This article has been accepted for publication in IEEE Transactions on Intelligent Vehicles. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIV.2022.3216318

A Crowd-Aided Vehicular Hybrid Sensing Framework for Intelligent Transportation Systems

Zhengqiu Zhu†, Yong Zhao†, Bin Chen*, Sihang Qiu, Zhong Liu, Kun Xie, *Member, IEEE,* and Liang Ma

***Abstract*—The development of Internet of Things (IoT) techniques enables the paradigm shift in traffic data collecting. In traditional practices of transportation system's constructions, traffic-related information is collected based on dedicated sensor networks, which are not only coverage-limited but also cost-consuming. With the enrichment of the concepts concerning "social sensors" and "social transportation", Sparse mobile crowdsensing (MCS) emerges as a promising sensing paradigm to collect data from only a few subareas by recruiting vehicles or mobile users with portable devices and to infer the data in unsensed subareas with acceptable errors at a low-cost manner. However, in real-world sensing campaigns, the Sparse MCS systems often fail to collect data from any subareas of interest since the assumption about sufficient participants is not always realistic. To be specific, the recruitment of participants is often limited by interest deficiency, privacy awareness, and distribution biases. To handle this problem, we introduce the dedicated sensing vehicles (DSVs) e.g., drones or driverless vehicles into traditional Sparse MCS to improve subarea coverage and inference performance. To achieve effective collaboration among DSVs and mobile users, we first design a crowd-aided vehicular hybrid sensing framework, which defines the order of task assignment for DSVs and mobile users as well as the budget allocation. In terms of DSVs route planning, we propose a three-step strategy, including optimal route searching, fused route selection, and final route determination. Moreover, mobile users are selected based on a proposed novel selection strategy. Experimental findings on two real-world datasets validate the effectiveness (with less inference error) of the hybrid sensing framework and the proposed strategies, in comparison with the user-only/DSV-only framework and five baselines. Results reveal important implications of applying the hybrid sensing paradigm in intelligent transportation systems to enhance data collection.**

***Index Terms*—Social sensors, Social transportation, Sparse mobile crowdsensing, Hybrid sensing framework, Intelligent transportation systems**

I. Introduction

By applying the social computing approach [1], many complex systems are managed in innovative ways. The traffic system is no exception, and the management mode of which is changed to optimize different systems (e.g., smart parking, intelligent transit) to provide better services (e.g., recommending parking spots, rescheduling trip plans) by leveraging IoT-enabled sensing techniques [2]. By connecting ubiquitous devices and facilities with various networks, IoT shows promising ability to provide efficient perception services for intelligent transportation systems (ITS). Besides the traditional way via infrastructure-supported sensor networks, "social sensors" was proposed to collect traffic-related information via social network and social media from a humanized perspective [3]. To conclude the emerging area in detecting traffic information from social networks, Prof. Wang introduced "social transportation" as a new direction for computational transportation study [4], [5].

Inspired by this promising direction, social sensing was proposed to leverage crowds as sensors to collect data in both physical space and cyber space with the increasingly ubiquitous and deep sensing capacity of portable devices. Based on the concepts of "social sensors" and "social sensing", mobile crowdsensing (MCS) [6] and vehicular crowdsensing (VCS) [7] have emerged as an alternative sensing paradigm for data collecting, which usually request a large number of participants to execute the sensing task with a full coverage [8], [9] or probabilistic coverage target [10]–[13] for ensuring high-quality sensing services. However, the practical application of such crowd sensing networks is often restricted by the availability of participants and the required large sensing costs. To further reduce sensing costs, current studies [14]–[17] investigated the inherent correlations embedded in the sensing data and proposed the Sparse MCS approaches. In this way, the number of required samples is reduced but a predefined data quality required by activity organizers is kept.

Though promising, Sparse MCS systems still face non-negligible problems when dealing with real-world participants. Most existing works assumed that enough participants in each subarea can be recruited [18], [19] or the user mobility can be estimated accurately [20]. However, the situation of insufficient participants often occurs due to interest deficiency, privacy awareness, and distribution biases [21]. Even after considering the mobility of users, few participants can be recruited at night or in sparsely populated rural areas [22]. Moreover, the sensors mounted in portable devices owned by mobile users are usually low-price and low-precision. It inevitably incurs non-negligible sensing errors to MCS systems. Therefore, these practical factors should be considered in the design of Sparse MCS for ITS.

The continuous emergence of intelligent mobile devices,

Manuscript received June 12, 2022; revised July 17, and October 15, 2022; accepted October 18, 2022. This study is supported by the National Natural Science Foundation of China (62173337, 21808181, 72071207), and Humanity and Social Science Youth Foundation of Ministry of Education of China (19YJCZH073). (Zhengqiu Zhu and Yong Zhao contributed equally to this work.) (Corresponding author: Bin Chen).

Zhengqiu Zhu, Yong Zhao, Bin Chen, Sihang Qiu, Zhong Liu and Liang Ma are with the College of Systems Engineering, National University of Defense Technology, Changsha, Hunan 410073, China. (e-mail: zhuzhengqiu12@nudt.edu.cn; zhaoyong15@nudt.edu.cn; chenbin06@nudt.edu.cn; sihangq@acm.org; philliPliu@263.net; maliang09a@nudt.edu.cn).

Kun Xie is with the College of Computer Science and Electronic Engineering, Hunan University, Changsha, China. (e-mail: cskxie@hnu.edu.cn).

Digital Object Identifier 10.1109/TIV.2022.XXXXXXX

such as unmanned (aerial) vehicles and driverless cars, provides significant support for sensing the traffic contexts. Different from mobile users, these devices are usually equipped with powerful onboard sensors and able to provide powerful processing ability [23]. Therefore, it is a trend to recruit idle IoT devices as dedicated sensing vehicles (DSVs) to complete sensing tasks [24]. However, involving different sensing resources in a same MCS task inevitably brings new problems. For example, recruiting the specialized sensor-equipped vehicles incurs extra costs. Moreover, how to utilize various sensing and computing abilities from different sensing resources to overcome the limitations of single-resource has not been explored. Therefore, in this paper, we aim to address the problem of *how to effectively schedule the DSVs and recruit suitable crowds to cooperatively complete sensing tasks under some practical constraints with the minimum inference errors in Sparse MCS paradigm for ITS*. In the next, we would elaborate upon the challenges of designing such a hybrid sensing framework and shed some light on the philosophies behind how we address them.

The first challenge is **how to design a hybrid sensing framework that fully exploits the capabilities of different sensing resources and further reduces inference errors**. The problem is definitely NP-hard since we not only assign tasks for mobile users, but also plan the paths for DSVs under a total budget constraint [15], [25]. Specifically, the inherent differences such as mobility and sensing accuracy embedded in the two kinds of sensing entities would incur great difficulties in task assignment. These factors inevitably influence the data inference in the unsensed subareas.

The second challenge is **how to determine the sensing route for each DSV for contributing maximum information**. In a sensing cycle, a DSV can collect data from several subareas along a route and may have numerous optional routes covering different subareas under the time constraint. The data in different subareas would incur different levels of error for the inference of missing data [18], [26], but it is difficult to recognize the subareas with maximum contribution in reducing inference error. The difficulty lies in multi-DSVs synergy and multi-factor coupling, such as the number of subareas on routes and information provided by subareas on routes.

The third challenge is **how to select proper mobile users with the information of DSVs' deployment**. User selection is constrained not only by the biased distribution of themselves, but also by the deployment of DSVs. It is a wise option to select the mobile users contributing data that is complementary to the information gathered by DSVs. Besides, the sensing error of users is ineluctable, and thus the difficulty also lies in how to carefully determine the selected users to alleviate the adverse impact of the sensing error.

To conquer these challenges, in this paper, we propose a crowd-aided vehicular hybrid sensing framework involving two different kinds of sensing resources, i.e., DSVs and mobile users. Specifically, we define the order of task allocation for the two sensing resources and design the budget allocation method. Different from the optimized path planning algorithms [27], [28], a three-step strategy is proposed to realize the DSVs route planning by taking into account the solution accuracy and searching efficiency. Moreover, mobile users are selected based on a novel selection strategy [18]. Extensive experiments are conducted on two real-world datasets, i.e., Flow [15] and TaxiSpeed [29]. Results not only show the effectiveness of our proposed framework, but also reveal the superiority of the novel strategies proposed in our framework for reducing the inference error over five baselines. In summary, this paper makes the following contributions:

- To address the first challenge, we propose a crowd-aided vehicular hybrid sensing framework to schedule sensing tasks for DSVs and mobile users under the budget and time constraints for each upcoming sensing cycle. To the best of our knowledge, this is the first work investigating the hybrid sensing framework for Sparse MCS involving different sensing resources. In this framework, the optimal budget allocation is determined via experiments. Moreover, we give the definitions of the informative subarea and the robust subarea. Based on that, we first plan the routes for each employed DSV; and then actively select the potential users under the rest budget constraint.
- To address the second challenge, we define the informative subarea based on the analysis of the relationship between selected subareas and the data inference since sampling the informative subarea can bring more information for data inference. Moreover, we propose a three-step strategy to determine a route for each DSV. Firstly, we find all optional routes for each DSV under the time constraint while only the routes that cover enough subareas are saved. Then, to contribute more information for inference, the optional routes of multiple DSVs are cross-fused in turn and we select fused routes based on the Local Beam Search (LBS) method. Finally, according to the selected fused route, we determine the corresponding route for each DSV.
- To address the third challenge, we define the robust subarea based on the analysis of the linear system from the geometric view, since sampling the robust subarea can alleviate the impact of sensing error. Then we propose the active user selection strategy, in which the number of users is determined under the rest budget by considering three factors (informative subarea, robust subarea, and the historical sampled times of a subarea). Notably, users located in the informative and robust subareas (with few sampled times) are preferentially selected since they are more conducive to data inference.

The remainder of the paper is organized as follows. We first review the related works in Section II. Then, the system model and problem formulation are presented in Section III. Next, we design the hybrid sensing framework in Section IV and illustrate the experimental setup in Section V. Finally, we show the performance evaluation in Section VI, and conclusions are drawn in Section VII.

II. Related works

In this section, we review the related work from three aspects: (1) Shifts in traffic data collecting, (2) Sparse MCS, and (3) Hybrid sensing scheme.

A. Shifts in Traffic Data Collecting

IoT technologies play a significant and fundamental role for traffic data collecting. With ubiquitous devices and sensors embedded in IoT-enabled ITS, traffic situations can be monitored in real time, and a vast amount of data is generated [2], [30]. The traditional way is to monitor the traffic context via infrastructure-supported devices. But these devices are expensive to deploy and maintain, and their functions are often limited due to their fixed locations. The emergence of "social sensors" [3] and "social transportation" [5] has greatly expanded the scope of collecting traffic data, besides from physical sensor networks, the applications of the two concepts can sense traffic context via social sensor networks. Social media and social networking platforms such as Weibo and Wechat provide ubiquitous chances for people to share ideas and information publicly about traffic. This crowdsourcing-based data collection paradigm provides diverse ITS services, such as geospatial data collection, road condition monitoring, urban traffic planning, social navigation, smart parking, and so on [31]. Further, IoT enables the detector embedded in moving devices, such as vehicles and mobile phones, thus provides a complete and alternative way for data collecting. Based on the development of "social transportation" and the concept of social sensing [32], leveraging both vehicles and crowds to collect traffic data in a low-cost manner is the current trend. Different from previous studies, we aim to unleash the potential of different sensing resources (DSVs and mobile users) collaboratively working in the same task to enhance data collection for ITS.

B. Sparse MCS

To ensure quality-aware sensing services, traditional MCS systems would recruit a large number of participants in the target sensing areas. As a consequence, these systems cost a lot in the task allocation process. As almost all physical conditions monitored are continuous, sensory data generally exhibits strong spatial-temporal correlations, thus the environment ground truth matrix often has a low-rank feature [33]. With this insight, a novel paradigm, namely Sparse MCS [34], was systematically proposed to collect data from only a few sensing subareas. In this way, many Sparse MCS applications have been developed to provide innovative data collection solutions in traffic systems [35], [36]. Notably, subarea selection is a general issue in Sparse MCS. Researchers proposed an online framework to prioritize subarea with greater uncertainty of the sensing data [15], [19], [26], such as SPACE-TA [37]. However, the direct relationship between the uncertainty and the quality of data inference is still unclear in this framework. Another kind of subarea selection strategy is based on reinforcement learning (RL) techniques [16], [38], [39], which achieves a global optimal quality of data inference. However, the RL-based strategies usually need extensive training data to obtain the weight of each subarea while the training data may be scarce in practice. Different from the above-mentioned frameworks, the active Sparse MCS scheme (AS-MCS) [18] was devised based on the bipartite-graph, in which a matrix completion algorithm was proposed to recover the unsensed data in the presence of sensing and communication errors robustly and accurately. Different from the aforementioned studies, we devise a hybrid sensing framework involving two different entities (i.e., DSVs and mobile users) for Sparse MCS systems. Based on the work of AS-MCS, we give the definitions of informative subarea and robust subarea, then propose our active subarea selection strategy in this paper.

C. Hybrid Sensing Scheme

To the best of our knowledge, few studies [24], [40]–[44] investigated the hybrid sensing scheme in the current ITS domain. Among which, researchers in [40], [43] proposed a hybrid task allocation framework to integrate the opportunistic and participatory sensing mode. This framework not only recruits opportunistic users to complete sensing tasks in their daily activities, but also assigns participatory users to move specifically to perform tasks that are not executed by opportunistic users. However, the participatory users in HyTasker only refer to the users who are willing to change their activities intentionally for tasks, thus, they are different from DSVs that are exclusively employed for sensing tasks. Apart from that, several works [41], [42] studied the task allocation problem in MCS with different entities that contains mobile participants and static participants (city camera, roadside infrastructure, etc.). Wang *et al.* [41] proposed a reverse auction-based method to allocate sensing tasks to two types of participants under the budget constraint. In [42], a framework called HySense was devised combining mobile users with static sensor nodes to generate uniformly distributed space-time data under the dynamic coverage constraint. As for Vehicular Crowd Sensing applications [24], [45]–[47], the literature [24] proposed a hybrid approach to leverage the for-hired vehicle (FHV) and the DSV to provide fine-grained Spatio-temporal sensing coverage. Different from the above-mentioned studies, in this paper, the different entities (mobile users and the DSVs) are different from previous studies, and the hybrid sensing framework is devised for Sparse MCS for the first time.

III. System model and problem formulation

In this section, we first present the system model involving different sensing resources. Then the active subarea selection strategy and data inference method are introduced in brief. Finally, we formulate the crowd-aided vehicular hybrid sensing problem in Sparse MCS and provide a running example to illustrate how our approach contributes to intelligent transportation systems. Table I shows the main concepts and notations used in this paper.

A. System model

The monitored area is divided into N subareas, denoted as $\mathbb{S} \triangleq \{s_1, s_2, \ldots, s_N\}$. Meanwhile, the whole sensing campaign is also uniformly split into T sensing cycles with duration t_d, denoted as $\mathbb{T} \triangleq \{t_1, t_2, \ldots, t_T\}$. Thus, the ground truth data of the whole monitored area over T sensing cycles can be organized as a matrix $G_{N\times T}$. In each sensing cycle,

TABLE I
MAIN NOTATIONS AND CONCEPTS

Notations	Explanations
$\mathbb{S}$, N	The subarea set, with N subareas
$\mathbb{T}$, T	The sensing cycle set, with T sensing cycles
$G_{N\times T}$, $\hat{F}_{N\times T}$	The ground truth data matrix of N subareas in T sensing cycles; and the inferred one
V	The employed DSV set $V \triangleq \{v_1, v_2, \dots v_{n_v}\}$
R	The route of a DSV (a sequence of subareas)
$\Gamma(R)$	The travel time of a DSV along a route
c_v	The cost for each DSV in a sensing cycle
U_r	The recruited user set $U_r \triangleq \{u_1, u_2, \dots u_{n_u}\}$
c_u	The cost for a recruited user in a sensing cycle
e_u	The sensing error of a user, which follows the normal distribution, i.e., $e_u \sim N(0, \sigma_u^2)$
B_u	The total cost budget for a MCS task within a sensing cycle
$error()$ function	The specific metric for calculating the overall sensing error and in this paper it is relative error: $error(G, \hat{F}) = \frac{\sqrt{\sum_{t=1}^{T}\sum_{s=1}^{N}(g_{s,t}-\hat{f}_{s,t})^2}}{\sqrt{\sum_{t=1}^{T}\sum_{s=1}^{N}(g_{s,t})^2}}$

the campaign initiator offers the system a budget B_u to cover the cost of participants.

The system employs n_v DSVs to exclusively collect data from any subareas of interest, and the set of the employed DSVs is denoted as $V \triangleq \{v_1, v_2, \dots v_{n_v}\}$. These DSVs can travel among the subareas at an average speed υ and collect data from the subareas along their route. In a sensing cycle, the route *R* of a DSV can be expressed as a sequence of subareas that are reached by the DSV, i.e., $R = \{s_1, s_2, \dots s_{n_R}\}$, and the travel time $\Gamma(R)$ along a route should be constrained within the duration t_d of the sensing cycle, calculated as:

$$\Gamma(R) = \sum\nolimits_{k=2}^{n_R} d(s_{k-1}, s_k)/\upsilon,\;\; s_{k-1}, s_k \in R \tag{1}$$

where $d(s_{k-1}, s_k)$ denotes the distance between subarea s_{k-1} and s_k. In each cycle, a DSV performs sensing tasks starting from the subarea where it was at the end of the last cycle. Since DSVs are employed by the system for a long period, their remuneration can be paid daily, weekly, or longer. However, to facilitate the budget management in each sensing cycle, the remuneration of DSVs is evenly shared in each cycle, that is, the system should expend a cost c_v for each DSV in each cycle. Besides, the sensing error of DSVs is assumed as negligible, which benefits from its high-accuracy sensors.

It is worth to mention that some users can be recruited to perform the sensing task in their located subarea and the proportion of the subarea covered by users in all subareas is denoted as user coverage p_u due to the biased distribution. In a sensing cycle, the Sparse MCS system will recruit n_u users, denoted as $U_r \triangleq \{u_1, u_2, \dots u_{n_u}\}$, and expend a cost c_u for each recruited user. The sensing error e_u of users is formulated as a random variable that follows the normal distribution, i.e., $e_u \sim N(0, \sigma_u^2)$.

B. Active subarea selection strategy and data inference

When selecting subareas to sense, two major problems should be solved: **(1) how many subareas to select and (2) where to sense to achieve better inference of missing data**. Inspired by this work [18], a novel subarea selection strategy is proposed to explore the inherent correlations between sampled subareas and missing data in terms of data inference directly. Then, we give the solutions for the two problems.

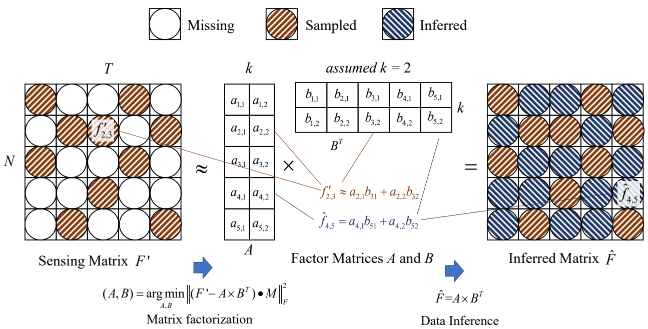


Fig. 1. Data inference based on matrix factorization.

The data inference based on matrix factorization is shown in Fig. 1. Specifically, a binary sampling matrix *M* is employed to mark whether a subarea is sampled or not. If the subarea s_i is sampled in the sensing cycle t_i, the element $m_{i,j} = 1$, otherwise $m_{i,j} = 0$. Correspondingly, a sensing matrix F' is used to record the sampled data, where the element $f'_{i,j}$ = 0 when the data in the subarea s_i in the sensing cycle t_i is missing. The sensing matrix $F'_{N\times T}$ generally has the low-rank feature, thus it can be factorized into factor matrices $A_{N\times k}$ and $B_{T\times k}$, where *k* is the rank of the sensing matrix. Using the Singular Value Decomposition [48] (SVD) or the Alternate Least Square (ALS) method [49], the factor matrices can be obtained by minimizing the loss as:

$$(A, B) = \underset{A,B}{\arg\min} \left\|(F' - A \times B^T) \bullet M\right\|_F^2 \tag{2}$$

where the $\|\|_F$ is the Frobenius norm. As the factor matrices *A* and *B* obtained, the inferred matrix $\hat{F}$ can be estimated as $\hat{F}$ = $A \times B^T$, and each missing data $\hat{f}_{i,j}$ can be inferred as $\hat{f}_{i,j} = a_i b_j^T$, where the $a_i = [a_{i,1}\; a_{i,2}\; \dots\; a_{i,k}]$ is the *i*-th row of matrix *A* and $b_j = [b_{j,1}\; b_{j,2}\; \dots\; b_{j,k}]$ is the *j*-the row of matrix *B*. Meanwhile, each sampled data $f'_{i,j}$ also corresponds to an approximated equation as $f'_{i,j} \approx a_i b_j^T = \sum_{r=1}^{k} a_{ir} b_{jr}$, where a_{ir}, b_{jr} are the *r*-th element of the *i*-th row and the *j*-th row of factor matrices *A* and *B*. We take *k*=2 as an example, a sample $f'_{i,j}$ can be expressed as $f'_{i,j} \approx \sum_{r=1}^{2} a_{ir} b_{jr} = a_{i1} b_{j1} + a_{i2} b_{j2}$.

In a new sensing cycle t_{n+1}, the factor matrix *B* will have a new row, that is, $B_{k\times(n+1)}^T = \begin{bmatrix} B_{k\times n}^T & b_{n+1}^T \end{bmatrix}$. Then, we can obtain the following theorem.

Theorem 1. *To determine the new row b_{n+1}^T in B, at least k subareas should be sampled in the sensing cycle t_{n+1}, where k is the rank of the sensing matrix.*

Proof. Without loss of generality, we simplify the relationship between the sampled data $f'_{i,n+1}$ and the dot product $a_i b_{n+1}^T$ as linear, i.e., $f'_{i,n+1}$ = $a_i b_{n+1}^T$. Suppose we sampled n_s subareas in sensing cycle t_{n+1}, each sampled data corresponds to an equation and we can build the following linear system with n_s equations:

$$\begin{cases} {f'}_{1,n+1} = a_1 b_{n+1}^T \\ {f'}_{2,n+1} = a_2 b_{n+1}^T \\ \quad \cdots \\ {f'}_{n_s,n+1} = a_{n_s} b_{n+1}^T \end{cases} \tag{3}$$

The linear system can be further transferred into matrix form, shown as follow:

$$\begin{bmatrix} {f'}_{1,n+1} \\ {f'}_{2,n+1} \\ \cdots \\ {f'}_{n_s,n+1} \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,k} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n_s,1} & a_{n_s,2} & \cdots & a_{n_s,k} \end{bmatrix} \begin{bmatrix} b_{n+1,1} \\ b_{n+1,2} \\ \cdots \\ b_{n+1,k} \end{bmatrix} \tag{4}$$

$$F'_{n+1} = C b_{n+1}^T \tag{5}$$

In the system, the matrix F'_{n+1} consists of sampled data that are known and the coefficient matrix *C* is composed of several row vectors of *A*, which is also known. Therefore, to determine the unknown vector b_{n+1}^T with *k* unknown elements, the coefficient matrix *C* should satisfy $rank(C) = k$. It means that the system requires at least *k* sampled data to build *k* equations. Here we also use an example of rank *k*=2 to illustrate the above equations. Given the sensing matrix F' (*k*=2) and its two samples (1,*n*+1), and (2, *n*+1) at subarea 1 and 2 in time cycle t_{n+1}, we can build a linear system with the following equation:

$$\begin{bmatrix} {f'}_{1,n+1} \\ {f'}_{2,n+1} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} b_{n+1,1} \\ b_{n+1,2} \end{bmatrix} \tag{6}$$

If $\begin{bmatrix} {f'}_{1,n+1} \\ {f'}_{2,n+1} \end{bmatrix}$ and $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ are known, it can be solved by using two equations since $\begin{bmatrix} b_{n+1,1} \\ b_{n+1,2} \end{bmatrix}$ only has two unknown variables. With the history samples, we can alternatively train the unknown variables in factor matrices *A* and *B* using samples taken in the new time cycle until their values converge to stable ones. □

From the above analysis, we can find that selecting different subareas to sample will bring different coefficient matrices *C*, which is key to solving the new row in *B*. Therefore, we decide which subareas to select by analyzing the corresponding *C*. In the linear system, if the row vector c in *C* is linearly dependent of other row vectors, the c is redundant for the linear system. Thus, the subarea corresponding to c is supposed to be unable to bring new information. Based on this analysis, we give the following definition of informative subareas.

Definition 1 (Informative subarea): *Given a set of subareas and the corresponding coefficient matrix, the subarea is informative if its corresponding row vector in the coefficient matrix is linearly independent of all row vectors corresponding to other subareas.*

Moreover, the potential sensing error, such as the sensing error of mobile users in our problem, may lead to the instability of the linear system and affect the accuracy of data inference adversely. **From the geometric view of the linear system, when the cosine values between the row vectors of the coefficient matrix are small, the linear system is more robust to the data with error**. Therefore, we give the following definition of robust subareas.

Definition 2 (Robust subarea): *Given a set of subareas and the corresponding coefficient matrix, the subarea is robust if the maximum cosine value between its corresponding row vector in the coefficient matrix and row vectors corresponding to other subareas is smaller than a certain threshold.*

Based on the above analysis and definitions, the active subarea selection strategy is described as: **the informative and robust subareas are preferentially selected and the total number of sampled subareas should exceed *k* (i.e., the rank of the sensing matrix)**. In fact, since the linear relationship between the sensing data and factor matrix is generally not complete, the number of sampled subareas is usually more than k. **Then, the new row in *B* is solved by using the Ridge Regression (RR) method [50] and we utilize the ALS method to update the factor matrix *A* and *B*. Finally, the inferred matrix $\hat{F} = A \times B^T$ can be obtained.** For more details about subarea selection and data inference, interested readers can be referred to [18].

C. Problem formulation

Based on the system model and data inference model, we define our hybrid sensing problem for Sparse MCS as follows.

Problem: *Given a MCS task with N subareas and T sensing cycles, we need to plan the route for each DSV and select mobile users at proper subareas to collect data cooperatively in each sensing cycle, then infer the missing data based on the current and historical sensing data. The objective is to minimize the inference errors while satisfying the budget and time constraints.*

$$\text{minimize} \qquad error(G, \hat{F}) \tag{7}$$

$$\text{subject to} \quad (c_v n_v + c_u n_u) \leqslant B_u \tag{8}$$

$$\Gamma(R_{t,v_i}) \leqslant t_d,\ t \in T,\ v_i \in V \tag{9}$$

$$\text{where } error(G, \hat{F}) = \frac{\sqrt{\sum_{t=1}^{T} \sum_{s=1}^{N} \left(g_{s,t} - \hat{f}_{s,t}\right)^2}}{\sqrt{\sum_{t=1}^{T} \sum_{s=1}^{N} \left(g_{s,t}\right)^2}} \tag{10}$$

The problem is formulated as Eq. (7-10). Relative error $error(G, \hat{F})$ between the inferred matrix $\hat{F}$ and the ground truth matrix G is used as the inference error (Eq. 10). The budget constraint in Eq. 8 limits the amount of employed DSVs and recruited users, and the time constraint in Eq. 9 limits the travel range of the DSVs within a sensing cycle.

D. Illustrative case

An illustrative case is shown in Fig. 2 to provide more details of our problem in a sensing cycle. A typical sensing scenario begins with a sensing task published by an organizer for obtaining fine-grained traffic information, e.g., traffic speeds in different road segments over a large-scale target area during a long time. Suppose the monitored area is split into

25 subareas, and two DSVs as well as five mobile users are distributed in different subareas. The budget in this sensing cycle is B_u=12 while the costs of DSVs and users are $c_v = 5$ and $c_u = 1$, respectively. The task assignment for DSVs and mobile users is completed before the start of this sensing cycle. We first search optional routes for each DSV under the time constraint, denoted as $R^1_{v_1}$, $R^2_{v_1}$, $R^1_{v_2}$, $R^2_{v_2}$. The route $R^2_{v_1}$ covers three subareas, which is more than that covered by $R^1_{v_1}$. Therefore, we would like to select $R^2_{v_1}$ for DSV-1 since it may be more conducive to inferring missing data than $R^1_{v_1}$. Besides, $R^2_{v_2}$ may not be a good choice for DSV-2 because it has an overlap with both $R^1_{v_1}$ and $R^2_{v_1}$, thus, $R^1_{v_2}$ is selected. After route planning for DSVs, appropriate users are selected to collect data with the rest budget $B_{urest} = 2$. User u_6 and u_8 are not selected because the subareas their resided are not informative or robust. The u_{14} is also excluded since the subarea s_{14} is covered by DSV-2. Finally, we recruit u_4 and u_{21} to execute the sensing task. When all the sensed data is submitted, the values of traffic speeds in the unsensed subareas are deduced by matrix completion algorithms. With the collected traffic information, the governor or the service initiator could adopt measures or impose influence on the urban context, for instance, encouraging citizens and intelligent vehicles to assist package delivery or suggesting intelligent vehicles to take other routes when meeting traffic congestion.

The above case is only an intuitive interpretation of our problem while more complicated factors should be considered in the DSVs route planning and user selection process.

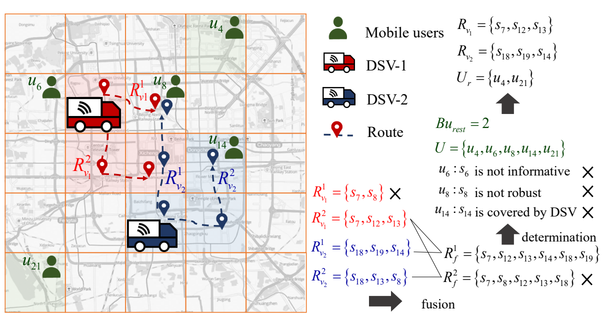


Fig. 2. An illustrative case of our problem.

IV. A Crowd-aided Vehicular Hybrid Sensing Framework

A. Framework design

DSVs route planning and user selection are two mutually influential tasks that share a total budget. More expenses are spent on DSVs, so that fewer users can be recruited. Therefore, the budget allocation is important, which are determined via experiments in this paper. Moreover, although the tasks are assigned to different sensing resources, the sensing data collected by these resources have inherent relationships. Therefore, it is important to determine the cooperative form of the two sensing resources. For DSVs route planning, we not only need to plan the route for each DSV under the time constraint, but also should consider the cooperation between multiple DSVs with the aim to contribute maximum information, such as avoiding the overlap between routes of different DSVs. For user selection, it is executed under the budget constraint. For one thing, we want mobile users to collect data that is complementary to the sensing data of DSVs. For another thing, we should select users located in the appropriate (informative and robust) subareas to reduce the negative impact of sensing error. The essence of task assignment is to select appropriate subareas for DSVs and users to sample. Based on the active subarea selection strategy described in section III B, we would like to select more informative and robust subareas. In addition, the historical sampled times of a subarea also have an impact on data inference [25]. Therefore, this factor is also considered in this work.

To effectively assign tasks for various sensing resources, this work proposes a crowd-aided vehicular hybrid sensing framework, namely DRPUS, which mainly contains two parts: DSVs route planning and user selection. The framework currently works offline for each sensing cycle for the sake of simplicity and practicality, in which all sensing tasks for both kinds of sensing resources are scheduled and assigned before a sensing cycle starts. In the DSVs route planning, a three-step strategy is devised, consisting of optional route searching, fused route selection, and final route determination. Then, the user selection is conducted to select appropriate mobile users under the rest budget constraint based on the result of DSVs route planning. The system architecture of the proposed framework is shown as Fig. 3. As we can see, the sensing tasks are motivated by practical demands, i.e., monitoring what is happening in a city, for instance, traffic situations. Our framework leverages the power of crowds, e.g., DSVs and mobile users, to collect useful information at a low-cost manner and to help understand how the traffic system is evolving. Based on that, a governor could further take actions to optimize different smart systems, e.g., smart parking and intelligent transit. In this paper, we design such a fixed task assignment order and perform the DSVs route planning before user selection based on the following reasons: (i) Subarea selection benefits from this way. With the support of powerful maneuverability, the DSVs are able to collect data in any subareas of interest and even collect at night. (ii) Miss data inference benefits from this way. In terms of sensing error, dedicated sensors with greater sizes and higher accuracy can be equipped on DSVs to collect more accurate sensing data. (iii) Privacy protection benefits from this way. Mobile users often risk their location and identification privacy when reporting data with actual positions. Therefore, recruiting mobile users first may face the privacy leakage problem or a shortage of participants. While our approach requires recruiting only a small number of willing users after the decision of the DSVs' routes. Based on our work, interested readers can consider designing more realistic and generic frameworks, for instance, a hybrid mode where DSVs are offline but mobile users are online recruited; or a framework that adapts to the number of DSVs and mobile users in different time periods.

B. DSVs route planning

To support the offline route planning, we have the following assumption about the travel time.

Assumption 1. *The travel time between each pair of subareas can be estimated in advance.*

The travel time between two subareas is related to the speed of a DSV and the distance between the subareas. In a sensing campaign, the distance between the subareas is usually constant, and the average speed of a DSV can also be obtained from the daily data. Therefore, the travel time between each pair of subareas can be estimated in advance. Considering that travel time may fluctuate due to real-time traffic conditions in the task execution stage, sufficient redundancy of travel time should be left to allow a DSV to move on time.

In the DSVs route planning, we mainly face two challenges. Firstly, a DSV may have numerous optional routes in a sensing cycle, even under the time constraint. If all optional routes are recorded, it will consume a lot of storage resources and more importantly complicate the route planning. Besides, in our system, there may be multiple DSVs. To take full advantage of multiple DSVs, we need to not only carefully select the appropriate route for each DSV, but also properly handle the relationship between the routes of multiple DSVs.

Therefore, we propose a three-step heuristic strategy to determine the route of each DSV, as shown in Fig. 4. The optional route searching is firstly performed for each DSV. In general, the more sampled subareas will bring more information. Therefore, we only record the optional routes that cover a considerable number of subareas. Then, the fused route selection is conducted based on the LBS method [20], in which the optional routes of multiple DSVs are cross-fused and only the best N_k fused routes are retained in each fused step. After that, a fused route is selected based on the weighted random selection. Finally, the route of each DSV is determined based on the selected fused route.

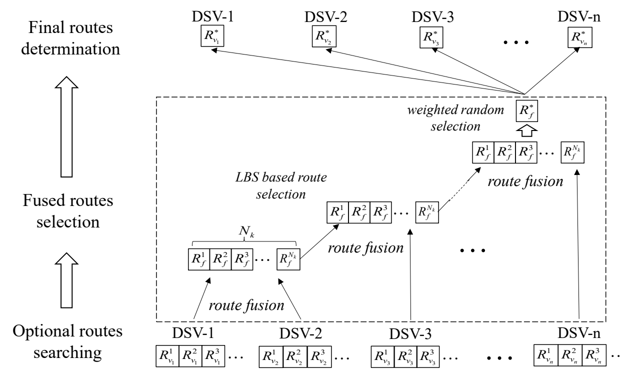


Fig. 4. Three-step DSVs route planning strategy.

a) optional route searching: Based on Breadth-First Search Method (BFS), we propose an optional route searching algorithm to find all optional routes meeting our requirement, as shown in Algorithm 1. Given the travel time matrix Φ, the initial subarea s_b of the DSV, and the threshold n_{th}, we use the open route list (*RL_open*) and the close route list (*RL_close*) to record the incomplete routes and complete routes, respectively. The definition of the incomplete route and the complete route is given as follows:

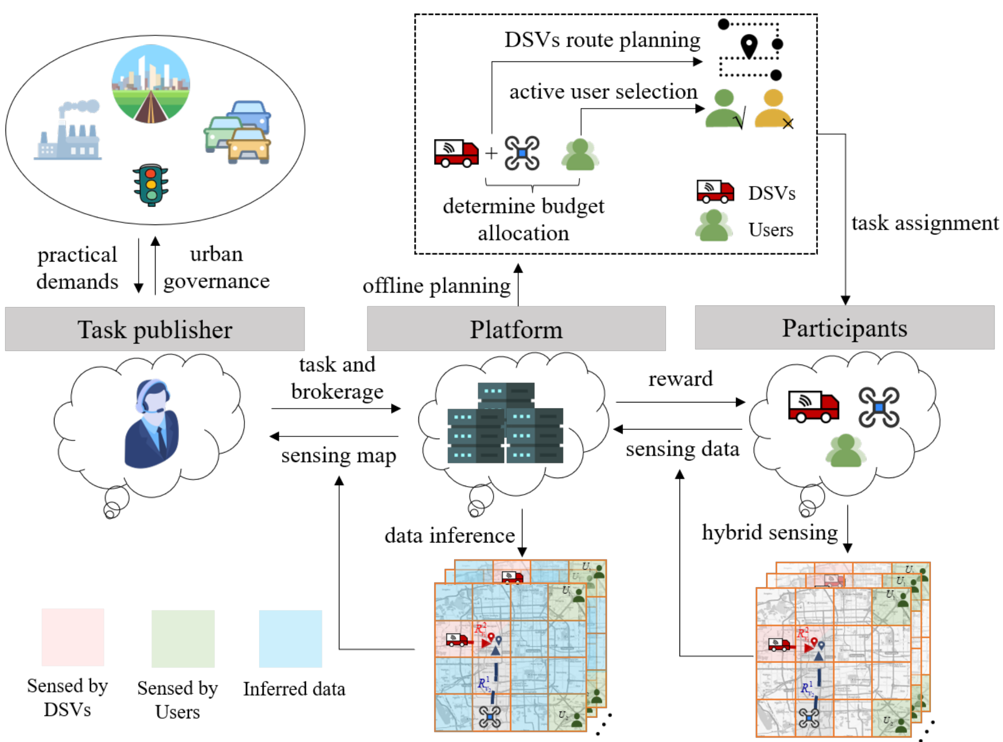


Fig. 3. System architecture of the crowd-aided vehicular hybrid sensing framework for urban actuation.

Algorithm 1 Optional route searching based on BFS

Input: Φ, s_b, t_d, n_{th}

add the route $R = \{s_b\}$ into *RL_open*
for *R* in *RL_open* **do**
  $S_{next} = \{s|\Gamma(R) + \Phi(s_l, s) \leqslant t_d, s \in \mathbb{S}\}$
  if $S_{next} = \emptyset$ **then**
    if $N(R) > n_{th}$ and $R \notin$ *RL_close* **then**
      add the route *R* into *RL_close*
  else
    for s_{next} in S_{next} **do**
      $R_{new} = \mathrm{add}(R,\ s_{next})$
      if $R_{new} \notin$ *RL_open* **then**
        add the route R_{new} into *RL_open*
return *RL_close*

Algorithm 2 Fused route selecting based on LBS

Input: *OR*, N_k, *A*

// search the fused route list
add $R_f = \emptyset$ into *FRL*
for OR_v in *OR* **do**
  for R_v in OR_v **do**
    for R_f in *FRL* **do**
      $R_{f\_new} = R_f \cap R_v$
      if $N_{info}(R_{f\_new}) > min\_N_{info}(FRL)$ and R_{f_new} not in *FRL* **then**
        add R_{f_new} into *FRL*
        $R_f^{\min} = \arg\min\_N_{info}(FRL)$
        delete $R_f^{\min}$ from *FRL*
// select with the random weighted method
for R_f in *FRL* **do**
  $weight(R_f) = N_{info}(R_f) \Big/ \sum_{R_f \in FRL} N_{info}(R_f)$
  $R_f^* = \mathrm{Random}(FRL, weight)$
return R_f^*

Definition 3 (Incomplete route and complete route): *Given a route $R = \{s_b, s_2 \dots s_l\}$, the successor subareas set $S_{next} = \{s|\Gamma(R) + \Phi(s_l, s) \leqslant t_d, s \in \mathbb{S}\}$. If $S_{next} \neq \emptyset$, the route R is incomplete and we can add new subareas into the route, otherwise, the route is complete.*

All routes in *RL_open* are handled successively. If the $S_{next} \neq \emptyset$, each subarea s_{next} in S_{next} is added into the incomplete route respectively to form the new route R_{new}, then the R_{new} will be added into *RL_open*. If the $S_{next} = \emptyset$ and the route coveres more than n_{th} subareas, the route will be added into *RL_close*. The above process will be performed iteratively until *RL_open* is empty, then all routes in *RL_close* are the optional routes for the next route fusion process.

Before adding the route into *RL_open* or *RL_close*, it is necessary to determine whether the same route has already existed in the list. We do not distinguish the order in which each subarea is sampled in an optional route, except for the first and last subarea of the route. For example, the route $R^1 = \{s_3, s_4, s_5, s_6\}$ and $R^2 = \{s_3, s_5, s_4, s_6\}$ are considered as the same route, while the R^1 and $R^3 = \{s_3, s_4, s_6, s_5\}$ are different because the last subarea in the route will affect the route planning in the next sensing cycle.

b) Fused route selection: The fused route selecting algorithm based on LBS is shown in Algorithm 2. The basic idea of our algorithm is to fuse the routes of each DSV in turn while only the best N_k fused routes are retained in the fused route list (*FRL*) in each step. According to the active subarea selection strategy, we use the number of informative subareas $N_{info}(R)$ covered by a route as the indicator, and a route with more informative subareas will be retained preferentially. If two routes have the same N_{info}, we count the total historical sampled times of the subareas covered by the two routes, respectively, and retain the route with fewer sampled times.

The best N_k optional fused routes can be obtained after optional routes of all DSVs are fused and we need to select one route R_f^* from them. A plain idea is to select the fused route with maximum N_{info}. However, there may be several routes that always contain more informative subareas than other routes, so they will always be selected as the plain idea suggesting. Then, DSVs can only collect data from the subareas covered by those routes alternately, which leads to the absence of data in many other subareas. To avoid this situation, the weighted random selection is adopted. The weight of each optional fused route R_f is calculated as $weight(R_f) = N_{info}(R_f) \Big/ \sum_{R_f \in FRL} N_{info}(R_f)$, and we randomly select the fused route R_f^* based on the weight.

c) final route determination: After the fused route is determined, the fused route needs to be split into the corresponding route for each DSV, so that the sensing task of each DSV can be truly determined. In the route fusion process, we record the index between the fused route and the corresponding routes of each DSV, thus, the route of each DSV can be determined by the index quickly.

C. Active user selection

The user selection is implemented after the determination of DSVs' routes, thus, we want to select the users to sample the subareas that are informative with regard to the subareas covered by DSVs, as the active subareas selection strategy suggests. Besides, the sensing error of the user is non-negligible, which may seriously impact the accuracy of data inference. Therefore, to reduce the negative impact, we need to select users in the robust subareas.

To solve the user selection problem, we propose the active user selection algorithm, as shown in Algorithm 3. Firstly, the number n_u of users that we can recruit is determined under the budget constraint. Given the total budget B_u and the cost of employed DSVs $c_v n_v$, the rest budget is calculated as $B_{urest} = B_u - c_v n_v$ and the number is $n_u = \mathrm{ROUNDDOWN}(B_{urest}/c_u)$. The optional user set is $U \triangleq \{u_1, u_2, \dots\}$ and the subareas covered by users are recorded in the set $S_u = \left\{s_u^1, s_u^2, \dots\right\}$. Some subareas in S_u may have already been covered by DSVs and they are excluded since we do not want to sample these subareas repeatedly, thus, the rest subareas only covered by users are recorded in $\bar{S}_u = S_u - R_f^*$. So far, we need to select n_u subareas from $\bar{S}_u$ and record them in the set S_{uv}. The historical sampling times of the unsampled subareas in the current sensing cycle are sorted.

Algorithm 3 Active user selection

Input: n_u, S_u, R_f^*, *A*, *M*, *U*

1: $\bar{S}_u = S_u - R_f^*$, $S_{uv} = R_f^*$
2: **while** $N(U_r) < n_u$ and $\bar{S}_u \neq \emptyset$ **do**
3: $\bar{s}_u^* = \arg\min_{\bar{s}_u \in \bar{S}_u}(N_{sampled}(\bar{s}_u))$
4: **if** $\bar{s}_u^*$ is informative and robust **then**
5: add $\bar{s}_u^*$ into S_{uv}
6: add $u_s = \text{locate\_in}(\bar{s}_u^*)$ into U_r
7: delete $\bar{s}_u^*$ from $\bar{S}_u$
return U_r

Those informative and robust subareas with few sampled times would be added to S_{uv} until the number $N(U_r) = n_u$ or $\bar{S}_u = \emptyset$ is satisfied.

D. Algorithm analysis

In the DSVs route planning, the runtime is mainly spent on optional route searching and fused route selection. Since the optional route searching is based on Breadth-First Search Method, the complexity of algorithm 1 can be formulated as $O(n_v \times T_{BFS})$ if the computation complexity of the BFS method is T_{BFS}. In terms of fused route selection, it is based on a greedy search method, namely Local Beam Search. The runtime of fused route selection is dependent on the beam width N_k and the number of DSVs n_v, and its complexity can be formulated as $O((n_v - 1) \times T_{LBS})$ if the computation complexity of LBS method is T_{LBS}. In addition, to successfully complete the data inference process, at least *k* samples should be collected in a sensing cycle. This requirement ensures that the matrix rank reaches *k*. Different subareas selected for sampling could impact the recovery performance. Therefore, informative subarea, robust subarea, and the number of historical sampled times are proposed as the guidelines for subarea selection.

V. Experimental Setup

A. Data sets

We use two real-world data sets, namely Flow and TaxiSpeed, to evaluate our proposed framework and strategies. The data sets contain different types of sensed data in representative IoT applications, like passenger flow index and traffic speeds. The detailed statistics of the two data sets are given in Table II.

TABLE II
STATISTICS OF TWO EVALUATION DATASETS

	Flow	**Traffic**
Project	DataFountain competitions	TaxiSpeed
City	Beijing	Beijing
Cycle length	1 hour	1 hour
Duration	7 days	4 days
Cell size	200m × 200m	road segments
Number of cells	171	100
Mean ± *Std.*	8.84 ± 13.68	13.01 ± 6.97 m/s

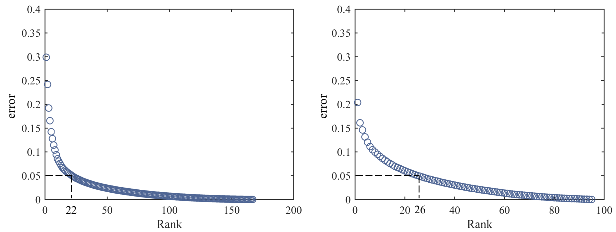
- The Flow data set contains the flow index of people, which were sensed from different regions in Beijing during the outbreak of COVID-19. The target area was divided into 997 subareas and the sensing campaign lasted from 2020-01-17 to 2020-02-15 with one hour as a sensing cycle. In this paper, we only adopt the sensing data in 171 subareas for 7 days after the data screening.
- The TaxiSpeed data set contains more than 33,000 trajectories of taxicabs in Beijing, and the average travel speed of taxicabs on each road segment was derived from trajectories. The sensing campaign lasted for four days from 2013-09-12 to 2013-09-15 and the sensing cycle is also one hour. According to [29], a road segment is considered as the subarea and we selected 100 road segments with valid data in this paper.

Strong spatial-temporal correlations in urban sensing data are the prerequisite for data inference. This point was discovered and verified in our previous work [15]. The two data sets serve as the ground truth matrix *G* with *N* subareas and *T* sensing cycles. Before a sensing cycle starts, our framework would select appropriate subareas and assign the tasks to DSVs and mobile users, for instance, collecting flow information in different subareas. Within a sensing cycle, DSVs and mobile users will collect information in their assigned subareas. After the sensed data are submitted in each time cycle, the sensing matrix F' is acquired. By leveraging matrix completion algorithms, the inferred matrix $\hat{F}$ is obtained. That is to say, the information in unsensed subareas is estimated with acceptable inference errors. With the information of $\hat{F}$, a governor can take action to optimize different smart systems. For instance, when the collected information in a subarea about passenger flow index exceeds the predefined threshold, the local governor will suggest citizens in other regions to not travel to this region and take strict isolation measures in this region during the COVID-19 outbreak.

In our approach, the rank *k* of the matrix is a key parameter. Thus, we implement the matrix factorization with a series of ranks to the ground truth matrix *G* and obtain the inferred matrix $\hat{F}$, then the rank is determined by analyzing the series of $error(G, \hat{F})$, as shown in Fig. 5. The series of *error* shown in the figure reveals the low-rank feature of the matrix *G*: small rank (relative to the dimension of data matrix) leads to small error and the *error* has a manifest drop with the increase of rank. When the *rank*(Flow) $>$ 22 and *rank*(TaxiSpeed) $>$ 26, the *error* is less than 0.05 and the decline rate of error gradually slows down with the increase of rank. Since selecting more subareas to sense would incur more costs and the *error* less than 0.05 reflects a satisfying result, we adopt *rank*(Flow) = 22 and *rank*(TaxiSpeed) = 26 for the two data sets, respectively.

B. Configurations

We conduct extensive Monte Carlo experiments to investigate the performance of our proposed hybrid sensing framework. To eliminate the factors of several initial conditions on the results, such as different initial positions of DSVs and different distribution of users, 50 groups of initial configurations



(a) Flow (b) TaxiSpeed

Fig. 5. The series of error with different rank.

are randomly generated, including the DSVs' initial positions and users' distribution, under each key parameters combination (DSVs number and user coverage). Then, the experiments of different algorithms with the same key parameters employ the same 50 groups of initial configurations. We use 1) the *inference error* to evaluate the overall performance, and use 2) the *optimal DSVs cost rate (ODCR)* to denote the optimal budget allocation. The inference error can be calculated using Eq. (10), while the *ODCR* is defined as follows: Given a total budget B_u and cost of each DSV c_v, the error is minimum when the number of employed DSVs is n_v^*, then *ODCR* = $c_v n_v^*/B_u$.

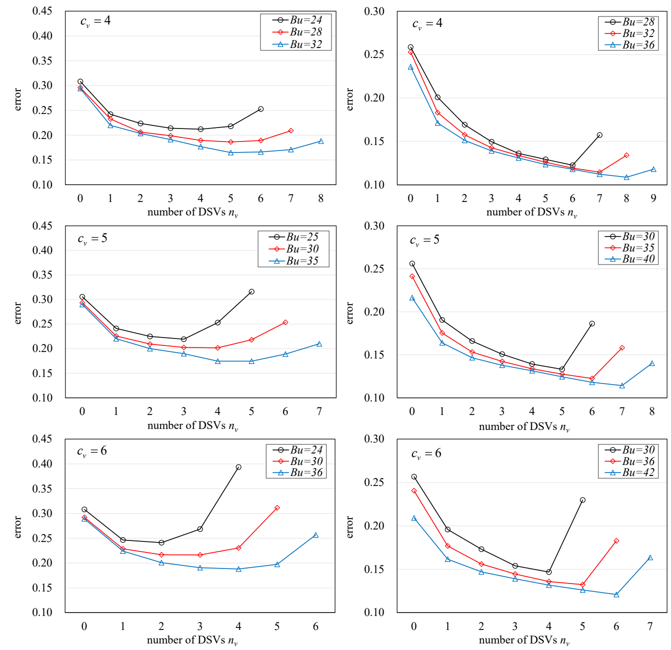
VI. Results and Discussion

We first investigate the effectiveness of the hybrid sensing framework by performing extensive experiments with different numbers of DSVs under different c_v and B_u. Then, we compare our approach with five baselines to verify the effectiveness of the novel strategies (subarea selection, route planning, and user selection) proposed in our framework. Finally, the factors such as fusion order of DSVs, user coverage, and sensing error of users are discussed to better understand the robustness of our framework.

A. *Effectiveness of our hybrid sensing framework*

In our proposed framework, both DSVs and mobile users are employed to complete sensing tasks, while only users or DSVs are employed in the all-user pattern and all-DSV pattern, respectively. The cost of recruiting a mobile user in each sensing cycle is set as $c_u = 1$ (unit). Considering the cost of energy, labor, and communication, the cost of a DSV c_v is determined as 4-6 times that of a user per cycle, that is $c_v = 4$, $c_v = 5$, or $c_v = 6$. According to the active subarea selection strategy, the budget should be satisfied to sample at least k subareas. For example, as the *rank*(Flow) = 22, the budget used on the Flow must satisfy B_u(Flow) ≥ 22. Moreover, the budget is set as multiples of the cost of DSV c_v, so that we can compare the hybrid pattern with the all-user and all-DSV patterns under the same budget. For instance, in experiments on the Flow dataset, when the $c_v = 4$, the budget is set as 24, 28, and 32. The results of six groups of experiments are exhibited in Fig. 6.

In all groups of experiments, the same trend is witnessed: the error at first decreases with the increment of the number of DSVs, which benefits from the suitable budget allocation



(a) error-Flow (b) error-TaxiSpeed

Fig. 6. The performance of different numbers of DSVs with different B_u and c_v.

and the high sensing accuracy of DSVs. However, as the n_v further increases, the error will arise. When all budget is used to employ DSVs (all-DSV pattern), there is a noticeable surge in error, especially on the TaxiSpeed dataset. This is because there are some informative subareas that DSV cannot reach in a sensing cycle (due to the time constraint in a time cycle), so some users are required to sample those subareas. However, the remaining budget for recruiting mobile users cannot afford the essential informative subareas, especially when all budgets are assigned to DSVs. In other words, several informative subareas cannot be sampled in this situation. Besides, under the same c_v and n_v, the enlargement of the budget will cause a decrease in error, which stands to reason. More budget enables the system to recruit more participants and collect more data.

With a larger vehicle cost c_v, the inference error will increase more sharply when employing more DSVs. For example, on the Flow, the error only increases by about 19% when $n_v = 6$, compared with the minimum error in this group ($c_v = 4$, B_u = 24). However, on the Flow, the error increases by about 63% when $n_v = 4$, compared with the minimum error in this group ($c_v = 6$, B_u = 24). It is because when c_v is large, the total number of participants would decrease, and further less informative and robust subareas can be sampled. It reveals that a reasonable budget allocation is essential to achieve optimal inference results.

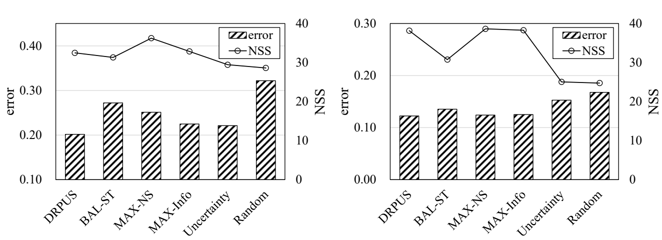
On the Flow, the average *ODCR* of all experiments is 0.6234 while the average *ODCR* is 0.8530 on the TaxiSpeed. These two values represent the optimal budget allocation on the two datasets respectively, providing the reference values for the actual sensing campaign on the two datasets.

B. Effectiveness of the strategies in our framework

Here we evaluate the effectiveness of the novel strategies in our framework, including subarea selection (Section III.B), route planning (Section IV.B), and user selection (Section IV.C). Specifically, we compare our strategy with five baselines as following:

- **Random**, which randomly selects a route from the optional routes set for each DSV and selects users to collect data.
- **Uncertainty**, which preferentially selects the subarea with greater uncertainty that is calculated based on the Query by Committee (QBC) [26] method. We use the data in the last sensing cycle as the pre-recovered data since the data in the upcoming sensing cycle is absent.
- **BLA-ST**, which only preferentially selects the subarea with few historical sampled times in route planning and user selection and tends to balance the historical sampled times of all subareas.
- **MAX-NS**, which only preferentially selects the route with more subareas in the fused route selection process and tends to maximize the number of subareas covered by DSVs.
- **MAX-Info**, which only preferentially selects the route with more informative subareas in the fused route selection process and tends to maximize the number of informative subareas covered by DSVs.

Since the results with different parameter configurations have similar trends, we present the results in Appendix (Table III and IV). Here we only show the results of one of the parameter configurations for the subsequent experiments. The cost of each DSV is selected as $c_v = 5$, and the budgets for the two data sets are determined at B_u(Flow) = 30, B_u(TaxiSpeed) = 35, respectively. According to results in Fig. 6, the number of employed DSVs are set as $n_v = 4$ and $n_v = 6$ for the two data sets. The experiments for our strategy and five baselines are conducted under the same parameter configurations, and the results are drawn in Fig. 7. In addition to the error, we also count the average number of sampled subareas (*NSS*) for each group of experiments.



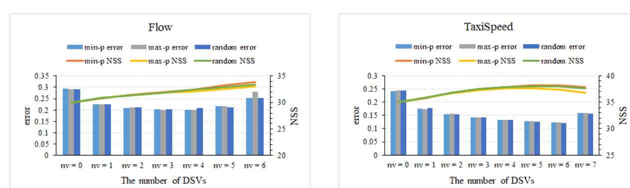
(a) error-Flow (b) error-TaxiSpeed

Fig. 7. The performance of different algorithms.

As shown in the figure, our strategy acquires the minimum error among five baselines on the two datasets, which validates the advantages of our framework. The BAL-ST has the highest error among our strategy and its three variants (BAL-ST, MAX-NS, MAX-Info), especially on the Flow dataset. Besides, the BAL-ST typically has fewer *NSS* than our strategy and the other two variants. The MAX-NS enables DSVs to sample more subareas, which has the highest *NSS* on the two datasets, but the error of MAX-NS is higher than that of our strategy. The MAX-Info abandons the weighted random selection process in the fused route selection, which leads to worse performance. This verifies the effectiveness of the weighted random selection process in obtaining high-quality inference results. The Uncertainty and Random usually have lower *NSS* than other algorithms, and the Random has the worst performance among all algorithms. Uncertainty has a smaller error than MAX-NS and MAX-Info on the Flow, but it is inferior to our strategy and its three variants on the TaxiSpeed.

C. Fusion order of DSVs, sensing error and coverage of users

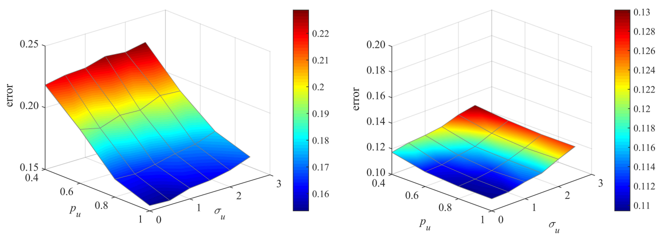
To test the robustness of the proposed hybrid framework, we first examine if the fusion order of DSVs would affect the final results. The cost of each DSV is selected as $c_v = 5$, and the budgets for the two data sets are determined at B_u(Flow) = 30, B_u(TaxiSpeed) = 35, respectively. As we can see in Fig. 8, three different fusion orders are compared, and they are fusing preferentially if the number of optional routes is small (min-p), fusing preferentially if the number of optional routes is large (max-p), and fusing randomly (random). It is concluded that different fusion orders of DSVs have little impact on the results. In general, the average result of min-p is the best, and its performance is relatively stable, so we adopt this strategy in this work. The min-p strategy is less likely to delete a large number of fused routes in the initial stage and finally, it can always cover more subareas (larger *NSS*).



(a) On Flow (b) On TaxiSpeed

Fig. 8. The performance of our framework with different fusion orders of DSVs.

We then investigate the performance of our framework with a different standard deviation of sensing error (σ_u) and coverage (p_u) of mobile users. As shown in Fig. 9, the inference error generally decreases with the increase of p_u and the decrease of σ_u. Greater user coverage p_u indicates that informative subareas can be covered by users with a higher probability. Therefore, when some informative subareas cannot be reached by DSVs in a sensing cycle, users can be recruited for sampling, so the error is reduced. On the two data sets, the error falls gently along with the change of σ_u. However, on the Flow, the error has a rapid drop with the increase of p_u while the decreasing trend of error is slow along with the change of p_u on the TaxiSpeed. This is because the Flow contains more subareas than the TaxiSpeed and the informative subareas that DSVs cannot reach in a sensing cycle will be more on the Flow, thus the sensing campaign on the Flow relies more on users to collect data from different informative subareas that cannot be reached by DSVs. It also explains why the *ODCR* on the Flow is smaller than that on TaxiSpeed. Therefore, the error drops sharply on the Flow when the user coverage increases.



(a) error-Flow (b) error-TaxiSpeed

Fig. 9. The performance of our framework with different σ_u and p_u.

D. Discussion

In this section, we conclude the research findings and discuss some drawbacks of this work.

By blending DSVs and mobile users in the MCS campaigns, the hybrid sensing framework successfully achieves lower inference error compared to the all-user and the all-DSV pattern. It indicates that the framework realizes a good balance of sensing costs and sensing error for different sensing resources and overcomes the problems incurred by only a single sensing resource. Moreover, by comparing with five baselines, our proposed strategies achieve superior performance on the number of sampled subareas and the inference errors. It reveals that more informative and robust subareas are selected by our strategies. Notably, our framework is robust to address the different fusion orders of DSVs, the uncertainty of sensing error as well as the coverage of mobile users. It indicates that the framework can complete the real-world Sparse MCS tasks.

However, there remain some drawbacks in the present work. The main lies in the hybrid pattern. We use an offline way, which assigns the tasks to DSVs and the mobile users in a defined order. This approach fails to fully exploit the potential of a hybrid sensing pattern in further reducing sensing costs and inference errors. The current DSVs route planning algorithm is based on an improved version of the greedy algorithm, which keeps the current best k candidates for subselection. The division of the spatiotemporal sensing map heavily affects the efficiency of our algorithm. In terms of task assignment, the current approach is an offline mode, in which DSVs and mobile users are arranged before a new cycle starts. However, considering a more practical scenario, mobile users can participate in or quit the sensing tasks at any time. Thus, an offline-online hybrid mode for different sensing resources (offline for DSVs, and online for mobile users) is more suitable. Moreover, our framework does not take into account changes in the number of available DSVs and mobile users over time. A dynamic mechanism would better be designed to adjust the recruitment of different sensing resources during different time cycles.

VII. Conclusion

Based on the development of social transportation and social sensing, a crowd-aided vehicular hybrid sensing framework is proposed, namely DRPUS. To the best of our knowledge, the framework is the first attempt to blend numerous devices and individuals in a unified target, which assigns sensing tasks to DSVs and mobile users effectively and achieves improved inference results by conquering three challenges. Firstly, we determine the order of task assignment based on the characteristics of DSVs and mobile users (mobility, sensing accuracy, etc.). Also, we define the informative subarea and robust subarea based on the analysis of data inference for subarea selection. Moreover, we devise a three-step strategy to determine the route for each DSV, including optimal route searching, fused route selection, and final route determination. Under the rest budget constraint, we propose an active user selection strategy to recruit appropriate users to complement the data. The effectiveness of our framework is validated by extensive experiments on two real-world datasets by comparing it with the user-only/DSV-only framework, and five baselines. Research findings reveal that our hybrid sensing framework enhances the data collection at a low-cost manner for ITS.

In the future, we would like to improve our hybrid sensing framework from the following aspects. Firstly, a more universal hybrid sensing framework should be devised, in which no fixed order of task assignment for DSVs and mobile users is required. Secondly, in the decision on DSVs routes, we would like to try evolutionary algorithms to provide better solutions. Lastly, we will consider more practical factors in the framework, such as the diversity of DSV's cost and user mobility prediction.

VIII. Appendix

Two tables in Section VI of Results are listed here.

TABLE III
THE RESULTS OF OUR STRATEGY AND FIVE BASELINES ON FLOW

	Flow							
	Indicators	$n_v = 0$	$n_v = 1$	$n_v = 2$	$n_v = 3$	$n_v = 4$	$n_v = 5$	$n_v = 6$
Our strategy	*error*	0.294	0.226	0.209	0.203	0.202	0.218	0.254
	NSS	30	30.842	31.462	31.991	32.476	33.277	33.867
BLA-ST	*error*	0.372	0.291	0.284	0.273	0.272	0.284	0.355
	NSS	30	30.736	31.036	31.380	31.308	30.696	29.474
MAX-NS	*error*	0.294	0.284	0.272	0.260	0.251	0.301	0.404
	NSS	30	31.371	33.138	34.723	36.237	37.498	38.502
MAX-Info	*error*	0.293	0.268	0.251	0.237	0.225	0.233	0.285
	NSS	30	30.703	31.413	32.097	32.857	33.560	34.099
Uncertainty	*error*	0.308	0.227	0.219	0.218	0.221	0.231	0.368
	NSS	30	30.634	30.732	30.343	29.439	28.151	26.663
Random	*error*	0.354	0.321	0.309	0.311	0.322	0.349	0.451
	NSS	30	31.060	31.080	30.225	28.626	26.937	24.475

References

[1] F.-Y. Wang, K. M. Carley, D. Zeng, and W. Mao, "Social computing: From social informatics to social intelligence," *IEEE Intell Syst*, vol. 22, no. 2, pp. 79–83, Mar. 2007, doi: 10.1109/MIS.2007.41.

[2] F. Zhu, Y. Lv, Y. Chen, X. Wang, G. Xiong, and F.-Y. Wang, "Parallel transportation systems: Toward iot-enabled smart urban traffic control and management," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 10, pp. 4063–4071, Aug. 2019, doi: 10.1109/TITS.2019.2934991.

[3] J. Cao, K. Zeng, H. Wang, J. Cheng, F. Qiao, D. Wen, and Y. Gao, "Web-based traffic sentiment analysis: Methods and applications," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 2, pp. 844–853, Dec. 2013, doi: 10.1109/TITS.2013.2291241.

TABLE IV
THE RESULTS OF OUR STRATEGY AND FIVE BASELINES ON TAXISPEED

TaxiSpeed									
	Indicators	$n_v = 0$	$n_v = 1$	$n_v = 2$	$n_v = 3$	$n_v = 4$	$n_v = 5$	$n_v = 6$	$n_v = 7$
Our strategy	*error*	0.241	0.175	0.153	0.142	0.134	0.128	0.123	0.158
	NSS	35	35.714	36.837	37.472	37.913	38.186	38.169	37.859
BLA-ST	*error*	0.257	0.176	0.159	0.151	0.144	0.139	0.135	0.150
	NSS	35	35.152	35.051	34.489	33.692	32.369	30.759	28.644
MAX-NS	*error*	0.244	0.201	0.159	0.145	0.135	0.128	0.124	0.137
	NSS	35	35.906	37.185	37.790	38.343	38.661	38.633	38.554
MAX-Info	*error*	0.244	0.203	0.172	0.154	0.143	0.131	0.125	0.184
	NSS	35	35.852	36.827	37.287	37.719	38.258	38.275	37.882
Uncertainty	*error*	0.257	0.180	0.162	0.153	0.149	0.149	0.153	0.188
	NSS	35	35.241	34.568	33.211	31.111	28.689	25.054	21.873
Random	*error*	0.258	0.201	0.174	0.167	0.165	0.162	0.168	0.243
	NSS	35	35.222	34.569	32.840	30.481	27.8	24.755	21.412

[4] F.-Y. Wang, "Scanning the issue and beyond: Crowdsourcing for field transportation studies and services," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 1, pp. 1–8, Feb. 2015. doi: 10.1109/TITS.2015.2393752.

[5] X. Zheng, W. Chen, P. Wang, D. Shen, S. Chen, X. Wang, Q. Zhang, and L. Yang, "Big data for social transportation," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 3, pp. 620–630, Dec. 2015, doi: 10.1109/TITS.2015.2480157.

[6] C. Jiang, L. Gao, L. Duan, and J. Huang, "Scalable mobile crowdsensing via peer-to-peer data sharing," *IEEE Trans. Mobile Comput*, vol. 17, no. 4, pp. 898–912, Apr. 2017, doi: 10.1109/TMC.2017.2743718.

[7] R. Chattopadhyay and C.-K. Tham, "Joint sensing and processing resource allocation in vehicular ad-hoc networks," *IEEE Trans. Intell. Vehicl.*, p. Early Access, Nov. 2021, doi: 10.1109/TIV.2021.3124208.

[8] X. Sheng, J. Tang, and W. Zhang, "Energy-efficient collaborative sensing with mobile phones," in *Proc. IEEE INFOCOM*, Orlando, FL, United states, 2012, pp. 1916–1924.

[9] H. Xiong, D. Zhang, L. Wang, and H. Chaouchi, "Emc 3: Energy-efficient data transfer in mobile crowdsensing under full coverage constraint," *IEEE Trans. Mobile Comput*, vol. 14, no. 7, pp. 1355–1368, Sep. 2014, doi: 10.1109/TMC.2014.2357791.

[10] S. Hachem, A. Pathak, and V. Issarny, "Probabilistic registration for large-scale mobile participatory sensing," in *IEEE Int. Conf. Pervasive Comput. Commun.*, San Diego, CA, United states, 2013, pp. 132–140.

[11] D. Zhang, H. Xiong, L. Wang, and G. Chen, "Crowdrecruiter: Selecting participants for piggyback crowdsensing under probabilistic coverage constraint," in *Proc. ACM Int. Jt. Conf. Pervasive Ubiquitous Comput.*, Seattle, United states, 2014, pp. 703–714.

[12] M. Zhang, P. Yang, C. Tian, S. Tang, X. Gao, B. Wang, and F. Xiao, "Quality-aware sensing coverage in budget-constrained mobile crowdsensing networks," *IEEE Trans. Veh Technol*, vol. 65, no. 9, pp. 7698–7707, Oct. 2015, doi: 10.1109/TVT.2015.2490679.

[13] F. Yucel, M. Yuksel, and E. Bulut, "Coverage-aware stable task assignment in opportunistic mobile crowdsensing," *IEEE Trans. Veh Technol*, vol. 70, no. 4, pp. 3831–3845, Mar. 2021, doi: 10.1109/TVT.2021.3065688.

[14] W. Liu, L. Wang, E. Wang, Y. Yang, D. Zeghlache, and D. Zhang, "Reinforcement learning-based cell selection in sparse mobile crowdsensing," *Comput Netw*, vol. 161, pp. 102–114, Jun. 2019, doi: 10.1016/j.comnet.2019.06.010.

[15] Z. Zhu, B. Chen, W. Liu, Y. Zhao, Z. Liu, and Z. Zhao, "A cost-quality beneficial cell selection approach for sparse mobile crowdsensing with diverse sensing costs," *IEEE Internet Things*, vol. 8, no. 5, pp. 3831–3850, Sep. 2020, doi: 10.1109/JIOT.2020.3024833.

[16] X. Wei, Z. Li, Y. Liu, S. Gao, and H. Yue, "Sdlsc-ta: subarea division learning based task allocation in sparse mobile crowdsensing," *IEEE Trans. Emerg Top Com*, vol. 9, no. 3, pp. 1344–1358, Dec. 2020, doi: 10.1109/TETC.2020.3045463.

[17] P. Sun, Z. Wang, L. Wu, H. Shao, H. Qi, and Z. Wang, "Trustworthy and cost-effective cell selection for sparse mobile crowdsensing systems," *IEEE Trans. Veh Technol*, vol. 70, no. 6, pp. 6108–6121, May. 2021, doi: 10.1109/TVT.2021.3077112.

[18] K. Xie, X. Li, X. Wang, G. Xie, J. Wen, and D. Zhang, "Active sparse mobile crowd sensing based on matrix completion," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, Amsterdam, Netherlands, 2019, pp. 195–210.

[19] L. Wang, D. Zhang, A. Pathak, C. Chen, H. Xiong, D. Yang, and Y. Wang, "Ccs-ta: Quality-guaranteed online task allocation in compressive crowdsensing," in *Proc. ACM Int. Jt. Conf. Pervasive Ubiquitous Comput.*, Osaka, Japan, 2015, pp. 683–694.

[20] W. Liu, Y. Yang, E. Wang, and J. Wu, "User recruitment for enhancing data inference accuracy in sparse mobile crowdsensing," *IEEE Internet Things*, vol. 7, no. 3, pp. 1802–1814, Dec. 2019, doi: 10.1109/JIOT.2019.2957399.

[21] C. Zhang, L. Zhu, C. Xu, and K. Sharif, "Prvb: Achieving privacy-preserving and reliable vehicular crowdsensing via blockchain oracle," *IEEE Trans. Veh Technol*, vol. 70, no. 1, pp. 831–843, Dec. 2020, doi: 10.1109/TVT.2020.3046027.

[22] X. Wang, W. Wu, and D. Qi, "Mobility-aware participant recruitment for vehicle-based mobile crowdsensing," *IEEE Trans. Veh Technol*, vol. 67, no. 5, pp. 4415–4426, Dec. 2017, doi: 10.1109/TVT.2017.2787750.

[23] W. Gong, B. Zhang, and C. Li, "Location-based online task assignment and path planning for mobile crowdsensing," *IEEE Trans. Veh Technol*, vol. 68, no. 2, pp. 1772–1783, Nov. 2018, doi: 10.1109/TVT.2018.2884318.

[24] G. Fan, Y. Zhao, Z. Guo, H. Jin, X. Gan, and X. Wang, "Towards fine-grained spatio-temporal coverage for vehicular urban sensing systems," in *Proc. IEEE INFOCOM*, Vancouver, BC, Canada, May 10-13, 2021.

[25] W. Liu, Y. Yang, E. Wang, and J. Wu, "Dynamic user recruitment with truthful pricing for mobile crowdsensing," in *Proc. IEEE INFOCOM*, Toronto, ON, Canada, 2020, pp. 1113–1122.

[26] S. Chakraborty, J. Zhou, V. Balasubramanian, S. Panchanathan, I. Davidson, and J. Ye, "Active matrix completion," in *Proc. IEEE Int. Conf. Data Min.*, Dallas, TX, USA, 2013, pp. 81–90.

[27] D. Zhu and S. X. Yang, "Bio-inspired neural network-based optimal path planning for uuvs under the effect of ocean currents," *IEEE Trans. Intell. Vehicl.*, vol. 7, no. 2, pp. 231–239, May. 2021, doi: 10.1109/TIV.2021.3082151.

[28] H. Li, W. Liu, C. Yang, W. Wang, T. Qie, and C. Xiang, "An optimization-based path planning approach for autonomous vehicles using dynefwa-artificial potential field," *IEEE Trans. Intell. Vehicl.*, vol. 7, no. 2, pp. 263–272, Oct. 2021, doi: 10.1109/TIV.2021.3123341.

[29] J. Shang, Y. Zheng, W. Tong, E. Chang, and Y. Yu, "Inferring gas consumption and pollution emission of vehicles throughout a city," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, New York, NY, United states, 2014, pp. 1027–1036.

[30] D. Cao, X. Wang, L. Li, C. Lv, X. Na, Y. Xing, X. Li, Y. Li, Y. Chen, and F.-Y. Wang, "Future directions of intelligent vehicles: Potentials, possibilities, and perspectives," *IEEE Trans. Intell. Vehicl.*, vol. 7, no. 1, pp. 7–10, Mar. 2022, doi: 10.1109/TIV.2022.3157049.

[31] X. Wang, X. Zheng, Q. Zhang, T. Wang, and D. Shen, "Crowdsourcing in its: The state of the work and the networking," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 6, pp. 1596–1605, Feb. 2016, doi: 10.1109/TITS.2015.2513086.

[32] Y. Liu, X. Liu, S. Gao, L. Gong, C. Kang, Y. Zhi, G. Chi, and L. Shi, "Social sensing: A new approach to understanding our socioeconomic environments," *Ann Assoc Am Geogr*, vol. 105, no. 3, pp. 512–530, Apr. 2015, doi: 10.1080/00045608.2015.1018773.

[33] L. Kong, M. Xia, X.-Y. Liu, M.-Y. Wu, and X. Liu, "Data loss and reconstruction in sensor networks," in *2013 Proceedings IEEE INFOCOM*. IEEE, 2013, pp. 1654–1662.

[34] L. Wang, D. Zhang, Y. Wang, C. Chen, X. Han, and A. M'hamed, "Sparse mobile crowdsensing: challenges and opportunities," *IEEE Commun Mag*, vol. 54, no. 7, pp. 161–167, Jul. 2016, doi: 10.1109/MCOM.2016.7509395.

[35] C. Wang, Z. Zhang, L. Shao, and M. Zhou, "Estimating travel speed via sparse vehicular crowdsensing data," in *IEEE World Forum Internet Things*, Reston, VA, United states, 2016, pp. 643–648.

[36] C. Wang, Z. Xie, L. Shao, Z. Zhang, and M. Zhou, "Estimating travel speed of a road section through sparse crowdsensing data," *IEEE Trans. Intell Transp*, vol. 20, no. 9, pp. 3486–3495, Dec. 2018, doi: 10.1109/TITS.2018.2877059.

[37] L. Wang, D. Zhang, D. Yang, A. Pathak, C. Chen, X. Han, H. Xiong, and Y. Wang, "Space-ta: Cost-effective task allocation exploiting intradata and interdata correlations in sparse crowdsensing," *ACM Trans. Intel Syst Tec*, vol. 9, no. 2, pp. 1–28, Oct. 2017, doi: 10.1145/3131671.

[38] L. Wang, W. Liu, D. Zhang, Y. Wang, E. Wang, and Y. Yang, "Cell selection with deep reinforcement learning in sparse mobile crowdsensing," in *Proc Int Conf Distrib Comput Syst*, Vienna, Austria, 2018, pp. 1543–1546.

[39] E. Wang, M. Zhang, X. Cheng, Y. Yang, W. Liu, H. Yu, L. Wang, and J. Zhang, "Deep learning-enabled sparse industrial crowdsensing and prediction," *IEEE Trans. Ind Inform*, vol. 17, no. 9, pp. 6170–6181, Oct. 2021, doi: 10.1109/TII.2020.3028616.

[40] J. Wang, F. Wang, Y. Wang, L. Wang, Z. Qiu, D. Zhang, B. Guo, and Q. Lv, "Hytasker: Hybrid task allocation in mobile crowd sensing," *IEEE Trans. Mobile Comput*, vol. 19, no. 3, pp. 598–611, Feb. 2019, doi: 10.1109/TMC.2019.2898950.

[41] G. Han, L. Liu, S. Chan, R. Yu, and Y. Yang, "Hysense: A hybrid mobile crowdsensing framework for sensing opportunities compensation under dynamic coverage constraint," *IEEE Commun Mag*, vol. 55, no. 3, pp. 93–99, Mar. 2017, doi: 10.1109/MCOM.2017.1600658CM.

[42] X. Wang, P. Li, and J. Xiao, "Budget constraint task allocation for mobile crowd sensing with hybrid participant," in *CollaborateCom*, Shanghai, China, 2020, pp. 506–517.

[43] X. Wei, Z. Li, C. Ren, T. Guo, and S. Gao, "Hsm-smcs: Task assignment based on hybrid sensing modes in sparse mobile crowdsensing," *IEEE Internet Things*, early access, doi: 10.1109/JIOT.2022.3150804.

[44] P. Zhao, X. Li, S. Gao, and X. Wei, "Cooperative task assignment in spatial crowdsourcing via multi-agent deep reinforcement learning," *J Syst Architect*, vol. 128, p. 102551, Jul. 2022, doi: 10.1016/j.sysarc.2022.102551.

[45] K. P. O'Keeffe, A. Anjomshoaa, S. H. Strogatz, P. Santi, and C. Ratti, "Quantifying the sensing power of vehicle fleets," *P Natl Acad Sci*, vol. 116, no. 26, pp. 12 752–12 757, Jun. 2019, doi: 10.1073/pnas.1821667116.

[46] R. Ding, Z. Yang, Y. Wei, H. Jin, and X. Wang, "Multi-agent reinforcement learning for urban crowd sensing with for-hire vehicles," in *Proc. IEEE INFOCOM*, Vancouver, BC, Canada, May 10-13, 2021.

[47] Y. Zhu, Z. Li, H. Zhu, M. Li, and Q. Zhang, "A compressive sensing approach to urban traffic estimation with probe vehicles," *IEEE Trans. Mobile Comput*, vol. 12, no. 11, pp. 2289–2302, Oct. 2012, doi: 10.1109/TMC.2012.205.

[48] A. G. Akritas and G. I. Malaschonok, "Applications of singular-value decomposition (svd)," *Math Comput Simulat*, vol. 67, no. 1-2, pp. 15–31, Jul. 2004, doi: 10.1016/j.matcom.2004.05.005.

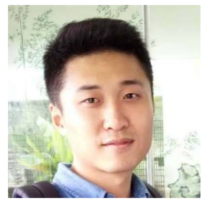
[49] A. De Juan, Y. Vander Heyden, R. Tauler, and D. Massart, "Assessment of new constraints applied to the alternating least squares method," *Anal Chim Acta*, vol. 346, no. 3, pp. 307–318, Dec. 1997, doi: 10.1016/S0003-2670(97)90069-6.

[50] H. D. Vinod, "A survey of ridge regression and related techniques for improvements over ordinary least squares," *REV ECON STAT*, vol. 60, no. 1, pp. 121–131, Feb. 1978, doi: 10.2307/1924340.

Zhengqiu Zhu received the B.E. and M.E. degree in Simulation Engineering from National University of Defense Technology, Changsha, China, in 2016 and 2018, where he is currently pursuing the Ph.D. degree in the College of Systems Engineering. He was also a visiting Ph.D. student in the research group of Multi-scale Networked Systems at University of Amsterdam (UvA). His research interests include mobile crowdsensing, ubiquitous computing, and social computing.

Yong Zhao received the B.E. degree in Simulation from National University of Defense Technology, China, in 2019, where he is currently pursuing the master degree with Control Science and Engineering. His research interests include Reinforcement Learning, autonomous searching and crowd sensing. He is now working on localization the odor source using mobile sensors.

Bin Chen the corresponding author, received his B.E., M.E. and Ph.D. degree in Control Science and Engineering from National University of Defense Technology (NUDT) in 2003, 2005 and 2010. He is currently an associate research fellow in the College of Systems Engineering at National University of Defense Technology. His current research focuses on ACP-based simulation running support, parallel experimental method, data mining, mobile crowdsensing and micro-task crowdsourcing.



Sihang Qiu received his Ph.D. from the Web Information Systems group of TU Delft in 2021 and is an assistant professor at Hunan Institute of Advanced Technology and College of Systems Engineering of National University of Defense Technology. He has published many papers on CHI, CSCW, and WWW. He is interested in developing novel human-AI interaction techniques. His research also focuses on human-centered AI, crowd computing, and conversational agents.

Zhong Liu obtained his Ph.D. in System Engineering and Mathematics from National University of Defense Technology and and also a professor of Science and Technology on Information Systems Engineering Laboratory. He is also the member of the National Artificial Intelligence Strategy Advisory Committee. His main research interests include artificial general intelligence, deep reinforcement learning, and multi-agent systems.

Kun Xie is currently a Professor of College of Computer Science and Electronic Engineering at Hunan University, Changsha, China. She is also the Associate Dean of the Department of Computer Engineering and the winner of National Science Fund for Outstanding Young Scholars. Her research interests include Computer Networks, Network Security, Big Data, and AI.

Liang Ma received his B.E., M.E. and Ph.D. degree in Control Science and Engineering from National University of Defense Technology(NUDT) in 2013, 2015, 2019, respectively. He is currently a lecturer in the College of Advanced Interdisciplinary Studies at National University of Defense Technology. His current research focuses on ACP-based massive pedestrian crowds management and simulation big data analysis.