IEEE Open Journal of the
Computer Society

Received 26 April 2022; accepted 19 May 2022. Date of publication 23 May 2022;
date of current version 23 June 2022. The review of this paper was arranged by Associate Editor Deze Zeng.

Digital Object Identifier 10.1109/OJCS.2022.3177290

A Survey of Sparse Mobile Crowdsensing: Developments and Opportunities

SHITING ZHAO[1], GUOZI QI[1], TENGJIAO HE [1], JINPENG CHEN [2], ZHIQUAN LIU [1], AND KAIMIN WEI [1]

[1]College of Information Science and Technology and National & Local Joint Engineering Research Center of Network Security Detection and Protection Technology, Jinan University, Guangzhou 510632, China
[2]School of computer science, Beijing University of Posts and Telecommunications, Beijing 100876, China

CORRESPONDING AUTHOR: KAIMIN WEI (e-mail: cswei@jnu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61972178, and in part by Guangdong Key R&D Plan under Grants 2019B010137005 and 2019A1515011753.

ABSTRACT Sparse mobile crowdsensing (SMCS) has emerged as a promising sensing paradigm for urban sensing, leveraging the spatial and temporal correlation among data sensed in distinct sub-areas to cut sensing expenses dramatically. It intelligently selects only a tiny portion of the target regions for sensing and accurately infers the data for the remaining unsensed areas. SMCS confronts numerous challenges, such as sensing cell selection and missing data inference, when compared to mobile crowdsensing. Researchers in recent years have proposed plenty of strategies to solve these challenges. From the perspective of comparing MCS, we aim to provide a comprehensive literature review of recent advances in SMCS in this paper. We begin by going over the preliminary of SMCS and MCS, including their evolution, characteristics, and life-cycle stages. We then go through their common key techniques and recent developments. Furthermore, we give a review of the unique key techniques as well as the most recent advancements. We finally identify existing applications and highlight potential research opportunities for SMCS. Our objective is to provide researchers with a comprehensive understanding of SMCS.

INDEX TERMS Sparse mobile crowdsensing, participant recruitment, privacy protection, incentive mechanism.

I. INTRODUCTION

Thanks to the popularity of mobile devices equipped with various sensors, including accelerometers, gyroscopes, GPS, cameras, microphones, etc., mobile crowdsensing (MCS) has emerged as a promising data collection paradigm in smart cities [1], [2]. It mainly leverages the strength of collaboration among regular users, as well as the sensing, computing, and wireless communication capabilities afforded by mobile devices, to complete large-scale tasks. Compared to traditional wireless sensor networks (WSNs) that rely on expensive specialized sensing infrastructure, MCS is fast deployment and has low maintenance costs. Consequently, MCS effectively expands the sensing scope from a single place to community and city-wide levels, such as air quality monitoring [3], noise level sensing [4], urban traffic conditions [5], indoor location [6], disaster relief [7], and so forth.

In MCS, data quality and sensing cost are two primary concerns that compete with one another [2], [8]. A straightforward way to enhance the quality of data is to employ a large number of participants, which guarantees sufficient and diverse sensed data. However, it comes with higher sensing costs, including worker cost, device cost, bandwidth cost, computing cost, and storage cost. Many academics have endeavored to minimize sensing costs while maintaining a certain level of data quality in data collection. Liu *et al.* [9] aimed to minimize the travel cost while maximizing the number of completed MCS tasks. Xiong *et al.* [10] minimized the number of redundant task assignments while ensuring the required number of participants to return the sensing results within each time slot.

In most MCS applications, the scale of sensing tasks is rather large, the task duration is generally long, and the coverage of the target scene is relatively wide. To this end, some

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

researchers believe that data collected at various places and time has a specific relationship, referred to as the *Spatio-temporal correlation*. They utilized the Spatio-temporal correlation between data from distinct sub-areas to pick just a small portion of the areas for sensing and then infer the data for unsensed regions, thereby reducing the overall sensing cost and guaranteeing the quality of the sensed data. Wang *et al.* [11] proposed a novel sensing framework, called *sparse mobile crowdsensing (SMCS)*, which exploits the spatial and temporal correlations between data to reduce the number of sensing tasks. SMCS includes three stages: sensing area selection, missing data inference, and data quality assessment. Since its emergence, SMCS has received extensive attention from both academia and industry.

Compared to MCS, SMCS confronts several challenges, e.g., *how to select suitable sensing areas from the target regions to sense data? how to effectively infer data for unsensed areas from collected data remains a mystery*. In recent years, a considerable amount of work on SMCS has been developed, such as sensing area selection in SMCS [12], [13], participant recruitment [14], and data reasoning [15]. To the best of our knowledge, no comprehensive investigation of SMCS has been conducted. In addition, although there are various tutorials or surveys for MCS [2], [16], [17], they concentrate on diverse areas and research topics in this field, giving us a broad picture of MCS. Capponi *et al.* [16] presented a comprehensive investigation and categorization of MCS. Wang *et al.* [2] focused on task assignment in MCS systems. Therefore, a tutorial or survey dedicated to summarizing up-to-date research results in SMCS would be extremely useful and beneficial for better understanding SMCS. In this paper, we provide a comprehensive literature overview of recent advances in SMCS. The significant contributions to this paper can be summarized as follows:

- clarifying the relationship and characteristics and lifecycle stages of MCS and SMCS, to provide an overall picture of SMCS.
- discussing common key techniques between MCS and SMCS and reviewing their recent developments, to help researchers get a better grasp of critical concerns in SMCS.
- analysing the unique key techniques of SMCS and examining the recent works, to better understand the trends and important factors of SMCS.
- investigating various applications and prospective research directions, to provide guidance on its future research and application in SMCS.

The remainder of this paper is organized as follows. In Section II, we go over the preliminaries. Section III summarizes and discusses common key techniques of MCS and SMCS, as well as their recent developments. Section IV analyzes the unique key techniques of SMCS and the related algorithms and technologies. In Section V, we identify various novel applications enabled by SMCS and outline potential research opportunities for SMCS. Finally, we conclude this paper in Section VI.

II. PRELIMINARY

A. EVOLUTION OF SENSING PARADIGM

Wireless Sensor Networks & Mobile Crowdsensing. With the fast expansion of cities, people's demand for refined urban management has increased, while such urban management heavily relies on the state data collected via urban sensing. As an example of urban sensing [18], WSNs leverage a large number of customized sensors and deploy them in an ad-hoc manner in the target region to monitor the system, physical, or environmental conditions. However, traditional WSNs are inherently resource-constrained; they have restricted processing speed, storage capacity, and communication bandwidth [19], limiting the application scope. Moreover, they have higher sensing costs, including device cost, deployment cost, maintenance cost, etc. Later, with the popularity of mobile devices and the widespread use of crowdsourcing, MCS emerges as a low-cost method for urban sensing. On the one hand, it makes use of a wide range of existing mobile devices, such as mobile phones, wearable devices, unmanned aerial vehicles, and so on. It significantly reduces the device cost. On the other hand, it utilizes the cooperation of a large number of ordinary users to execute large-scale sensing tasks. It implies that MCS can be rapidly deployed at a lower cost and more flexible manner. Therefore, MCS has a promising future in urban sensing.

Crowdsourcing & Mobile Crowdsensing. The term "crowdsourcing" [20] was used by Jeff Howe in an essay to describe how businesses use the Internet to outsource work to the masses, using the power of groups to collaborate on a complex task. The core concept behind "crowdsourcing" is to leverage the collective intelligence of a crowd to perform a challenging activity collaboratively, with each person doing much smaller micro-tasks. "Mobile crowdsensing," initially termed by Ganti *et al.* [1], is a subclass of crowdsourcing enabled by the proliferation of mobile devices and the rising sensing needs in the city. The term "crowdsourcing" is defined from a business standpoint, while "mobile crowdsensing" is defined from the perspective of mobile apps. Crowdsourcing has been widely used in many fields, such as knowledge graph refinement, data mining, software engineering, urban sensing, etc. MCS is mostly employed in urban sensing contexts as a specific instance of crowdsourcing. The relationship between crowdsourcing and mobile crowdsensing is depicted in Fig. 1.

Mobile Crowdsensing & Sparse Mobile Crowdsensing. MCS is made up of multiple task requesters, a large number of participants, and an MCS platform, as illustrated in Fig. 3. Task requesters publish perception tasks to the MCS platform in the hopes of minimizing task sensing costs while maximizing data quality. Task participants receive tasks from the MCS platform and wish to maximize the reward (i.e., sensing cost) for data collection at a lower cost, diminishing data quality. Therefore, data quality and sensing cost have certain inherent tensions in MCS. To reduce sensing costs, some researchers merely sense a portion of the target regions and infer the data for the remaining unsensed areas by using the Spatio-temporal correlation between distinct regions, as shown in Fig. 2. The "sparse mobile crowdsensing" paradigm (SMCS), a subgroup

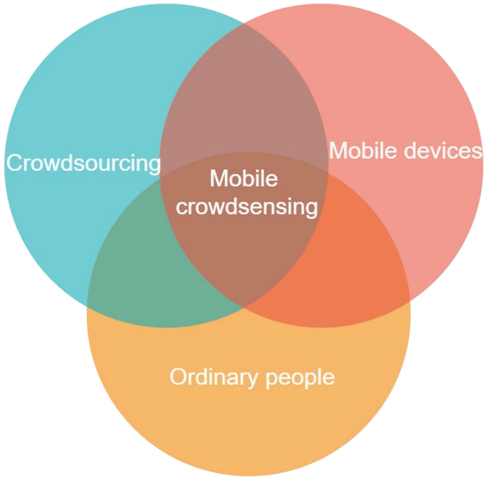

FIGURE 1. The relationship between crowdsourcing and mobile crowdsensing.

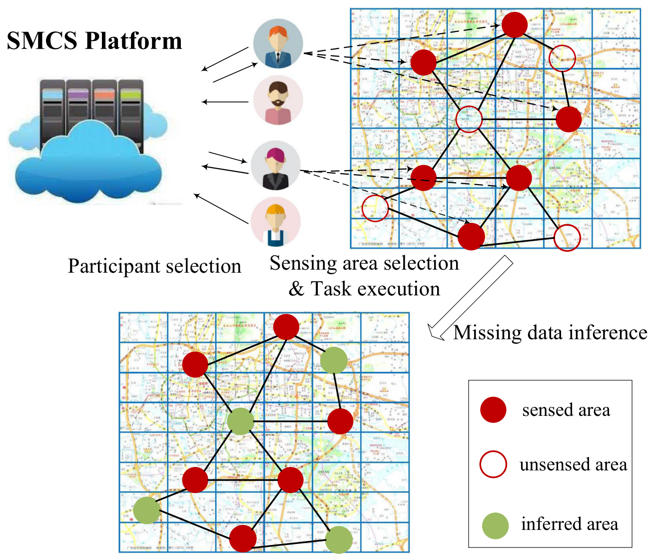

FIGURE 2. The system flow chart of SMCS.

of MCS, was explicitly described by Wang *et al.* [11]. They developed an SMCS framework with three processes: sensing area selection, missing data inference, and quality evaluation. “Sparse” may be interpreted here: (i) collecting data from just a few sub-regions, and (ii) choosing only a few individuals to gather data from rather than all of them [21].

In the field of data sensing, some terms are close to crowdsensing, such as crowd computing, participatory sensing, social computing, crowdsourcing, citizen sensing, opportunistic sensing, and so on. They gather a large number of people and leverage the collective intelligence of a crowd to perform a challenging activity collaboratively, with each person doing much smaller micro-tasks. Compared with crowdsensing, crowd computing covers a broader range, and crowdsourcing refers to task distribution mechanisms. In principle, crowdsensing is similar to other sensing schemes, although they focus on different aspects.

B. CHARACTERISTICS OF MCS AND SMCS

MCS and SMCS both leverage a certain number of mobile devices carried by participants to obtain data about urban life and human activities. Through analyzing their sensing frameworks, MCS and SMCS share **several common characteristics**, summarized as follows:

- **Wide sensing range**: The most prominent feature of MCS and SMCS is the extensive sensing coverage. As the primary sensing paradigms of urban sensing, MCS and SMCS can be used in large-scale perception tasks that generally cover a broad area.
- **Mobility and unconsciousness**: One of the most important requirements for MCS and SMCS is participant mobility, which is used to deploy smart terminals into target scenarios and subsequently obtain data from the surrounding environment. After the mobile devices are placed in the target situations, the target data is automatically sensed and stored by their built-in sensors. In other words, sensors gather data without the need for human involvement, and all of these actions take place in the unconscious state of users [22].
- **Unreliability of sensing data**: The acquired data in MCS and SMCS is collected from users’ mobile devices. These mobile terminals have diverse sensing capacities. The precision of their built-in sensors, for example, is significantly various. Meanwhile, sensors collect some noise data that is hard to identify. Furthermore, some malicious users may upload fake data to disrupt MCS systems. All of the factors mentioned above cause the collected data to be unreliable.

The unique features of SMCS: SMCS exploits the spatial and temporal correlations between the sensing data of different sub-areas and infers the data for other unperceived areas from the collected data of sensed regions. Its distinguishing features are as follows:

- **Sparse**: The “sparse” characteristic in SMCS has two interpretations. On the one hand, this suggests that just a tiny portion of areas will be detected. On the other side, it indicates that just a part of the participants will be chosen rather than all.
- **Low sensing cost**: As previously stated, just a few participants are chosen, and only a few target locations are sensed, resulting in a lower sensing cost.
- **Spatio-temporal**: It refers to some similarity in data acquired at different locations or at various time. For example, in environmental monitoring, the PM 2.5 value of an area is closer to that of the surrounding areas, or the PM 2.5 taken at a particular time is also closer to that collected soon in the same spot.
- **Delay**: In SMCS, missing data inference for unsensed regions requires an amount of time, which is dependent on

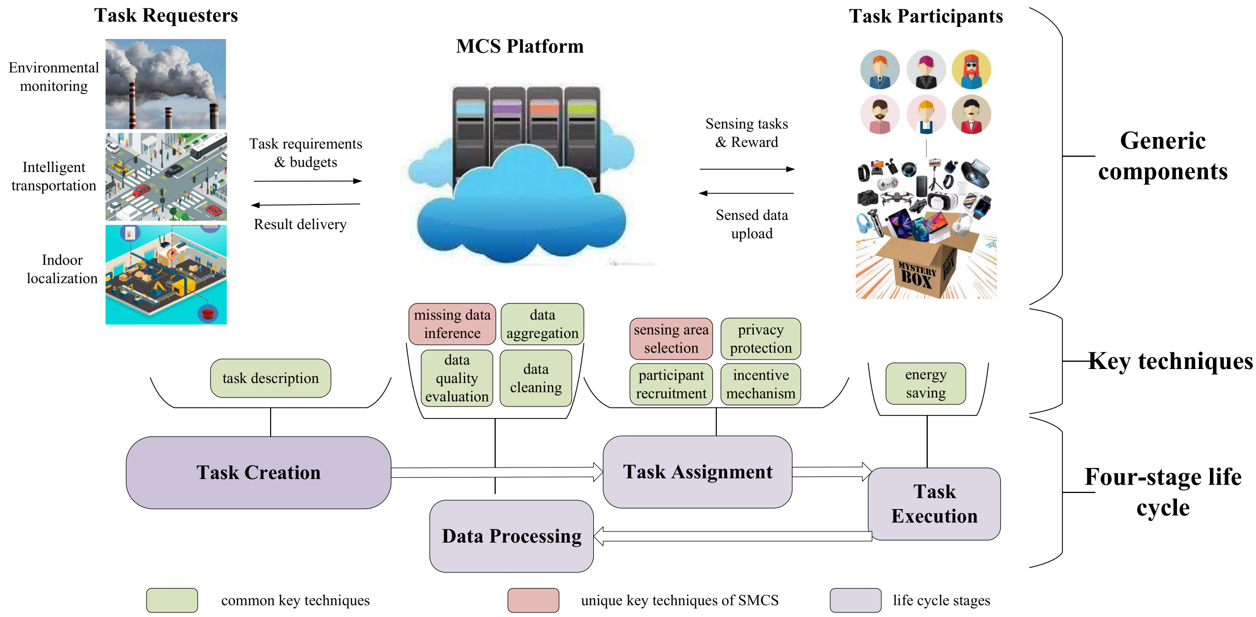

FIGURE 3. The overall framework of MCS and SMCS.

TABLE 1. A Comparison of MCS and SMCS

Paradigm	Number of Participants	Coverage of Sensing Areas	Sensing Cost	Spatiotemporal Correlation	Data Sources
MCS	Massive	High	High	Less Consideration	Sensing Data
SMCS	Modicum	Low	Low	Strong	Sensing Data and Inferred Data

the performance of the missing data inference method. As a result, SMCS cannot be employed in time-critical situations.

- **Uncertainty**: In SMCS, the data of unsensed areas is primarily inferred from the collected data of the selected perception regions. However, this process has considerable ambiguity, which results in inferred data being uncertain. Furthermore, we cannot yet determine whether the inferred data and the genuine data in the unperceived region are identical.

Table I offers comparison and summary of various characteristics of MCS and SMCS.

C. LIFE-CYCLE OF MCS AND SMCS

Generally, the life-cycle of MCS and SMCS consists of four stages: task creation, task assignment, task execution, and data processing, as shown in Fig. 3. The key functionalities and research issues of each phase are described in the following.

- **Task creation**: After receiving task requesters' requirements (such as time, location, type, budgets, etc.), the MCS platform will construct MCS tasks. At this stage, the primary research issue is how to describe tasks without specialized knowledge effectively. The majority of existing efforts at this stage make use of a simple domain-specific language.
- **Task assignment**: At this stage, the MCS platform recruits workers and assigns them specific sensing tasks to do on their terminal devices. The most important difficulty at this point is to find enough qualified individuals to help with sensing tasks. Due to a variety of factors, some workers may reject those assigned tasks. As a result, participant recruiting, incentive systems, and privacy protection have been extensively studied at this stage. In addition, the selection of sensing areas is also a major challenge in SMCS.
- **Task execution**: Once the assigned individual sensing task is received, the workers will complete it within the pre-defined criteria, e.g., duration, location, etc. TThis state can be further broken down into three stages: sensing, calculating, and uploading, and its fundamental challenge is to reduce energy consumption.
- **Data processing**: At this stage, the MCS platform mainly processes the data streams from the crowd, including data cleaning, aggregation, and quality evaluation. Furthermore, the missing data inference is also a significant concern of SMCS. After that, the task requesters will get what they need in the appropriate manner from the MCS platform.

The key technologies mentioned above and their recent advancements will be introduced in detail in Sections III and IV.

III. COMMON KEY TECHNIQUES AND ADVANCEMENTS

SMCS is a novel sensing paradigm that is emerged from the development of MCS to reduce sensing costs. These two sensing paradigms face common challenges, e.g., *How to recruit participants and allocate them sub-tasks efficiently and cost-effectively? How to protect participants' private information, such as ID, geographic information, sensors, etc. How to motivate participants to engage in perception tasks and complete them well? How to assess the quality of collected data to judge the completion of sensing tasks effectively?* In this section, we provide a brief overview of the common key techniques that two sensing paradigms adopt, as well as their recent advances.

A. PARTICIPANT RECRUITMENT

Participant recruitment refers to the process of selecting suitable participants from candidates and completing the given tasks with certain goals, such as minimizing sensing costs, maximizing task completion ratio, maximizing data quality, and so on. It directly impacts the quality of completed activities, which in turn has an impact on the performance of sensing systems. The majority of studies concentrate on participant recruitment, and there are a lot of recent works in this field. When developing participant recruiting strategies, a variety of factors were examined, including user interests, reputation, social relationships, cost fairness, and so on. Wang *et al.* [23] considered the user's preference for tasks and turned it into a task-worker binary classification problem, determining the parameter weights to further calculate the matching probability between workers and tasks. Considering the reputation of users, Truong *et al.* [24] proposed a reputation evaluation model framework to recruit trusted participants to complete sensing tasks. Wang *et al.* [25] employed social networks as a recruitment platform to find users and assign them tasks.

Compared to MCS, participant recruitment is especially crucial for SMCS because only a tiny number of candidates are selected to participate in the perception task. Liu *et al.* [14] aimed to increase the accuracy of missing data inference in SMCS by employing appropriate users. They studied this problem from the two aspects of users and partitions and proposed a three-step strategy: selecting some candidate users and filtering out those who cover more sub-areas; selecting some effective sub-areas; utilizing the weighted calculation to ensure that those users who have a greater probability of passing through more selected sub-areas have greater weight, and finally recruiting these users with greater weight. Tu *et al.* [26] studied a reinforcement learning-based strategy for SMCS participant recruitment. They modeled the states, actions, and rewards of the recruitment system by taking into account the regularity of participants' movement trajectories and the differences in their contributions and used the Deep Q Network to determine which users are the best candidates to recruit.

Based on the above analysis, participant selection in MCS and SMCS is critical. Selecting high-quality users can ensure the task completion rate and improve the quality of collected data. In MCS and SMCS, how to assign tasks to high-quality users is still a significant research problem that needs to be solved urgently under the circumstances of malicious users or lack of user evaluation methods.

B. PRIVACY PROTECTION

Participants may be compelled to share their location, ID, and other information when executing tasks, posing a risk of privacy leakage [27]. Users, for example, often move to the target region to collect and upload data, and the submitted data may contain sensitive information, such as user ID, location, and so forth. Malicious users may exploit these to attack MCS systems. Due to concerns about privacy leaks, some users may refuse to participate in perception tasks, which is not conducive to completing large-scale tasks. Therefore, privacy protection is an essential problem for MCS and SMCS.

With the rise of a significant variety of location-based services, MCS has actively investigated location privacy in recent years [28], [29]. There are three types of location privacy protection strategies [30], [31]: location cloaking, obfuscation, and encryption-based. Cloak-based approaches [32], [33] depict fine-grained locations with coarse-grained location ranges. Instead of receiving a specific location, the server only gets a range value. Cloaking mechanisms are often built on k-anonymity, which ensures that a user's reported location matches the other k-1 users. Obfuscation-based methods consistently report a fake location instead of the real one. Differential privacy is an example of an obfuscation-based technique [34], [35]. It mostly maps real locations to obfuscated ones and uploads them to the server rather than the actual ones. Encryption-based approaches treat physical locations as data and then encrypt them [36], [37]. Liu *et al.* [38] designed a privacy-preserving truth discovery scheme to effectively achieve strong privacy protection for participants by using an additive homomorphic cryptosystem and secret sharing between two non-collusion servers.

The location cloaking, obfuscation, and encryption-based privacy protection approaches used by SMCS are the same as those utilized by MCS. Furthermore, in SMCS, the spatial-temporal correlation between data is critical for missing data inference. Some studies focus on privacy protection by taking Spatio-temporal correlation into consideration. By merging activity privacy with location differential privacy, Yang *et al.* [39] devised a novel technique to keep your whereabouts hidden. Wang *et al.* [35] decreased data quality loss in SMCS using differential privacy and then created a differential distortion privacy protection architecture to reduce data quality loss caused by location obfuscation.

Based on the aforementioned research work, MCS and SMCS confront the same privacy protection dilemma and use comparable strategies to defend against privacy leakage. Cloaking, obfuscation, and encryption are three of the most frequently used methods for privacy protection. The cloak-based method is significantly less effective when the adversary knows the location distribution of participants in advance [35], while differential privacy can reduce the harm of those opponents with prior knowledge. Differential privacy is one of the most widely used methods in privacy protection.

C. INCENTIVE MECHANISM

MCS relies on mobile users' wisdom and devices to explore the world, which incurs explicit costs (such as power) and implicit costs (such as the risk of privacy leakage) to users. Due to concerns about explicit or implicit costs, some participants may decline to take part in perception tasks, resulting in failed tasks or poor quality of collected data [40]. Therefore, it is critical to devise a mechanism to compensate users for the loss of participation in sensing activities and motivate them to participa in perception tasks.

To encourage more users to engage in sensing activities actively, researchers have designed a variety of incentive mechanisms that can be divided into two categories: monetary and non-monetary. Monetary incentives mainly motivate participants to complete tasks through payments, and they always adopt the auction mechanism [41], [42] to determine payments. Non-monetary incentives mainly use users' interests, hobbies, or social relationships to attract them to participate in tasks, including game incentives, virtual credit incentives, and social relationship incentives. Talasila *et al.* [43] designed a game called "Alien vs. Mobile User" where users are motivated to participate in tasks in surrounding areas by engaging in the game. They can also obtain monetary rewards after completing sensing tasks. Yu *et al.* [44] presented a virtual credit-based data flow sharing system whose goal is to encourage users with extra data flow to share it with others to alleviate the lack of data flow. Xu *et al.* [45] considered the collaborative compatibility of participants for multiple collaborative tasks and designed an incentive mechanism for multi-task collaboration that allows each collaborative activity to be performed by a group of compatible users. Liu *et al.* [46] introduced the concepts of "capital accumulation" and "intertemporal selection" in behavioral economics. They proposed an incentive mechanism to promote users' indulgence in cooperative behaviors and thus participate in tasks for a long time.

SMCS, like MCS, generally uses monetary or non-monetary incentives to encourage people to participate in sensing tasks. The incentive mechanism in SMCS, in particular, must consider not just recruiting more participants to perform tasks, but also improving the quality of the sparse data obtained for inferring global knowledge. Liu *et al.* [47] presented an incentive mechanism for air quality monitoring systems under SMCS. It assesses the data based on the difference between the data obtained in two consecutive rounds of perception cycles and gives rewards to attract enough users to engage in the perception activities. Gao *et al.* [48] proposed an incentive mechanism to maximize the quality of collected data and enable the platform to provide bonus rewards based on task completion levels and participants' previous performance.

In summary, incentive mechanisms are critical for MCS and SMCS. Existing studies usually focus on short-term sensing task allocation in MCS and SMCS. In reality, however, some large-scale and long-time sensing tasks need to recruit more users to participate in sensing activities over an extended period. In addition to reward strategies, cultural recognition should also be used to keep the long-term participant excited.

D. DATA QUALITY

In MCS, data quality is often quantified by coverage [49], which is calculated as a ratio of the number of sensing sub-areas to the total number of sub-areas. A straightforward method is to recruit more participants to the target areas to improve the coverage, while it will incur higher sensing costs. Moreover, sensors built into mobile devices have varying sensing accuracy, and participants typically have different experiences and expertise, all of which will alter the quality of obtained data. In addition, some malicious users may report fake data, causing data evaluation to be skewed. All of the aforementioned factors will degrade the quality of the data obtained and the accuracy of future data analysis. To obtain high-quality data, reputation evaluation mechanisms [50]–[52] have been developed to evaluate participants and assure the credibility of gathered data in order to get high-quality data. Truong *et al.* [24] proposed a reputation evaluation mechanism that selects credible participants for delivering high-quality data while preventing malevolent users from supplying fraudulent or incorrect data. Truskinger *et al.* [53] assessed reputation based on participants' previous performance and the opinions of other participants. Ren *et al.* [54] introduced bid prices into the reputation system to evaluate the cost-effectiveness of participants.

Since SMCS significantly depends on collected data to infer data for unsensed areas, the quality of the acquired data directly impacts the accuracy of the inferred data and, as a result, on the overall performance of the SMCS system. Therefore, data quality is one of the most concerning aspects of SMCS. Similarly, the above-mentioned factors also affect the quality of the obtained data in SMCS. In addition, the quality of inferred data is influenced by missing data inference. Truth discovery has attracted a lot of interest in recent years because of its capacity to measure user dependability and infer truthful information. Its goal is to utilize the gathered information to determine the underlying truth and assess the participants' trustworthiness. It may also be utilized in SMCS to increase data quality. Gao *et al.* [55] adopted the truth inference method for data inference and investigated the problem of efficient task allocation based on data quality inference to obtain sufficient high-quality data. Liu *et al.* [38] devised a privacy-preserving truth discovery scheme in SMCS scenarios. However, there are two important challenges to overcome in the present truth discovery. On the one hand, "the spread of misinformation" makes it harder to distinguish between accurate and false data. On the other hand, "data scarcity" can not supply sufficient evidence to find the truth. Zhang *et al.* [56] designed a robust truth discovery scheme to address these two challenges.

In addition, data quality evaluation is also another challenge in SMCS. The work in [11] emphasizes the challenge of inferred data evaluation and its importance. SMCS must assess the quality of inferred data to decide whether the sensing technique and missing data inference strategy need to be adjusted. There are two scenarios for data quality assessment in MCS and SMCS: underestimating or overestimating data quality. An incorrect number of sensing regions may be chosen as a consequence of underestimating or overestimating. Underestimating data quality, for example, will cause the selection of additional sensing regions, but overestimating will result in data redundancy. As a result, data quality evaluation is another critical problem that must be addressed immediately.

In reality, some other factors affect data quality, such as the subjectivity of participants uploading data, the remaining power, the performance and accuracy of sensors, etc. The quality of the acquired data determines the performance of the whole sensor system. So many factors must be considered to get high-quality data and determine how accurate the data is.

IV. UNIQUE KEY TECHNIQUES AND ADVANCEMENTS

Sparse mobile crowdsensing is a representative example of sensing cost-cutting for mobile crowdsensing. As illustrated in Fig. 2, it primarily relies on the Spatio-temporal correlation between data to infer data for unsensed sub-areas, lowering total sensing costs and ensuring data quality [11]. In reality, SMCS faces two significant challenges: *(i) how to choose a set of suitable sensing areas from which collected data can be used to infer data for unsensed areas, (ii) and how to infer data for unsensed areas by only involving obtained data*. This section will take a close look at these two major challenges and their recent developments.

A. SENSING AREA SELECTION

Sensing area selection is vital for SMCS because the quality of data collected from the selected areas directly determines the inferred data. On the one hand, data obtained from various sub-areas leads to different levels of inferred data. On the other hand, a combination of distinct sub-areas will have different effects on inferred data. If suitable sensing regions are chosen, a tiny quantity of acquired data may offer adequate knowledge about the target, allowing the number of sensing areas to be reduced and hence the sensing cost to be reduced.

Existing studies often consider different factors or scenarios in sensing area selection, such as cost, participant characteristics, multi-dimensional urban scenarios, and so on. Zhu *et al.* [57] considered the diversity of costs (such as routing, measurement, and sensing) to determine the target sensing areas. They devised a three-stage selection method, which included information modeling, cost calculation, and a cost-quality benefit option to increase the quantity of data in the chosen sub-areas while lowering the overall cost. Sun *et al.* [13] considered the heterogeneity of sensing areas and malicious users when picking target sensing locations. To begin, reliable users are identified using an iterative statistical spatial interpolation approach. Then, the regularized mutual coherence is used to define the contribution of sensing data given by various users to inference accuracy. Finally, an optimization problem with regularized mutual coherence constraints determines the sensing areas. Liu *et al.* [61] investigated multi-dimensional city sensing in multi-task settings, including missing data inference and sensing area selection. To infer the data for unsensed regions, they first employed intra-task and inter-task correlations and multi-task compressed sensing. Following that, the most successful <area, task> pairings are chosen using a reinforcement learning technique.

According to our survey, the QBC (Query-by-Committee) approach was employed to tackle the issue of sensing area selection in the majority of prior research [58], [59]. To infer a full map of the target region, QBC initially employs various missing data inference techniques. Then it chooses the most uncertain sensing regions and continues the procedure until a sufficient number of sub-areas have been chosen. Here, the most uncertain sensing areas are not always the most informative or helpful for missing data inference. As a result, the current work on sensing area selection focuses on identifying the most valuable target regions. Xie *et al.* [60] devised a sensing scheduling algorithm that carefully selects sampling locations in each time slot, resulting in a more informative result.

Reinforcement learning has recently been shown to be effective in sensing area selection [12], [14]. It begins with the modeling of states, actions, and rewards. Here, the state symbolizes the current area, time, and place as a whole. The action specifies the places that will be chosen. The reward refers to what you'll get if you choose a sensing region. Then, a reinforcement learning-based system is created to choose the most appropriate sensing regions by using the aforesaid data. It also trains a Q-function to calculate the reward generated by each action in the model. If a better reward is obtained, the corresponding action is considered a good action. Liu *et al.* [61] employed reinforcement learning to combine missing data inference and sensing area selection to decide the best <area, task> pair. In addition, Han *et al.* [63] presented a reinforcement learning-based region module update approach from SMCS because the social environment evolves over time.

Furthermore, specific current approaches, such as mobile edge computing, subarea division learning, transfer learning, and so on, have lately been used in sensing area selection. Xia *et al.* [64] utilized mobile edge computing to assist sensing area selection. Wei *et al.* [65] made use of subarea division learning to establish an uneven regional division to guide task allocation (sensing area selection) by considering historical data and Spatio-temporal correlations. Wang *et al.* [59] used intra-data correlation within the same type of collected data and inter-data correlation between different types of obtained data to minimize sensing expenses. They designed a new task allocation framework in which a set of sensing areas is selected at each period (cycle) to collect data by using compressed sensing, statistical analysis, active learning, and

TABLE 2. A Summary of Related Work on Sensing Area Selection

Issue	Reference	Research Scenario	Objective	Strategy
Sensing Area Selection	Zhu *et al.* [57]	SMCS with diverse sensing costs	Maximizing the informativeness in the selected subareas and minimizing the total sample costs	Greedy selection and the Pareto optimization selection
	Sun *et al.* [13]	Sensing area heterogeneity and malicious participants	Meeting the desired level of sensing quality with minimized sensing cost	Optimization selection
	Wang *et al.* [58]	Quality-guaranteed online task allocation	Reducing the required number of sensing tasks allocated and ensuring the data quality	Query-By-Committee
	Wang *et al.* [59]	Cost-efficient task allocation	Reducing the number of sensing task assignments and ensuring data quality	Query-By-Committee
	Xie *et al.* [60]	Active SMCS based on matrix completion	Reducing the cost while ensuring the quality of missing data inference	Sampling scheduling algorithm
	Liu *et al.* [61]	Multi-dimensional urban sensing	Addressing sensing area selection for multi-task scenarios	Reinforcement Learning
	Liu *et al.* [12]	User recruitment on both the users and subareas sides	Enhancing data inference accuracy	Reinforcement Learning
	Liu *et al.* [14]	Sensing area selection in SMCS	Ensuring a certain level of data quality	Reinforcement Learning
	Wang *et al.* [62]	Sensing area selection in SMCS	Reducing sensing cost and ensuring data quality	Deep Reinforcement Learning
	Han *et al.* [63]	Time-dependent urban environment	Keeping sensing area selection model up-to-date	Reinforcement Learning
	Xia *et al.* [64]	Quality-aware sparse data collection	Reducing the data redundancy and selecting the appropriate users' group	Mobile Edge Computing
	Wei *et al.* [65]	No participants perform the task in vital areas	Reducing sensing cost	Subarea Division Learning

TABLE 3. A Summary of Related Work on Missing Data Inference

Issue	Reference	Research Scenario	Objective	Strategy
Missing Data Inference	Zhu *et al.* [66]	Road traffic conditions	Solving the missing data problem	Compressed Sensing
	Quer *et al.* [67]	Real WSNs scenarios	Compressing large and distributed signals monitored by WSNs	Compressed Sensing
	Kong *et al.* [68]	Environment reconstruction	Reconstructing the massive missing data	Compressed Sensing
	Rana *et al.* [4]	Urban noise map	Recovering the noise map from incomplete and random samples obtained	Compressed Sensing
	Xu *et al.* [69]	Mobile crowdsourcing scenarios.	Reducing the high burden on the users	Compressed Sensing
	Liu *et al.* [15], [47]	Monitoring urban air quality conditions	Designing an efficient incentive mechanism to encourage users with low costs participating.	Compressed Sensing
	He *et al.* [70]	Construction of signal maps	Balancing signal quality and crowdsourcing costs	Bayesian Compressed Sensing
	Xie *et al.* [60]	The large scale deployment of Mobile Crowd Sensing (MCS)	Recovering the unsampled data and low cost sample selection	Matrix Completion
	Zhang *et al.* [71]	Inductive prediction	Resolving no additional information available	An Inductive Graph-based Matrix Completion (IGMC) model
	Wang *et al.* [72]	Urban application scenarios	Considering outlier effects	Matrix Completion
	Pan *et al.* [73]	Wireless sensor networks	Estimating missing sensor data	K-nearest Neighbor algorithm
	Yan *et al.* [74]	The Internet of things	Resolving missing value imputation problem in IoT	Gaussian mixture model
	Stekhoven *et al.* [75]	Missing data based on high-throughput technology	Handling missing values	Random forest model algorithm
	Alimpertis *et al.* [76]	Signal strength maps	Improving signal strength maps from limited measurements	Random forest model algorithm

transfer learning. Table 2 summarizes the related efforts on sensing area selection in SMCS.

B. MISSING DATA INFERENCE

Missing data inference is an essential part of SMCS, which is directly influenced by the quality of inferred data. The majority of extant research in this field now relies on distinct sub-areas' temporal and geographical correlation to infer data for the remaining unsensed areas. As a result, it is vital and challenging to infer data for unsensed regions and ensure their quality using the obtained data. Table 3 summarizes a collection of related works on missing data inference.

Researchers have conducted in-depth research on missing data inference before the emergence of SMCS. In WSNs, a large number of sensors are deployed in each area for data collection. Quer *et al.* [67] recovered large-scale distributed signals across the sensing area by collecting a small number of samples. Kong *et al.* [68] suggested a compressed sensing-based massive data reconstruction approach to solve data loss in WSNs. Zhu *et al.* [66] employed a compressed sensing algorithm to overcome the problem of missing data and predict road traffic conditions.

As far as we know, much effort has been devoted to data inference, such as compressed sensing, matrix completion, k-nearest neighbour, parametric Gaussian mixture model, random forest model, etc. Compressed sensing is a signal reconstruction method for acquiring and reconstructing sparse or compressible signals. In other words, the process of signal sampling uses a few samples to produce the same impact as complete sampling. It has been utilized extensively in traffic monitoring, sensor networks, and MCS. Rana *et al.* [4] utilized compressive sensing to deal with the obtained incomplete and a random sample data collections, thereby recovering a complete urban noise map. Xu *et al.* [69] applied compressed sensing to MCS scenarios to tackle data loss problems. Liu *et al.* [15], [47] leveraged compressed sensing as a missing data inference algorithm to monitor urban air quality. It merely collects data in a few locations and then infers the severity of air pollution in areas where it hasn't been observed. He *et al.* [70] proposed a BCCS mechanism to reconstruct signal maps by combining crowdsourcing with Bayesian compressed sensing.

Some other researchers employ matrix completion algorithms to solve the problem of missing data inference. Considering the Spatio-temporal correlation of environmental monitoring, Xie *et al.* [60] proposed a bipartite graph-based active SMCS scheme and matrix completion algorithm to recover unsampled data in the presence of sensing and communication errors. Zhang *et al.* [71] proposed a graph neural network-based matrix completion (IGMC) model without using side information, which can be used to infer data for SMCS. Wang *et al.* [72] offered an outlier-based matrix completion method to predict unencrypted data efficiently.

In addition, there are other techniques for missing data inference, such as the k-nearest neighbor algorithm, the Gaussian mixture model, the random forest model algorithm, and so on. Pan *et al.* [73] proposed a k-nearest neighbors-based missing data estimation algorithm that employs a linear regression model to characterize the spatial correlation of data between distinct sensor nodes and the data of many neighboring nodes to jointly further estimate the missing data. Yan *et al.* [74] divided missing data into three categories according to the characteristics of IoT and missing data. They proposed three new models to solve the corresponding missing data imputation problem. Stekhoven *et al.* [75] devised an iterative imputation method based on random forests and demonstrated its superiority in dealing with missing values in nonlinear relationships. Alimpertis *et al.* [76] adopted the random forest to overcome missing data inference concerns. They developed a random forest-based prediction framework to improve finitely predicted signal strength maps, which are important for network planning and operations of cellular network providers.

In a nutshell, much effort has been devoted to missing data inference, and plenty of related works have been recently developed. However, there are still several issues to be further investigated. It is very challenging to design an efficient missing data inference algorithm. Some fundamental SMCS features, such as sensing area selection and sensing devices, are exploited to enhance the performance of missing data inference methods.

V. APPLICATIONS AND FUTURE RESEARCH OPPORTUNITIES

A. APPLICATIONS

As the last "mile" of SMCS projects, the collected data from the crowd and its processing provide solutions to various applications for constructing a smart city. Since its unique perception, SMCS is naturally low-cost and ideal for urban sensing. In general, the SMCS-implemented smart city strives to better assist and improve our everyday lives from three critical perspectives, infrastructure and social services. This section summarizes possible and existing applications to aid in defining the requirements of future SMCS.

1) ENVIRONMENTAL MONITORING

Environmental monitoring is critical for the long-term growth of cities and the improvement of their inhabitants' quality of life. Its goal is to monitor resource consumption, infrastructure conditions, and urban phenomena, which are well-known issues affecting residents' daily lives. For example, air pollution causes a range of respiratory ailments and may lead to cancer if people are exposed for extended periods of time. Traditional environmental monitoring sensing paradigms (WSNs and MCS) always have high sensing costs. SMCS can fully exploit the spatial-temporal correlation among environmental data to reduce the number of sensing areas, dramatically lowering sensing expenses. As a result, SMCS is particularly well suited to environmental monitoring. An environmental monitoring scenario is shown in Fig. 4.

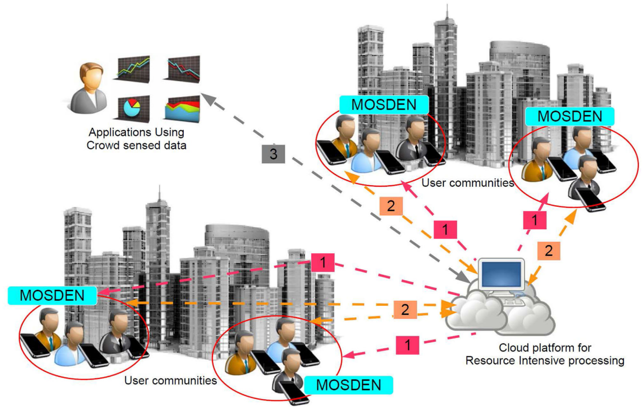

FIGURE 4. An environmental monitoring scenario [78].

Hao *et al.* [77] developed a system for assessing traffic-related air pollution that can effectively, quickly, and affordably estimate the environmental impact of traffic congestion. Air quality data is acquired in certain places and inferred for other unsensed regions by using the Spatio-temporal correlation. Finally, a comprehensive air quality map of the target region is created. Ear-phone [4] is widely used in SMCS projects. It collects noise data from the surrounding area and calculates the noise level using the smartphone's microphone and GPS sensor. Due to the limitation of participant mobility, it is not possible to gather data from the whole target region. To infer noise data for such unsigned regions, a compressed sensing approach is provided, and a comprehensive noise map for the whole target region is eventually constructed.

2) INTELLIGENT TRANSPORTATION

The refined management of smart cities relies on a large amount of data sensed from urban areas. Since SMCS can obtain a great number of urban data at a low cost, it has lately been applied to assist city management, particularly for intelligent transportation. The GPS traces are always selected to be detected to monitor the traffic conditions for all modes of transportation. To correct existing maps, CrowdAtlas [79] devised an automated road inference system using GPS probes. This system can automatically solve the drawbacks of existing

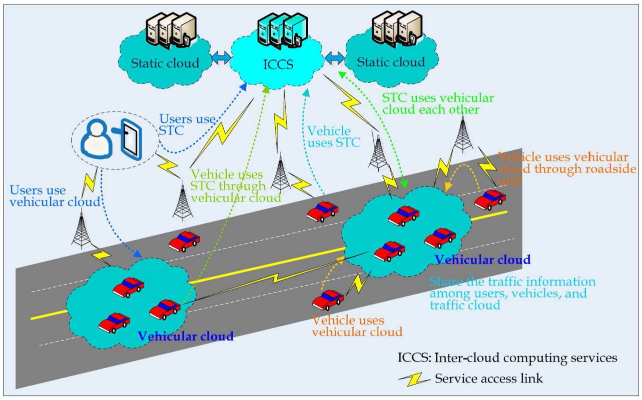

FIGURE 5. **An intelligent transportation scenario [81].**

digital road maps by updating them in real-time. The participants in CrowdAtlas may choose to submit only mismatched segments to the server, limiting their privacy exposure. Zhu *et al.* [66] presented a modified compressive sensing approach to estimate urban traffic speeds based on periodically collected data by probe vehicles. Wang *et al.* [80] investigated how to estimate the speed of a road section using sparse sensing data. In the temporal and geographical dimensions, they acquired the complete coverage distribution of the average driving speed of the road section. SMCS has shown its superiority in intelligent transportation and may extend to other aspects of the city administration in the future. Fig. 5 shows an intelligent transportation scenario.

3) SOCIAL SERVICES

Mobile devices have become an indispensable part of our lives with their popularity. They are often used to fulfill various activities in people's everyday lives. When individuals run, for example, their movement trajectories and body status data are logged; when they interact with social networks, their personal preferences and views are recorded. Consequently, a vast quantity of information about humans has been collected and can be used to exploit the characteristics of individuals or groups of people [82]. If the target is a person, individual features such as daily habits, preferences and health are exposed. If the target is a group of people, it can reveal or predict social relationships and even safety issues. However, due to the variety and scale of social networks, the data obtained from these networks is often fragmented and incomplete and thus cannot be used directly for analysis and prediction. SMCS can compensate for this deficiency and be used to infer the missing data. By analyzing the relationship between people and place, Zhou *et al.* [83] propose a three-step travel planning process: activity preparation, group activity mining, and activity suggestion. Liu *et al.* [84] investigated the problem of fine-grained city prediction. They proposed an urban prediction scheme via SMCS consisting of the matrix completion and the near-future prediction to predict the fine-grained complete picture of future sensing data.

B. FUTURE RESEARCH DIRECTIONS

Although SMCS has made great success in recent years, there are still some limitations to be overcome. The gap between the ideal problem setting and real-world applications still inhibits the widespread adoption of SMCS. In this section, we highlight some innovative and fascinating research directions in the field of SMCS.

1) REINFORCEMENT LEARNING-BASED SENSING AREA SELECTION

Sensing area selection and missing data inference are two major difficulties in SMCS. The front's quality has a direct impact on the latter. Recent work has demonstrated that reinforcement learning is a powerful technique for evaluating sensing areas [12], [14], [61]. Moreover, reinforcement learning and deep learning have been exploited to choose those sensing regions with distinct properties [62]. It seems that reinforcement learning is a promising approach for guiding the selection of sensing regions. The majority of existing SMCS work focuses on single-characterized perceptual domains. On the other hand, there are many varied sensing regions, such as traffic flow changes over time. In the future, reinforcement learning should be used to pick sensing zones in diverse settings.

2) SPATIO-TEMPORAL CORRELATION IN MULTI-TASK

In existing work, researchers usually consider the Spatio-temporal correlation of single-task scenarios in SMCS [4], [11], [58], [62], [66], [70]. However, in reality, the SMCS platform often publishes numerous tasks. And there is a specific correlation between various sensing data in multiple tasks. For example, higher temperatures generally result in lower humidity, and poor air quality may result in higher PM2.5. Therefore, we can make full use of the Spatio-temporal correlation within single-task data and between multi-task data to meet the needs of SMCS in more complex real-world scenarios.

3) MULTI-DIMENSIONAL DATA INFERENCE AND FUTURE FORECASTING

Here, "multi-dimensional" can be understood from two aspects. On the one hand, "multi-dimensional" refers to different data collected from a single sensing task. These data contains intrinsic correlations that may be leveraged to minimize sensing costs and increase inferred data accuracy [84]. For example, large-scale urban sensing systems often require multiple kinds of sensing data (such as temperature and humidity) to construct a multi-dimensional urban sensing map. On the other hand, "multi-dimensional" denotes a variety of missing data inference methodologies. Each of them has pros and cons. Taking multiple techniques into account, using their strengths, and avoiding their flaws, is one alternative strategy to constructing data reasoning. We believe that multi-dimensional data inference is still an undiscovered gem deserving further investigation.

In addition, the existing data inference work of SMCS is mainly used to infer unsensed areas' data under the current sensing state.We believe that it is possible to strengthen the use of Spatio-temporal correlation in SMCS by accumulating collected and inferred data. These data are used to infer the data information of a certain period in the future of the sensing task and to predict the future sensing task to save costs better.Therefore, utilizing SMCS for future sensing task prediction is also an important direction to promote in both academia and industry.

VI. CONCLUSION

In this paper, we provide a comprehensive literature review of recent advances in sparse mobile crowdsensing (SMCS) from the perspective of comparing MCS. First, we provide an overview of SMCS, clarifying its evolution, characteristics, and life-cycle states. Then, we go through some of the common key techniques between MCS and SMCS, as well as their recent advancements. After that, we give a review of the unique key techniques of SMCS and their advances. Finally, we summarize current SMCS applications and highlight some fascinating research possibilities in the field of SMCS.

REFERENCES

[1] R. K. Ganti, F. Ye, and H. Lei, "Mobile crowdsensing: Current state and future challenges," *IEEE Commun. Mag.*, vol. 49, no. 11, pp. 32–39, Nov. 2011.

[2] J. Wang, L. Wang, Y. Wang, D. Zhang, and L. Kong, "Task allocation in mobile crowd sensing: State-of-the-art and future opportunities," *IEEE Internet Things J.*, vol. 5, no. 5, pp. 3747–3757, Oct. 2018.

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

[4] R. K. Rana, C. T. Chou, S. S. Kanhere, N. Bulusu, and W. Hu, "Ear-phone: An end-to-end participatory urban noise mapping system," in *Proc. 9th ACM/IEEE Int. Conf. Inf. Process. Sensor Netw.*, 2010, pp. 105–116.

[5] K. Farkas, G. Feher, A. Benczur, and C. Sidlo, "Crowdsending based public transport information service in smart cities," *IEEE Commun. Mag.*, vol. 53, no. 8, pp. 158–165, Aug. 2015.

[6] R. Gao *et al.*, "Sextant: Towards ubiquitous indoor localization service by photo-taking of the environment," *IEEE Trans. Mobile Comput.*, vol. 15, no. 2, pp. 460–474, Feb. 2016.

[7] Y. Wu, Y. Wang, W. Hu, and G. Cao, "Smartphoto: A resource-aware crowdsourcing approach for image sensing with smartphones," *IEEE Trans. Mobile Comput.*, vol. 15, no. 5, pp. 1249–1263, May 2016.

[8] K. Wei *et al.*, "High-performance UAV crowdsensing: A deep reinforcement learning approach," *IEEE Internet Things J.*, early access, Mar. 21, 2022, doi: 10.1109/JIOT.2022.3160887.

[9] Y. Liu *et al.*, "Multi-task allocation in mobile crowd sensing," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2016, pp. 403–414.

[10] H. Xiong, D. Zhang, L. Wang, J. P. Gibson, and J. Zhu, "EEMC: Enabling energy-efficient mobile crowdsensing with anonymous participants," *ACM Trans. Intell. Syst. Technol.*, vol. 6, no. 3, pp. 1–26, 2015.

[11] 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.

[12] 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, 2019.

[13] 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, Jul. 2021.

[14] W. Liu, Y. Yang, E. Wang, and J. Wu, "User recruitment for enhancing data inference accuracy in sparse mobile crowdsensing," *IEEE Internet Things J.*, vol. 7, no. 3, pp. 1802–1814, Mar. 2020.

[15] T. Liu, Y. Zhu, Y. Yang, and F. Ye, "ALC^2 : When active learning meets compressive crowdsensing for urban air pollution monitoring," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 9427–9438, Dec. 2019.

[16] A. Capponi, C. Fiandrino, B. Kantarci, L. Foschini, D. Kliazovich, and P. Bouvry, "A survey on mobile crowdsensing systems: Challenges, solutions, and opportunities," *IEEE Commun. Surv. Tut.*, vol. 21, no. 3, pp. 2419–2465, Jul.–Sep. 2019.

[17] K. Wei, X. Liang, and K. Xu, "A survey of social-aware routing protocols in delay tolerant networks: Applications, taxonomy and design-related issues," *IEEE Commun. Surv. Tut.*, vol. 16, no. 1, pp. 556–578, Mar. 2014.

[18] J. Yu, S. Wan, X. Cheng, and D. Yu, "Coverage contribution area based k-coverage for wireless sensor networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 8510–8523, Sep. 2017.

[19] W. Guo, W. Zhu, Z. Yu, J. Wang, and B. Guo, "A survey of task allocation: Contrastive perspectives from wireless sensor networks and mobile crowdsensing," *IEEE Access*, vol. 7, pp. 78406–78420, 2019.

[20] J. Howe *et al.*, "The rise of crowdsourcing," *Wired Mag.*, vol. 14, no. 6, pp. 1–4, 2006.

[21] G. Bajaj and P. Singh, "An in-depth analysis of the impact of battery usage patterns on performance of task allocation algorithms in sparse mobile crowdsensing," in *Proc. 22nd Int. ACM Conf. Model., Anal. Simul. Wireless Mobile Syst.*, 2019, pp. 297–306.

[22] H. Hong, X. Chaocan, X. Shucheng, S. Xin, Y. Panlong, and G. Jibin, "Survey on crowd–sensing networks," *J. Jilin Univ. (Inf. Sci. Ed.)*, vol. 34, no. 3, pp. 374–383, 2016.

[23] Z. Wang *et al.*, "Towards personalized task-oriented worker recruitment in mobile crowdsensing," *IEEE Trans. Mobile Comput.*, vol. 20, no. 5, pp. 2080–2093, May 2021.

[24] N. B. Truong, G. M. Lee, T.-W. Um, and M. Mackay, "Trust evaluation mechanism for user recruitment in mobile crowd-sensing in the Internet of Things," *IEEE Trans. Inf. Forensics Secur.*, vol. 14, no. 10, pp. 2705–2719, Oct. 2019.

[25] J. Wang, F. Wang, Y. Wang, D. Zhang, L. Wang, and Z. Qiu, "Social-network-assisted worker recruitment in mobile crowd sensing," *IEEE Trans. Mobile Comput.*, vol. 18, no. 7, pp. 1661–1673, Jul. 2019.

[26] T. Chun-Yu *et al.*, "Direct participant recruitment strategy in sparse mobile crowdsensing," *Chin. J. Comput.*, 2022.

[27] B. Guo *et al.*, "Mobile crowd sensing and computing: The review of an emerging human-powered sensing paradigm," *ACM Comput. Surv.*, vol. 48, no. 1, pp. 1–31, 2015.

[28] Z. Yu, H. Xu, Z. Yang, and B. Guo, "Personalized travel package with multi-point-of-interest recommendation based on crowdsourced user footprints," *IEEE Trans. Hum.-Mach. Syst.*, vol. 46, no. 1, pp. 151–158, Feb. 2016.

[29] J. Krumm, "A survey of computational location privacy," *Pers. Ubiquitous Comput.*, vol. 13, no. 6, pp. 391–399, 2009.

[30] C. Huang, D. Wang, and S. Zhu, "Where are you from: Home location profiling of crowd sensors from noisy and sparse crowdsourcing data," in *Proc. IEEE Conf. Comput. Commun.*, 2017, pp. 1–9.

[31] L. Pournajaf, D. A. Garcia-Ulloa, L. Xiong, and V. Sunderam, "Participant privacy in mobile crowd sensing task management: A survey of methods and challenges," *ACM Sigmod Rec.*, vol. 44, no. 4, pp. 23–34, 2016.

[32] C. Cornelius, A. Kapadia, D. Kotz, D. Peebles, M. Shin, and N. Triandopoulos, "Anonysense: Privacy-aware people-centric sensing," in *Proc. 6th Int. Conf. Mobile Syst., Appl., Serv.*, 2008, pp. 211–224.

[33] L. Pournajaf, L. Xiong, V. Sunderam, and S. Goryczka, "Spatial task assignment for crowd sensing with cloaked locations," in *Proc. 15th Int. Conf. Mobile Data Manage.*, vol. 1, 2014, pp. 73–82.

[34] X. Jin, R. Zhang, Y. Chen, T. Li, and Y. Zhang, "Dpsense: Differentially private crowdsourced spectrum sensing," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2016, pp. 296–307.

[35] L. Wang, D. Zhang, D. Yang, B. Y. Lim, and X. Ma, "Differential location privacy for sparse mobile crowdsensing," in *Proc. 16th Int. Conf. Data Mining*, 2016, pp. 1257–1262.

[36] A. Alanwar, Y. Shoukry, S. Chakraborty, P. Martin, P. Tabuada, and M. Srivastava, "Proloc: Resilient localization with private observers using partial homomorphic encryption," in *Proc. 16th ACM/IEEE Int. Conf. Inf. Process. Sensor Netw.*, 2017, pp. 41–52.

[37] L. Kong, L. He, X.-Y. Liu, Y. Gu, M.-Y. Wu, and X. Liu, "Privacy-preserving compressive sensing for crowdsensing based trajectory recovery," in *Proc. 35th Int. Conf. Distrib. Comput. Syst.*, 2015, pp. 31–40.

[38] F. Liu, B. Zhu, S. Yuan, J. Li, and K. Xue, "Privacy-preserving truth discovery for sparse data in mobile crowdsensing systems," in *Proc. IEEE Glob. Commun. Conf.*, 2021, pp. 1–6.

[39] Q. Yang, Y. Chen, M. Guizani, and G. M. Lee, "Spatiotemporal location differential privacy for sparse mobile crowdsensing," in *Proc. Int. Wireless Commun. Mobile Comput.*, 2021, pp. 1734–1741.

[40] H. Gao *et al.*, "A survey of incentive mechanisms for participatory sensing," *IEEE Commun. Surv. Tut.*, vol. 17, no. 2, pp. 918–943, Apr.–Jun. 2015.

[41] D. Yang, G. Xue, X. Fang, and J. Tang, "Incentive mechanisms for crowdsensing: Crowdsourcing with smartphones," *IEEE/ACM Trans. Netw.*, vol. 24, no. 3, pp. 1732–1744, Jun. 2016.

[42] F. Restuccia, P. Ferraro, S. Silvestri, S. K. Das, and G. L. Re, "*IncentMe*: Effective mechanism design to stimulate crowdsensing participants with uncertain mobility," *IEEE Trans. Mobile Comput.*, vol. 18, no. 7, pp. 1571–1584, Jul. 2019.

[43] M. Talasila, R. Curtmola, and C. Borcea, "Crowdsensing in the wild with aliens and micropayments," *IEEE Pervasive Comput.*, vol. 15, no. 1, pp. 68–77, Jan.–Mar. 2016.

[44] T. Yu, Z. Zhou, D. Zhang, X. Wang, Y. Liu, and S. Lu, "INDAPSON: An incentive data plan sharing system based on self-organizing network," in *Proc. IEEE Conf. Comput. Commun.*, 2014, pp. 1545–1553.

[45] J. Xu, Z. Rao, L. Xu, D. Yang, and T. Li, "Incentive mechanism for multiple cooperative tasks with compatible users in mobile crowd sensing via online communities," *IEEE Trans. Mobile Comput.*, vol. 19, no. 7, pp. 1618–1633, Jul. 2020.

[46] J. Liu, S. Huang, D. Li, S. Wen, and H. Liu, "Addictive incentive mechanism in crowdsensing from the perspective of behavioral economics," *IEEE Trans. Parallel Distrib. Syst.*, vol. 33, no. 5, pp. 1109–1127, May 2022.

[47] T. Liu, Y. Zhu, Y. Yang, and F. Ye, "Incentive design for air pollution monitoring based on compressive crowdsensing," in *Proc. IEEE Glob. Commun. Conf.*, 2016, pp. 1–6.

[48] H. Gao, C. H. Liu, J. Tang, D. Yang, P. Hui, and W. Wang, "Online quality-aware incentive mechanism for mobile crowd sensing with extra bonus," *IEEE Trans. Mobile Comput.*, vol. 18, no. 11, pp. 2589–2603, Nov. 2019.

[49] Z. He, J. Cao, and X. Liu, "High quality participant recruitment in vehicle-based crowdsourcing using predictable mobility," in *Proc. IEEE Conf. Comput. Commun.*, 2015, pp. 2542–2550.

[50] K. L. Huang, S. S. Kanhere, and W. Hu, "On the need for a reputation system in mobile phone based sensing," *Ad Hoc Netw.*, vol. 12, pp. 130–149, 2014.

[51] H. Amintoosi and S. S. Kanhere, "A reputation framework for social participatory sensing systems," *Mobile Netw. Appl.*, vol. 19, no. 1, pp. 88–100, 2014.

[52] X. O. Wang, W. Cheng, P. Mohapatra, and T. Abdelzaher, "Artsense: Anonymous reputation and trust in participatory sensing," in *Proc. IEEE Conf. Comput. Commun.*, 2013, pp. 2517–2525.

[53] A. Truskinger, H. Yang, J. Wimmer, J. Zhang, I. Williamson, and P. Roe, "Large scale participatory acoustic sensor data analysis: Tools and reputation models to enhance effectiveness," in *Proc. 7th Int. Conf. eScience*, 2011, pp. 150–157.

[54] J. Ren, Y. Zhang, K. Zhang, and X. S. Shen, "Sacrm: Social aware crowdsourcing with reputation management in mobile sensing," *Comput. Commun.*, vol. 65, pp. 55–65, 2015.

[55] X. Gao, H. Huang, C. Liu, F. Wu, and G. Chen, "Quality inference based task assignment in mobile crowdsensing," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 10, pp. 3410–3423, Oct. 2021.

[56] D. Y. Zhang, R. Han, D. Wang, and C. Huang, "On robust truth discovery in sparse social media sensing," in *Proc. IEEE Int. Conf.*, 2016, pp. 1076–1081.

[57] 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 J.*, vol. 8, no. 5, pp. 3831–3850, Mar. 2021.

[58] L. Wang *et al.*, "CCS-TA: Quality-guaranteed online task allocation in compressive crowdsensing," in *Proc. 2015 ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2015, pp. 683–694.

[59] L. Wang *et al.*, "SPACE-TA: Cost-effective task allocation exploiting intradata and interdata correlations in sparse crowdsensing," *ACM Trans. Intell. Syst. Technol.*, vol. 9, no. 2, pp. 1–28, 2017.

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

[61] W. Liu, Y. Yang, E. Wang, L. Wang, D. Zeghlache, and D. Zhang, "Multi-dimensional urban sensing in sparse mobile crowdsensing," *IEEE Access*, vol. 7, pp. 82066–82079, 2019.

[62] 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. 38th Int. Conf. Distrib. Comput. Syst.*, 2018, pp. 1543–1546.

[63] L. Han, Z. Yu, L. Wang, Z. Yu, and B. Guo, "Keeping cell selection model up-to-date to adapt to time-dependent environment in sparse mobile crowdsensing," *IEEE Internet Things J.*, vol. 8, no. 18, pp. 13914–13925, Sep. 2021.

[64] X. Xia, Y. Zhou, J. Li, and R. Yu, "Quality-aware sparse data collection in mec-enhanced mobile crowdsensing systems," *IEEE Trans. Computat. Social Syst.*, vol. 6, no. 5, pp. 1051–1062, Oct. 2019.

[65] 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. Topics Comput.*, vol. 9, no. 3, pp. 1344–1358, Jul.–Sep. 2021.

[66] 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, Nov. 2013.

[67] G. Quer, R. Masiero, G. Pillonetto, M. Rossi, and M. Zorzi, "Sensing, compression, and recovery for WSNs: Sparse signal modeling and monitoring framework," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3447–3461, Oct. 2012.

[68] L. Kong, M. Xia, X.-Y. Liu, M.-Y. Wu, and X. Liu, "Data loss and reconstruction in sensor networks," in *Proc. IEEE Conf. Comput. Commun.*, 2013, pp. 1654–1662.

[69] L. Xu, X. Hao, N. D. Lane, X. Liu, and T. Moscibroda, "More with less: Lowering user burden in mobile crowdsourcing through compressive sensing," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2015, pp. 659–670.

[70] S. He and K. G. Shin, "Steering crowdsourced signal map construction via Bayesian compressive sensing," in *Proc. IEEE Conf. Comput. Commun.*, 2018, pp. 1016–1024.

[71] M. Zhang and Y. Chen, "Inductive matrix completion based on graph neural networks," in *Proc. 8th Int. Conf. Learn. Representations*, 2020.

[72] E. Wang, M. Zhang, Y. Yang, Y. Xu, and J. Wu, "Exploiting outlier value effects in sparse urban crowdsensing," in *Proc. 29th Int. Symp. Qual. Serv.*, 2021, pp. 1–10.

[73] L. Pan *et al.*, "K-nearest neighbor based missing data estimation algorithm in wireless sensor networks," *Wireless Sensor Netw.*, vol. 2, no. 2, pp. 115–122, 2010.

[74] X. Yan, W. Xiong, L. Hu, F. Wang, and K. Zhao, "Missing value imputation based on gaussian mixture model for the Internet of Things," *Math. Problems Eng.*, vol. 2015, 2015.

[75] D. J. Stekhoven and P. Bühlmann, "Missforest–non-parametric missing value imputation for mixed-type data," *Bioinformatics*, vol. 28, no. 1, pp. 112–118, 2012.

[76] E. Alimpertis, A. Markopoulou, C. Butts, and K. Psounis, "City-wide signal strength maps: Prediction with random forests," in *Proc. World Wide Web Conf.*, 2019, pp. 2536–2542.

[77] P. Hao, C. Wang, G. Wu, K. Boriboonsomsin, and M. Barth, "Evaluating the environmental impact of traffic congestion based on sparse mobile crowd-sourced data," in *Proc. IEEE Conf. Technol. Sustainability*, 2017, pp. 1–6.

[78] P. P. Jayaraman, C. Perera, D. Georgakopoulos, and A. B. Zaslavsky, "Efficient opportunistic sensing using mobile collaborative platform MOSDEN," in *Proc. 9th IEEE Int. Conf. Collaborative Comput.: Netw., Appl. Worksharing*, 2013, pp. 77–86.

[79] Y. Wang, X. Liu, H. Wei, G. Forman, C. Chen, and Y. Zhu, "Crowdatlas: Self-updating maps for cloud and personal use," in *Proc. 11th Annu. Int. Conf. Mobile Syst., Appl., Serv.*, 2013, pp. 27–40.

[80] 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. Syst.*, vol. 20, no. 9, pp. 3486–3495, Sep. 2019.

[81] X. Kang, L. Liu, and H. Ma, "Enhance the quality of crowdsensing for fine-grained urban environment monitoring via data correlation," *Sensors*, vol. 17, no. 1, 2017, Art. no. 88.
[82] K. K. Rachuri, C. Mascolo, M. Musolesi, and P. J. Rentfrow, "Sociablesense: Exploring the trade-offs of adaptive sampling and computation offloading for social sensing," in *Proc. 17th Annu. Int. Conf. Mobile Comput. Netw.*, 2011, pp. 73–84.
[83] P. Zhou, Y. Zheng, and M. Li, "How long to wait? predicting bus arrival time with mobile phone based participatory sensing," *IEEE Trans. Mobile Comput.*, vol. 13, no. 6, pp. 1228–1241, Jun. 2014.
[84] W. Liu, Y. Yang, E. Wang, and J. Wu, "Fine-grained urban prediction via sparse mobile crowdsensing," in *Proc. IEEE 17th Int. Conf. Mobile Ad Hoc Sensor Syst.*, 2020, pp. 265–273.

SHITING ZHAO is currently working toward the master's degree with the College of Information Science and Technology, Jinan University, Guangzhou, China. Her research focuses on mobile crowdsensing systems.

GUOZI QI is currently working toward the master's degree at the College of Information Science and Technology of Jinan University. Her research interest includes mobile crowdsensing systems.

TENGJIAO HE received the Bachelor of Engineering degree (with Hons.) in computer engineering from the University of Wollongong, NSW, Australia, and Zhengzhou University, Zhengzhou, China, in 2013, and the Ph.D. degree from the University of Wollongong in 2017. He is currently a Lecturer with the College of Information Science and Technology, Jinan University, Guangzhou, China. His research focuses on resource management issues in rechargeable wireless sensor networks.

JINPENG CHEN is currently an Associate Professor with the School of Computer Science (National Pilot Software Engineering School), Beijing University of Posts and Telecommunications, Beijing, China. His research interests include social network analysis, recommendation system, data mining, and machine learning.

ZHIQUAN LIU received the B.S. degree from the School of Science, Xidian University, Xi'an, China, in 2012, and the Ph.D. degree from the School of Computer Science and Technology, Xidian University in 2017. He is currently an Associate Professor with the College of Cyber Security, Jinan University, Guangzhou, China. His current research interests include trust management and privacy preservation in vehicular networks and UAV networks.

KAIMIN WEI received the Ph.D. degree in computer science from Beihang University, Beijing, China. He is currently an Associate Professor with the College of Information Science and Technology, Jinan University, Guangzhou, China. His primary research interests include mobile computing and artificial intelligence, with focus on algorithm optimization and security technology in these areas.