality* improvement, our *PAS* need 200 USD while *RND* and *RNDRQ* needs 2000, which is $10\times$ of our expense. This shows our *PAS*'s ability to save incentive cost.

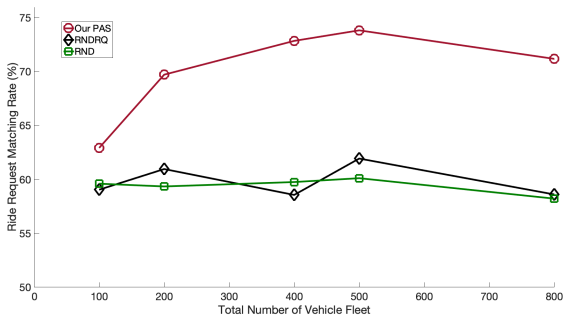
The effects of vehicle number on *sensing coverage quality* are also investigated as shown in Fig. 9(b) and Table II. Comparing the *sensing coverage quality* and its improvement of different methods, we see that $PAS > RND \approx RNDRQ > NA$. The *sensing coverage quality* of all methods improves with total number of vehicle fleet increases given the same amount of budget. In addition, our *PAS* shows consistent advantage of *sensing coverage quality* and improvement over baseline with variant total number of vehicle fleet. This proves that our *PAS* effectively selects "correct" vehicles and trajectories combinations that bring higher *sensing coverage quality* improvement. Recall the definition of *sensing coverage quality* $\phi(C)$, which is the trade-off between the sensed area $Q(C)$ and the evenness level of the spatial distribution $E(C)$. For $Q(C)$, more vehicles mean potentially more areas being sensed. For $E(C)$, more vehicles increase supply over demand and cause competition among vehicles, and the competition forces some vehicles to drive to areas with fewer vehicles to increase their probability of getting new passengers. Thus, the distribution of vehicles would be more balanced and $E(C)$ increases with total vehicle number. Therefore, the *sensing coverage quality* increases in all methods. It is noticed that our *PAS* still outperforms all baselines. Our *PAS* achieves up to 36.6% improvement at the vehicle number of 200, which is $5.1\times$ and $5.0\times$ higher than that of RND and $RNDRQ$ respectively. This shows the robustness and advantage of our *PAS* under variant uncertainties in vehicle availability.

To illustrate the practical meaning of *sensing coverage quality*, we take an improvement of 0.05 *sensing coverage quality* as an example. Since the default value of β is 0.5 in our experiment, an improvement of 0.05 on *sensing coverage quality* is equivalent to the improvement of the Q value of $e^{0.05/0.5} - 1 = 11\%$. The Q value denotes the size of area being senses. Given the map of $15km$ by $15km$, a 0.05 improvement on *sensing coverage quality* means that $124km^2$ more areas are sensed within each actuation period.

2) Ride Request Matching Rate: To evaluate how the budget affects the ride request matching rate of different methods, we plot r_{mat} with variant amounts of budgets in Fig. 10(a). First, a large budget leads to more actuated vehicles for all methods, but does not ensure large ride request matching rate, which is different from *sensing coverage quality*. This is because the first priority of our *PAS* is to improve *sensing coverage quality*. To the improvement of *sensing coverage quality*, *PAS* sacrifices ride request matching. Second, for different budgets, our *PAS* has up to $\sim 20\%$ larger ride request matching rate and than *RND* and *RNDRQ*. This shows that even though our *PAS* sacrifices ride request matching rate to guarantee the improvement of *sensing coverage quality*, it still

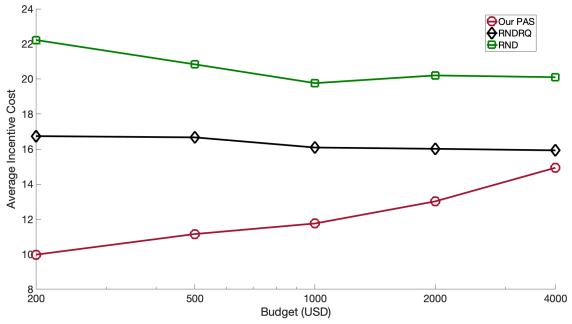


(a) Ride Request Matching Rate with Variant Budgets

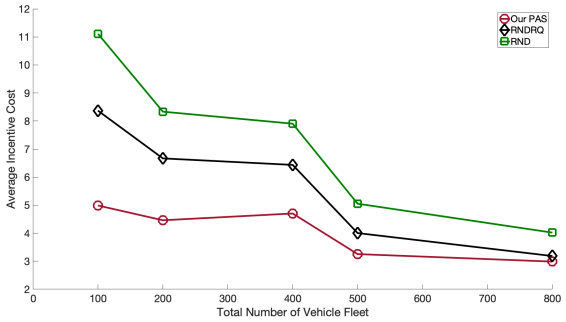


(b) Ride Request Matching Rate with Variant Total Number of Vehicle Fleet

Fig. 10. This figure shows our *PAS* consistently achieves higher ride request matching rate than *RND* and *RNDRQ* with variant budgets and total number of vehicle fleet.



(a) Average Incentive Cost with Variant Budgets



(b) Average Incentive Cost with Variant Total Numbers of Vehicle Fleet

Fig. 11. This figure shows our *PAS* consistently has consistent lower average cost than *RND* and *RNDRQ* with variant budgets and total numbers of vehicle fleet.

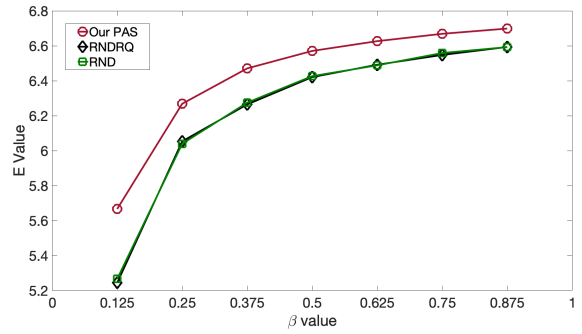
keeps a higher matching rate than other methods. This shows that compared to other methods our *PAS* are more attractive to vehicles since our *PAS* can find them new passengers with much higher probabilities.

The effect of the number of vehicles on ride request matching rate is also evaluated in Fig. 10(b). For all methods, the total numbers of vehicle fleet does not affect the ride request matching rate to much. In addition, our *PAS* consistently achieves higher ride request matching rates over baselines with variant total numbers of vehicle fleet. Our *PAS* achieves up to 73.8% ride request matching rate with 500 vehicles, which is 11.6% and 13.7% higher than *RNDRQ* and *RND*. This shows that the ride request prediction model, working with our prediction based actuation planning algorithm, does help more vehicles to improve the possibility of finding passengers. It it noticed that *RNDRQ* does not achieve similar ride request matching rate even with the ride request prediction model. This is because this method randomly select vehicles for actuation, and the route selection is based on the vehicle selection. This shows that ride request matching rate is decided by both route selection and vehicle selection.

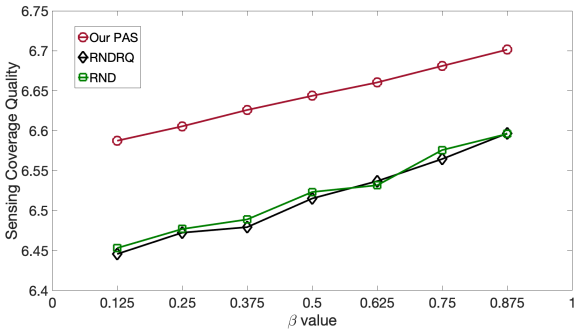
3) Average Incentive Cost: To evaluate how the budget affects the average incentive cost of different methods, we plot $\overline{b}$ with variant amounts of budgets in Fig. 11(a). First, *RNDRQ* shows consistent lower cost than *RND*, which shows that the ride request prediction does help reduces the average incentive cost. Second, our *PAS* shows consistent lower cost than *RNDRQ*. This shows that only adopting the mobility prediction model can further help lower the average incentive cost, since it helps our prediction based actuation planning algorithm select 'correct' vehicles, which have potentially lower incentive cost for actuation. Finally, with a higher budget, the average incentive cost of our *PAS* increases. This is because to improve *sensing coverage quality*, our *PAS* chooses the routes with higher average cost, which echoes the analysis of Fig. 10(a).

The effect of the number of vehicles on average incentive cost is also evaluated in Fig. 11(b). The average incentive cost trend of three methods are similar as that in Fig. 11(a), which validates the effectiveness of both mobility prediction and ride request prediction models. In addition, different from Fig. 11(a), the average incentive cost of all method decreases with higher total number of vehicle fleet. This is because more vehicles provide more available vehicle and route combinations, which potentially allows routes with lower incentive cost being selected. Furthermore, the average incentive cost converges to $\sim 3USD$, which is very close to lower bound of incentive cost r_{min} $(2USD)$. Therefore, increasing the total number of vehicle fleet is meaningful by providing higher flexibility on selecting vehicles and routes combinations with lower incentive cost.

4) The Impact of β Value: In order to check how the β value in Eq. 1 affects *sensing coverage quality* and its components, we plot the values of E and *sensing coverage quality* for our *PAS* and another two baselines in Fig. 12(a) and Fig. 12(b) respectively. As shown in Eq. 1, the E value denotes



(a) Information Entropy with Variant β Values



(b) Sensing Coverage Quality with Variant β Values

Fig. 12. This figure shows our *PAS* consistently has larger sensed area and higher *sensing coverage quality* than *RND* and *RNDRQ* with variant β values.

the information entropy of sensed area distribution, which measures how balanced the collected data is distributed. The maximum value of E is $log(N_x * N_y * T)$, when all vehicles are evenly distributed over space and time within the actuation period. First, our *PAS* consistently derives higher E values than *RND* and *RNDRQ* with different β values. This shows that our *PAS* is able to achieve more balanced distribution of data collection than baselines with the help of our prediction based actuation planning algorithm, which integrating two prediction models. Second, for all three methods, the E value increases with the increase of the β value. This is because that larger β value biases the system to focus more on the balanced distribution of data collection. Finally, with increase of the β value, the improvement of the E value decrease. This is because with large β value, the system puts more attentions on improving E and its value getting closer to saturation value with larger β.

With variant β, our *PAS* shows consistently advantages over two baseline concerning *sensing coverage quality*, which proves the robustness of our actuation system. The advantages of our *PAS* come from 1) the mobility prediction model helping our prediction based actuation planning algorithm to actuate vehicles with more *sensing coverage quality* improvement, and 2) the ride request prediction model helping our prediction based actuation planning algorithm selecting routes with lower incentive cost by matching the ride requests with vehicles. In addition, the increasing of β leads to improvement of *sensing coverage quality*. This is because larger β value bias the system to focus more on the information entropy improvement, which mainly comes from the improvement of data distribution balance (E value).

VI. CONCLUSION

This paper presents *PAS*, a prediction based actuation system that dispatches the ride sharing vehicle fleet for optimal *sensing coverage quality*. A near-optimal prediction based actuation planning algorithm is proposed which integrates 1) a mobility prediction model that guides the selection of vehicles to actuate and 2) a ride request prediction model to help match ride requests with vehicles, lowers average incentive cost and improves vehicles' motivation. Experiments on city scale deployment and physical feature based simulation shows that our *PAS* can achieve up to 40% more *sensing coverage quality* improvement and up to 20% more ride request matching rate than baselines. Additionally, to achieve a similar level of *sensing coverage quality* as the baseline, our *PAS* needs only 10% budget.

REFERENCES

[1] Jinwei Liu, Haiying Shen, Husnu S Narman, Wingyan Chung, and Zongfang Lin. A survey of mobile crowdsensing techniques: A critical component for the internet of things. *ACM Transactions on Cyber-Physical Systems*, 2(3):18, 2018.

[2] Charith Perera, Arkady Zaslavsky, Peter Christen, and Dimitrios Georgakopoulos. Sensing as a service model for smart cities supported by internet of things. *Transactions on Emerging Telecommunications Technologies*, 25(1):81–93, 2014.

[3] Weixi Gu, Kai Zhang, Zimu Zhou, Ming Jin, Yuxun Zhou, Xi Liu, Costas J Spanos, Zuo-Jun Max Shen, Wei-Hua Lin, and Lin Zhang. Measuring fine-grained metro interchange time via smartphones. *Transportation Research Part C: Emerging Technologies*, 81:153–171, 2017.

[4] Xinlei Chen, Yulei Zhao, and Yong Li. Qoe-aware wireless video communications for emotion-aware intelligent systems: A multi-layered collaboration approach. *Information Fusion*, 47:1–9, 2019.

[5] Robert Ighodaro Ogie. Adopting incentive mechanisms for large-scale participation in mobile crowdsensing: from literature review to a conceptual framework. *Human-centric Computing and Information Sciences*, 6(1):24, 2016.

[6] Xinlei Chen, Zheqi Zhu, Min Chen, and Yong Li. Large-scale mobile fitness app usage analysis for smart health. *IEEE Communications Magazine*, 56(4):46–52, 2018.

[7] Jing Lian, Yang Li, Weixi Gu, Shao-Lun Huang, and Lin Zhang. Mining regional mobility patterns for urban dynamic analytics. *Mobile Networks and Applications*, pages 1–15, 2019.

[8] Xinlei Chen, Aveek Purohit, Carlos Ruiz Dominguez, Stefano Carpin, and Pei Zhang. Drunkwalk: Collaborative and adaptive planning for navigation of micro-aerial sensor swarms. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*, pages 295–308. ACM, 2015.

[9] Shenggong Ji, Yu Zheng, and Tianrui Li. Urban sensing based on human mobility. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 1040–1051. ACM, 2016.

[10] Jing Yuan, Yu Zheng, and Xing Xie. Discovering regions of different functions in a city using human mobility and pois. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 186–194, 2012.

[11] Osamu Masutani. A sensing coverage analysis of a route control method for vehicular crowd sensing. In *Pervasive Computing and Communication Workshops (PerCom Workshops), 2015 IEEE International Conference on*, pages 396–401. IEEE, 2015.

[12] Susu Xu, Xinlei Chen, Xidong Pi, Carlee Joe-Wong, Pei Zhang, and Hae Young Noh. Vehicle dispatching for sensing coverage optimization in mobile crowdsensing systems. In *2019 18th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, pages 311–312. IEEE, 2019.

[13] Xinlei Chen, Susu Xu, Haohao Fu, Carlee Joe-Wong, Lin Zhang, Hae Young Noh, and Pei Zhang. Asc: actuation system for city-wide crowdsensing with ride-sharing vehicular platform. In *Proceedings of the Fourth Workshop on International Science of Smart City Operations and Platforms Engineering*, pages 19–24. ACM, 2019.

[14] Xinlei Chen, Xiangxiang Xu, Xinyu Liu, Shijia Pan, Jiayou He, Hae Young Noh, Lin Zhang, and Pei Zhang. Pga: Physics guided and adaptive approach for mobile fine-grained air pollution estimation. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, pages 1321–1330. ACM, 2018.

[15] Xinyu Liu, Xinlei Chen, Xiangxiang Xu, Enhan Mai, Hae Young Noh, Pei Zhang, and Lin Zhang. Delay effect in mobile sensing system for urban air pollution monitoring. In *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*, page 73. ACM, 2017.

[16] Xinlei Chen, Yu Wang, Jiayou He, Shijia Pan, Yong Li, and Pei Zhang. Cap: Context-aware app usage prediction with heterogeneous graph embedding. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(1):4, 2019.

[17] Xinyu Liu, Xiangxiang Xu, Xinlei Chen, Enhan Mai, Hae Young Noh, Pei Zhang, and Lin Zhang. Individualized calibration of industrial-grade gas sensors in air quality sensing system. In *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*, pages 1–2, 2017.

[18] Xiangxiang Xu, Xinlei Chen, Xinyu Liu, Hae Young Noh, Pei Zhang, and Lin Zhang. Gotcha ii: Deployment of a vehicle-based environmental sensing system. In *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM*, pages 376–377. ACM, 2016.

[19] Susu Xu, Xinlei Chen, Xidong Pi, Carlee Joe-Wong, Pei Zhang, and Hae Young Noh. ilocus: Incentivizing vehicle mobility to optimize sensing distribution in crowd sensing. *IEEE Transactions on Mobile Computing*, 2019.

[20] Raghu K Ganti, Fan Ye, and Hui Lei. Mobile crowdsensing: current state and future challenges. *IEEE Communications Magazine*, 49(11), 2011.

[21] Dejun Yang, Guoliang Xue, Xi Fang, and Jian Tang. Incentive mechanisms for crowdsensing: crowdsourcing with smartphones. *IEEE/ACM Transactions on Networking*, 24(3):1732–1744, 2016.

[22] Dong Zhao, Xiang-Yang Li, and Huadong Ma. How to crowdsource tasks truthfully without sacrificing utility: Online incentive mechanisms with budget constraint. In *INFOCOM, 2014 Proceedings IEEE*, pages 1213–1221. IEEE, 2014.

[23] Lingjie Duan, Takeshi Kubo, Kohei Sugiyama, Jianwei Huang, Teruyuki Hasegawa, and Jean Walrand. Incentive mechanisms for smartphone collaboration in data acquisition and distributed computing. In *INFOCOM, 2012 Proceedings IEEE*, pages 1701–1709. IEEE, 2012.

[24] Dejun Yang, Guoliang Xue, Xi Fang, and Jian Tang. Crowdsourcing to smartphones: Incentive mechanism design for mobile phone sensing. In *Proceedings of the 18th annual international conference on Mobile computing and networking*, pages 173–184. ACM, 2012.

[25] Xinglin Zhang, Zheng Yang, Wei Sun, Yunhao Liu, Shaohua Tang, Kai Xing, and Xufei Mao. Incentives for mobile crowd sensing: A survey. *IEEE Communications Surveys & Tutorials*, 18(1):54–67, 2016.

[26] Xingyu Huang, Yong Li, Yue Wang, Xinlei Chen, Yu Xiao, and Lin Zhang. Cts: A cellular-based trajectory tracking system with gps-level accuracy. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(4):140, 2018.

[27] Yu Zheng, Xing Xie, and Wei Ying Ma. Geolife: A collaborative social networking service among user, location and trajectory. *Bulletin of the Technical Committee on Data Engineering*, 33(2):32–39, 2010.

[28] Lili Cao and John Krumm. From gps traces to a routable road map. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '09, pages 3–12, New York, NY, USA, 2009. ACM.

[29] Yihua Chen and John Krumm. Probabilistic modeling of traffic lanes from gps traces. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '10, pages 81–88, New York, NY, USA, 2010. ACM.

[30] Marco Veloso, Santi Phithakkitnukoon, and Carlos Bento. Urban mobility study using taxi traces. In *International Workshop on Trajectory Data Mining and Analysis*, pages 23–30, 2011.

[31] L. I. Xiaolong, Gang Pan, W. U. Zhaohui, Q. I. Guande, L. I. Shijian, Daqing Zhang, Wangsheng Zhang, and Zonghui Wang. Prediction of urban human mobility using large-scale taxi traces and its applications. *Frontiers of Computer Science*, 6(1):111–121, 2012.

[32] B. Jiang, J. Yin, and S. Zhao. Characterizing the human mobility pattern in a large street network. *Physical Review E Statistical Nonlinear & Soft Matter Physics*, 80(1):1711–1715, 2009.

[33] Genlang Chen, Xiaogang Jin, and Jiangang Yang. Study on spatial and temporal mobility pattern of urban taxi services. In *Intelligent Systems and Knowledge Engineering (ISKE), 2010 International Conference on*, pages 422 – 425, 2010.

[34] Xidong Pi and Zhen Sean Qian. A stochastic optimal control approach for real-time traffic routing considering demand uncertainties and travelers' choice heterogeneity. *Transportation Research Part B: Methodological*, 104:710–732, 2017.

[35] Xinlei Chen, Xiangxiang Xu, Xinyu Liu, Hae Young Noh, Lin Zhang, and Pei Zhang. Hap: Fine-grained dynamic air pollution map reconstruction by hybrid adaptive particle filter. In *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM*, pages 336–337. ACM, 2016.

[36] G. Pan, G. Qi, Z. Wu, and D. Zhang. Land-use classification using taxi gps traces. *Intelligent Transportation Systems IEEE Transactions on*, 14(1):113–123, 2013.

[37] Guande Qi, Xiaolong Li, Shijian Li, and Gang Pan. Measuring social functions of city regions from large-scale taxi behaviors. In *IEEE International Conference on Pervasive Computing and Communications Workshops*, pages 384–388, 2011.

[38] Azam Ramazani and Hamed Vahdat-Nejad. Cans: context-aware traffic estimation and navigation system. *IET Intelligent Transport Systems*, 11(6):326–333, 2017.

[39] Nan Li and Guanling Chen. Analysis of a location-based social network. In *IEEE Cse'09, IEEE International Conference on Computational Science and Engineering, August 29-31, 2009, Vancouver, Bc, Canada*, pages 263–270, 2009.

[40] Xinlei Chen, Yulei Zhao, Yong Li, Xu Chen, Ning Ge, and Sheng Chen. Social trust aided d2d communications: Performance bound and implementation mechanism. *IEEE Journal on Selected Areas in Communications*, 36(7):1593–1608, 2018.

[41] Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.

[42] Alexandre De Brébisson, Étienne Simon, Alex Auvolat, Pascal Vincent, and Yoshua Bengio. Artificial neural networks applied to taxi destination prediction. In *Proceedings of the 2015th International Conference on ECML PKDD Discovery Challenge - Volume 1526*, ECMLPKDDDC'15, pages 40–51, Aachen, Germany, Germany, 2015. CEUR-WS.org.

[43] Abhinav Jauhri, Brian Foo, Jerome Berclaz, Chih Chi Hu, Radek Grzeszczuk, Vasu Parameswaran, and John Paul Shen. Space-time graph modeling of ride requests based on real-world data. *arXiv preprint arXiv:1701.06635*, 2017.

[44] Fang He and Zuo Jun Max Shen. Modeling taxi services with smartphone-based e-hailing applications. *Transportation Research Part C*, 58:93–106, 2015.

[45] Michal Maciejewski, Joschka Bischoff, and Nagel Kai. An assignment-based approach to efficient real-time city-scale taxi dispatching. *IEEE Intelligent Systems*, 31(1):68–77, 2016.

[46] Shuo Ma, Yu Zheng, and Ouri Wolfson. T-share: A large-scale dynamic taxi ridesharing service. In *IEEE International Conference on Data Engineering*, pages 410–421, 2013.

[47] Haoyi Xiong, Daqing Zhang, Guanling Chen, and Leye Wang. icrowd : Near-optimal task allocation for piggyback crowdsensing. *IEEE Transactions on Mobile Computing*, 15(8):2010–2022, 2016.

[48] Yohan Chon, Nicholas D Lane, Yunjong Kim, Feng Zhao, and Hojung Cha. Understanding the coverage and scalability of place-centric crowdsensing. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pages 3–12. ACM, 2013.

[49] Asaad Ahmed, Keiichi Yasumoto, Yukiko Yamauchi, and Minoru Ito. Distance and time based node selection for probabilistic coverage in people-centric sensing. In *Sensor, Mesh and Ad Hoc Communications and Networks*, pages 134–142, 2011.

[50] Chao Huang, Fengli Xu, Yong Li, Xinlei Chen, and Pei Zhang. Locally differentially private participant recruitment for mobile crowdsourcing. In *Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems*, pages 323–324. ACM, 2018.

[51] H. Xiong, D. Zhang, G. Chen, and L. Wang. Crowdtasker: Maximizing coverage quality in piggyback crowdsensing under budget constraint. In *IEEE International Conference on Pervasive Computing and Communications*, pages 55–62, 2015.

[52] Juong-Sik Lee and Baik Hoh. Sell your experiences: a market mechanism based incentive for participatory sensing. In *Pervasive Computing*

and Communications (PerCom), 2010 IEEE International Conference on, pages 60–68. IEEE, 2010.
[53] Georgios Amanatidis, Georgios Birmpas, and Evangelos Markakis. *Coverage, Matching, and Beyond: New Results on Budgeted Mechanism Design*. Springer Berlin Heidelberg, 2016.
[54] Zhenzhe Zheng, Fan Wu, Xiaofeng Gao, Hongzi Zhu, Guihai Chen, and Shaojie Tang. A budget feasible incentive mechanism for weighted coverage maximization in mobile crowdsensing. *IEEE Transactions on Mobile Computing*, PP(99):1–1, 2017.
[55] Luis G. Jaimes, Idalides J. Vergara-Laurens, and Andrew Raij. A survey of incentive techniques for mobile crowd sensing. *IEEE Internet of Things Journal*, 2(5):370–380, 2015.
[56] T. Luo, S. S. Kanhere, J. Huang, S. K. Das, and F. Wu. Sustainable incentives for mobile crowdsensing: Auctions, lotteries, and trust and reputation systems. *IEEE Communications Magazine*, 55(3):68–74, March 2017.
[57] Fangxin Wang, Feng Wang, Xiaoqiang Ma, and Jiangchuan Liu. Demystifying the crowd intelligence in last mile parcel delivery for smart cities. *IEEE Network*, 33(2):23–29, 2019.
[58] Fangxin Wang, Yifei Zhu, Feng Wang, and Jiangchuan Liu. Ridesharing as a service: Exploring crowdsourced connected vehicle information for intelligent package delivery. In *2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)*, pages 1–10. IEEE, 2018.
[59] Jing Yuan, Yu Zheng, Chengyang Zhang, Wenlei Xie, Xing Xie, Guangzhong Sun, and Yan Huang. T-drive: driving directions based on taxi trajectories. In *Proceedings of the 18th SIGSPATIAL International conference on advances in geographic information systems*, pages 99–108. ACM, 2010.
[60] Walter R Gilks, Sylvia Richardson, and David Spiegelhalter. *Markov chain Monte Carlo in practice*. CRC press, 1995.
[61] Siyu Chen, Yong Li, Wenyu Ren, Depeng Jin, and Pan Hui. Location prediction for large scale urban vehicular mobility. In *Wireless Communications and Mobile Computing Conference (IWCMC), 2013 9th International*, pages 1733–1737. IEEE, 2013.
[62] Openstreetmap 2016. openstreetmap contributors. https://www.openstreetmap.org/.
[63] Gps logger for android. 2017. google play shop. https://play.google.com/store/apps/details?id=com.mendhak.gpslogger.
[64] Jørgen Aarhaug and Silvia Olsen. Implications of ride-sourcing and self-driving vehicles on the need for regulation in unscheduled passenger transport. *Research in Transportation Economics*, 69:573–582, 2018.
[65] Zhe Xu, Zhixin Li, Qingwen Guan, Dingshui Zhang, Qiang Li, Junxiao Nan, Chunyang Liu, Wei Bian, and Jieping Ye. Large-scale order dispatch in on-demand ride-hailing platforms: A learning and planning approach. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 905–913. ACM, 2018.
[66] Beijing taxi data from computational sensing lab tsinghua university, 2016. http://sensor.ee.tsinghua.edu.cn/datasets.html.

Xinlei Chen is currently a postdoctoral research associate in Electrical Engineering Department at Carnegie Mellon University. He received the B.E. and M.S. degrees in Electronic Engineering from Tsinghua University, China, in 2009 and 2012, respectively, and Ph.D degrees in Electrical Engineering from Carnegie Mellon University, Pittsburgh, PA, USA. His research interests lie in mobile computing, crowd intelligence, large scale mobile cyber physical system, mobile embedded system etc.

Susu Xu received the B.E. degree from Tsinghua University, China in 2014, and M.S. degree from Carnegie Mellon University, Pittsburgh, PA, USA. She received her Ph.D. degree in Civil and Environmental Engineering in Carnegie Mellon University, Pittsburgh, PA, USA in 2019. Her research interests are mobile crowd sensing system, incentive mechanism, optimization, and time-series signal processing.

Jun Han is an Assistant Professor at the National University of Singapore with appointment in the Computer Science Department, School of Computing. His research interests lie at the intersection of sensing systems and security, and focuses on utilizing contextual information for security applications in the Internet-of-Things and Cyber-Physical Systems. He publishes across various research communities spanning security, sensing systems, and mobile computing (including S&P/Oakland, CCS, IPSN, TOSN, and HotMobile). He received his Ph.D. from the Electrical and Computer Engineering Department at Carnegie Mellon University as a member of Mobile, Embedded, and Wireless (MEWS) Group. He received his M.S. and B.S. degrees in Electrical and Computer Engineering also at Carnegie Mellon University. Jun also worked as a software engineer at Samsung Electronics.

Haohao Fu is pursuing a BA degree in University of California at Berkeley with an academic focus in Computer Science. His research interests are recommendation system, mobile network.

Xidong Pi received the B.E. degrees from Tsinghua University, China in 2014, and the M.S. degree from Carnegie Mellon University, Pittsburgh, PA, USA. He is now a Ph.D. candidate in Civil and Environmental Engineering in Carnegie Mellon University. His research interests are large-scale transportation system modeling and optimization, and mobility data mining.

Carlee Joe-Wong is an assistant professor of Electrical and Computer Engineering at Carnegie Mellon University. She received her A.B. degree (magna cum laude) in Mathematics, and M.A. and Ph.D. degrees in Applied and Computational Mathematics, from Princeton University in 2011, 2013, and 2016, respectively. Her research interests lie in optimizing various types of networked systems, including applications of machine learning and pricing to cloud computing, mobile/wireless networks, and ridesharing networks. From 2013 to 2014, she was the Director of Advanced Research at DataMi, a startup she co-founded from her research on mobile data pricing. She received the INFORMS ISS Design Science Award in 2014 and the Best Paper Award at IEEE INFOCOM 2012.

Yong Li is currently an associate professor of Electronic Engineering at the Tsinghua University. His research interests are in the areas of networking and communications, including mobile opportunistic networks, device-to-device communication, software-defined networks, network virtualization, and future Internet. He has published over 100 research papers, and has 10 Chinese and International patents. He has served as TPC Chair for WWW workshop of Simplex 2013 and TPC member of several international workshops and conferences.

Lin Zhang received all his degrees from Tsinghua University in Beijing (B.Sc. '98, M.Sc. '01, Ph.D. '06). He was the associate professor between 2008 and 2014 in Electronic Engineering Department in Tsinghua University. He was a visiting professor at UC Berkeley and Stanford University between 2011 and 2014. He is currently a professor and co-director of Tsinghua-Berkeley Shenzhen Institute since Sep.,2017. He is also the deputy director of Tsinghua Shenzhen International Graduate School since Jan.,2019 and the director of Talent Bureau of Shenzhen Municipality since May, 2019. He is a senior member of IEEE and a committee member of Chinese Information Theory Society.

Hae Young Noh is an associate professor in the Department of Civil and Environmental Engineering at Stanford University. Her research focuses on indirect sensing and physics-guided data analytics to enable low-cost and non-intrusive monitoring of cyber-physical-human systems. She is particularly interested in developing smart structures and systems to be self-, user-, and surrounding-aware to provide safe and resilient environments and improve user's quality of life, while reducing maintenance and operational costs. The result of her work has been deployed in a number of real-world applications from trains, to the Amish community, to eldercare centers, to pig farms. She received her Ph.D. and M.S. degrees in Civil and Environmental Engineering and the second M.S. degree in Electrical Engineering at Stanford University. She earned her B.S. degree in Mechanical and Aerospace Engineering at Cornell University. She received a number of awards, including the Google Faculty Research Awards in 2013 and 2016, the Dean's Early Career Fellowship in 2018, and the National Science Foundation CAREER award in 2017.

Pei Zhang is an associate research professor in the ECE departments at Carnegie Mellon University. He received his bachelor's degree with honors from California Institute of Technology in 2002, and his Ph.D. degree in Electrical Engineering from Princeton University in 2008. While at Princeton University, he developed the ZebraNet system, which is used to track zebras in Kenya. It was the first deployed, wireless, ad- hoc, mobile sensor network. His recent work includes SensorFly (focus on groups of autonomous miniature-helicopter based sensor nodes) and MARS (Muscle Activity Recognition). Beyond research publications, his work has been featured in popular media including CNN, Science Channel, Discovery Channel, CBS News, CNET, Popular Science, BBC Focus, etc. He is also a co-founder of the startup Vibradotech. In addition, he has won several awards including the NSF CAREER award, SenSys Test of Time Award, Google faculty award, and a member of the Department of Defense Computer Science Studies Panel.