ISSN 2007-9737

A Survey on Sparse Mobile Crowdsensing: Functionalities and Research Issues

Wangjam Niranjan Singh*, Rosy Swami, Tapodhir Acharjee, Abhijit Biswas

Assam University,
Department of Computer Science and Engineering,
India

{niranwang, dam.rosy, tapacharjee, abhi.021983}@gmail.com

Abstract. Mobile Crowdsensing (MCS), also known as participatory sensing, refers to the process of utilizing the collective power of mobile devices (such as smartphones and wearables) to collect and share data about the environment, user behavior, and other phenomena. MCS leverages the ubiquity of mobile devices and their embedded sensors to create dynamic and large-scale sensing networks. Participants willingly contribute data from their devices, which are often equipped with sensors like Global Positioning System (GPS), cameras, accelerometers, and microphones. The collected data is aggregated and analyzed to gain insights into various aspects of urban environments, healthcare, transportation, and more. MCS offers real-time data collection, wide geographical coverage, and the potential for crowd-driven insights. Sparse Mobile Crowdsensing (SMCS) is a specialized variant of MCS that deals with the challenges posed by intermittent and irregular data collection from mobile devices. In SMCS scenarios, participants contribute data sporadically due to factors like device availability, user engagement, or incentives. The sporadic nature of data collection leads to gaps in the dataset, requiring innovative techniques for data imputation, prediction, and analysis. SMCS aims to overcome the limitations of sparse data to still provide valuable insights and meaningful applications. This paper provides a comprehensive and detailed study of the MCS and SMCS framework, the relevant research work carried out in sparse mobile crowdsensing, and at the end, we provide open research problems with the help of future research in this domain.

Keywords. Mobile crowdsensing, sparse mobile crowdsensing, features, functionalities, research issues.

1 Introduction

In today's interconnected world, smartphones and mobile devices have become an integral part of our lives, serving as more than just communication tools. These de-vices are equipped with an array of sensors that can capture data about our environment, movements, and interactions.

Mobile crowdsensing harnesses this collective power by enlisting a multitude of individuals to act as distributed sensors, revolutionizing data collection and analysis. This paradigm offers unique advantages over traditional wireless sensor networks (WSNs), which are built using static sensor nodes. Mobile crowdsensing capitalizes on the widespread adoption of smartphones to create a dynamic and flexible network of data collectors. Instead of relying solely on specialized sensor nodes, mobile crowdsensing taps into the sensors embedded within smartphones, such as GPS, cameras, accelerometers, and microphones. This empowers individuals to contribute real-time data about their surroundings, enabling applications ranging from urban planning and environmental monitoring to healthcare research and disaster response [1].

Mobile crowdsensing has various advantages when compared to traditional wireless sensor networks. Unlike static sensor nodes in traditional WSNs, the "sensors" in mobile crowdsensing are carried by people who move through diverse environments. This mobility allows for data collection across wider geographic areas and varying contexts.

Mobile crowdsensing leverages the sheer number of smartphones in use, making it inherently scalable. Traditional WSNs often require careful deployment and management of sensor nodes, limiting their scalability. Mobile devices come with a rich variety of sensors, enabling the collection of diverse data types beyond what traditional WSNs can offer. This includes image, audio, and video data, enriching the information collected. Mobile crowdsensing networks form dynamically based on the movements and interactions of users. Traditional WSNs are usually designed with a fixed topology, which can be harder to adapt to changing conditions. Mobile crowdsensing places users at the center of data collection. Participants willingly contribute data from their personal devices, which requires addressing privacy concerns and ensuring proper incentives for participation. Mobile crowdsensing excels at providing real-time data due to the continuous movement of users.

Traditional WSNs might have delays due to factors like energy management and infrequent data aggregation. While traditional WSNs often optimize for energy efficiency due to limited power resources, mobile crowdsensing must consider the energy im-pact on participants' devices, potentially affecting user experience. Traditional WSNs can provide highly controlled environments for data collection, ensuring data quality. In mobile crowdsensing, data quality control becomes more complex due to the dynamic and diverse nature of contributors. Mobile crowdsensing revolves around harnessing the collective power of a large number of individuals' mobile devices to gather data about their surroundings and then utilize this data for various applications. Mobile crowdsensing treats smartphones and other mobile devices as distributed sensors. These devices are equipped with a variety of sensors, such as GPS, cameras, accelerometers, and microphones, which can capture information about the environment, user behaviors, and other relevant phenomena [2].

In mobile crowdsensing, individuals voluntarily contribute their device-generated data to a shared pool. This data can include location information, images, audio recordings, sensor readings, and more. User participation is a key aspect of crowdsensing, and proper incentives are often provided to encourage involvement. The data collected from multiple users' devices is aggregated to create a comprehensive and holistic view of the phenomenon under consideration. Aggregation can happen in real-time or periodically, and it involves combining data from various sources to generate meaningful insights. Mobile crowdsensing networks are dynamic in nature, as they form based on the movements and interactions of users. These networks can expand or contract depending on the density of participants in different areas. Different types of data collected from various sensors and devices can be fused to create a more accurate and complete understanding of the environment. Data fusion techniques help mitigate errors and enhance the overall quality of the collected information. Since mobile crowdsensing involves collecting data from individuals' personal devices, privacy and security are paramount.

Techniques like data anonymization, encryption, and access control are employed to protect users' sensitive information. Ensuring the quality and reliability of the collected data is a significant challenge in mobile crowdsensing. Mechanisms for data validation, error detection, and outlier removal are implemented to maintain data accuracy. To encourage user participation, incentive mechanisms are often employed. These incentives could include monetary rewards, gamification, reputation systems, or simply contributing to a cause that users care about. Mobile crowdsensing can be applied across various domains, including urban planning, transportation management, environmental monitoring, disaster response, healthcare research, and more. The collected data can provide valuable insights and support informed decision-making in these areas. Since mobile crowdsensing involves continuous data collection on users' devices, energy consumption is a crucial concern.

Balancing the need for data collection with the impact on device battery life is essential to ensure a positive user experience. The collected data is analyzed using various data processing techniques, such as machine learning, data mining, and statistical analysis, to extract meaningful pat-terns, trends, and insights. The interpretation of the data is often domain-specific and depends on the application's objectives. In essence, the basic theory of mobile crowdsensing revolves around leveraging the ubiquity of mobile devices, the willingness of individuals to contribute data, and the potential for meaningful insights that can be derived from the aggregated information. This approach has the power to transform data collection and analysis across a wide range of

ISSN 2007-9737

fields, offering valuable benefits to society, research, and decision-making processes.

The main contributions of this paper are outlined as:

– Basic concept of MCS and various domains where MCS finds its application.

– The challenges faced in MCS that led to the development of SMCS.

– Some of the common characteristics shared by MCS and SMCS and also the unique features of SMCS as well as a comparative study on the various features of MCS and SMCS.

– Key functionalities and research issues identified for the various stages in the life cycle of MCS and SMCS.

– Key techniques used in SMCS and the advancements in the area of SMCS research.

– And finally, the open research problems required to be addressed.

The rest of the article is organized as follows: we provide the various applications of mobile crowdsensing from the literature in Section 2; we identify the need for development of sparse mobile crowdsensing from the broader paradigm of mobile crowdsensing in Section 3; we provide the common features of MCS and SMCS as well as the unique features of SMCS in Section 4; we provide the key functionalities and research issues in each phase of MCS and SMCS life cycle in Section 5; we provide a detailed study of the key techniques and advancements in SMCS research in Section 6; and finally, we conclude the survey in Section 7.

2 Applications of MCS

Mobile crowdsensing finds application in various domains where real-time data collection from a large number of individuals can provide valuable insights.

Table 1. Some notable applications of MCS

Application	Description
Urban Planning and Management	Traffic congestion, air quality, noise levels, public transport usage
Environmental Monitoring	Air quality, temperature, humidity, pollution levels [3]
Traffic Monitoring and Management	Traffic patterns, congestion points, road conditions [4]
Healthcare and Wellbeing	Physical activity, sleep patterns, health metrics
Disaster Response and Management	Real-time information during emergencies [5]
Civic Engagement and Participatory Sensing	Reporting potholes, litter, urban concerns
Social Sciences Research	Human behavior, social interactions, cultural trends [6]
Tourism and Cultural Heritage	Tourist movements, personalized recommendations, managing visitor flows [7]
Retail and Marketing Insights	Foot traffic patterns, store layouts, marketing strategies [8]
Smart Energy Management	Energy consumption patterns, energy efficiency
Wildlife Monitoring and Conservation	Tracking animal populations, migratory patterns, habitat changes
Agriculture and Precision Farming	Soil conditions, crop growth, pest infestations [9]
Air Quality Management	Pollution levels, health risks [10]
Public Safety and Crime Prevention	Reporting incidents, suspicious activities, safety concerns [11]

These applications highlight the versatility and potential of mobile crowdsensing to provide insights, inform decision-making, and create innovative solutions across various sectors. As technology continues to advance, the scope of mobile crowdsensing is likely to expand even further.

3 Sparse Mobile Crowdsensing

Sparse mobile crowdsensing refers to a situation where the data collection process is not continuous or uniform but rather sporadic and irregular. In this approach, only specific users or devices contribute data at certain times or locations, which creates gaps or sparsity in the collected data. The development of sparse mobile crowdsensing is an ongoing area of research and innovation, as it requires addressing both technical and user-related challenges. As technology advances, more efficient algorithms, improved energy-saving strategies, and better user engagement tactics are likely to emerge, enhancing the overall effectiveness of sparse mobile crowdsensing approaches [12].

The development of sparse mobile crowdsensing from the broader mobile crowdsensing paradigm involves several considerations. Mobile devices have limited battery life. In sparse mobile crowdsensing, it is important to optimize data collection schedules and transmission protocols to minimize energy consumption while still capturing valuable information. Since data collection is not continuous, ensuring the quality of the collected data becomes crucial. Techniques to verify and validate data, as well as incentivizing users to provide accurate information during their limited participation, are essential. Task assignment to users becomes more challenging in sparse crowdsensing scenarios. Efficient algorithms are needed to match tasks with available users, taking into account their locations, capabilities, and willingness to participate.

Encouraging users to participate in sparse crowdsensing requires effective incentive mechanisms. Gamification, monetary rewards, social recognition, or providing useful insights derived from the collected data can motivate users

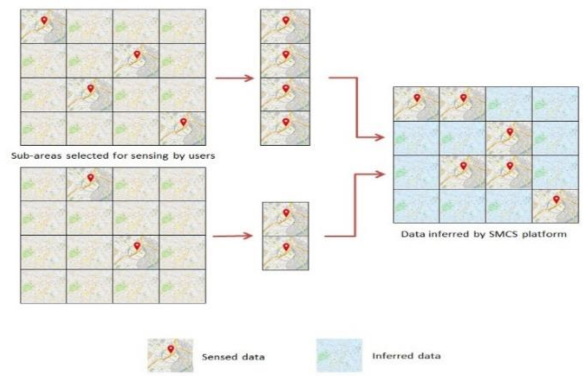

Fig. 1. Sparse Mobile Crowdsensing (SMCS) framework

to contribute despite the sporadic nature of the tasks [13]. Understanding the spatio-temporal patterns of user participation is the key in sparse mobile crowdsensing. Analyzing when and where users are more likely to contribute data can help design better data collection strategies. Advanced machine learning and AI techniques can be employed to predict when and where data is needed the most. Predictive models can help guide data collection efforts and prioritize certain areas or times for better coverage [14]. Due to the sparsity of data, techniques like data fusion and interpolation are used to estimate values in regions or time intervals where data is missing. These techniques help create a more complete and continuous dataset for analysis. Sparse mobile crowdsensing must address privacy concerns, as data might be collected from users' personal devices. Implementing privacy-preserving techniques and ensuring secure data transmission are critical. Figure 1 gives a layout of the SMCS framework.

4 Features of MCS and SMCS

Mobile crowdsensing and sparse mobile crowdsensing share many common characteristics, but sparse mobile crowdsensing introduces certain unique features due to its intermittent and irregular nature. Some of the common characteristics shared by both concepts are summarized below:

Table 2. Comparative study of MCS and SMCS

Paradigm	MCS	SMCS
Participation	Massive	Medium
Sensed areas	High	Low
Cost of sensing	High	Low
Spatio-temporal Correlation	Less Considered	Highly considered
Data Sources	Sensing Data	Sensing Data and Inferred Data

- **Utilization of Mobile Devices:** Both mobile crowdsensing and sparse mobile crowdsensing leverage the widespread availability of smartphones, tablets, wearables, and other mobile devices to collect and transmit data.
- **User Participation:** In both cases, users actively participate by contributing data using their mobile devices. This can involve sharing sensor readings, photos, videos, location information, and more.
- **Data Variety:** Both concepts allow for the collection of diverse types of data, such as environmental measurements, traffic patterns, noise levels, and other contextual information.
- **Data Transmission:** Data collected by participants is transmitted to a central server or platform for processing, analysis, and storage.
- **Scalability:** Both mobile crowdsensing and sparse mobile crowdsensing can scale up to a large number of participants, enabling the collection of data from various geographic areas and time periods.
- **Real-Time Capabilities:** Both approaches can offer real-time or near-real-time data collection and analysis, providing timely insights.
- **Community Engagement:** Both concepts can engage communities and individuals in data-driven decision-making processes for urban planning, environmental management, disaster response, and more. While both mobile crowdsensing and sparse mobile crowdsensing share foundational characteristics, sparse mobile crowdsensing utilizes the spatio-temporal correlation to gather the missing data from the data collected from different sensing sub-areas. Some of the unique features of SMCS are:
- **Intermittent Participation:** Sparse mobile crowdsensing involves sporadic and irregular user participation. Users contribute data only at certain times or locations, leading to gaps in the data collection timeline.
- **Data Sparsity:** As a result of intermittent participation, there are gaps in the collected data, making the dataset more sparse compared to continuous data collection.
- **Task Dynamics:** Task assignment and data collection schedules can be more challenging in sparse mobile crowdsensing due to the unpredictable nature of user participation.
- **Energy Efficiency Focus:** Given the sporadic nature of data collection, energy efficiency becomes a critical concern in sparse mobile crowdsensing to ensure that users' devices are not overly burdened by participating in data collection tasks.
- **Data Fusion and Interpolation:** Techniques like data fusion and interpolation are commonly used to fill in gaps in the collected data and provide a more complete dataset for analysis.
- **Incentive Mechanisms:** Incentivizing users to participate in sparse crowdsensing is especially important due to the irregularity of their involvement. Creative incentive mechanisms are needed to motivate participation.
- **Predictive Modeling:** Predictive models may be developed to anticipate when and where data is needed the most, aiding in optimizing task assignments and data collection efforts.
- **Privacy Concerns:** While privacy is a concern in both concepts, sparse mobile crowdsensing might need to address privacy issues more strategically due to the intermittent nature of

ISSN 2007-9737

data collection, ensuring that users' privacy is maintained even during their infrequent participation. Table 2 presents a comparative study of MCS and SMCS.

5 Key Functionalities and Research Issues of MCS and SMCS

The life cycle of MCS and SMCS involves several stages, from initialization to data collection, processing, and analysis. Table 3 presents an overview of the life cycle for both concepts.

The research issues mentioned for both MCS and SMCS highlight the challenges and opportunities at each phase of the life cycle. These issues often require interdisciplinary collaboration, involving aspects of computer science, human-computer interaction, data analysis, privacy, and domain-specific knowledge. Addressing these issues contributes to the successful implementation and utilization of both MCS and SMCS paradigms.

6 Key Techniques and Advancements in SMCS Research

Sparse Mobile Crowdsensing (SMCS) faces challenges in both sensing area selection and inferring missing data. Determining where and when to activate data collection tasks in SMCS is complex due to participants' intermittent contributions. Selecting the right sensing areas is crucial to ensure meaningful insights and coverage in data-sparse regions. Inaccurate sensing area selection can lead to biased or in-complete data, affecting the validity of analyses and decision-making based on the collected data. SMCS data often contains gaps and missing values due to participants' irregular contributions. Inferring the missing data accurately is challenging, especially when traditional interpolation or imputation methods may not be directly applicable. Accurate inference of missing data is essential for maintaining data continuity and ensuring that analyses and applications are based on as complete and representative a dataset as possible.

6.1 Sensing Area Selection

Sensing area selection in SMCS involves determining where and when to activate data collection tasks based on participant availability, task priorities, and the areas of interest. Since data collection is intermittent, selecting the right sensing areas becomes crucial to maximize the utility of collected data. Different factors and re-search frameworks have been considered in existing studies in selection of the sensing areas, such as characteristics of participants in sensing task, cost of sensing, type of sensing task, urban scenarios, and so on. Table 4 presents an outline of the relevant research work on sensing area selection. The challenges in sensing area selection are given below:

– **Data Sparsity:** Irregular data contributions lead to gaps in the collected dataset, making it challenging to ensure sufficient coverage in sensing areas.
– **Uncertain Availability:** Participants' availability for data collection varies, making it difficult to predict when and where data will be contributed.
– **Task Prioritization:** Assigning priorities to different data collection tasks be-comes crucial to focus efforts on areas of higher importance.
– **Contextual Variability:** Participant context, such as location and activity, constantly changes, requiring real-time adaptability in sensing area selection.
– **Privacy Concerns:** Balancing the need for accurate sensing with participants' privacy concerns requires careful consideration.

Approaches to Sensing Area Selection:

– **Predictive Modeling:** Develop predictive models based on historical data to estimate when and where participants are likely to contribute. Machine learning algorithms can capture patterns in participants' behaviors and predict their future contributions.
– **Context-Aware Algorithms:** Leverage contextual information, such as participants' locations, schedules, and routines, to make informed decisions about optimal sensing areas.

ISSN 2007-9737

Table 3. Overview of life cycle of MCS and SMCS

Section	Key Functionalities	Research Issues
Problem Definition & Task Design	- Define problem or task. - Specify objectives, data types, desired outcomes.	- **Task formulation:** Designing meaningful tasks. - **Incentive mechanisms:** Motivating participants.
Data Collection Setup	- Develop apps/platforms. - Configure sensors and parameters.	- **App usability:** Designing user-friendly apps. - **Sensor calibration:** Ensuring accurate readings.
Data Collection	- Collect data from devices.	- **Data accuracy:** Addressing inaccuracies. - **Data completeness:** Encouraging comprehensive data.
Data Transmission & Aggregation	- Transmit data to a server. - Aggregate data from multiple participants.	- **Data transmission efficiency:** Optimizing battery use. - **Data fusion:** Combining diverse data types.
Data Quality Control	- Detect and correct erroneous data.	- **Anomaly detection:** Identifying outliers. - **Data validation:** Ensuring quality and reliability.
Data Processing & Analysis	- Process aggregated data using analysis techniques.	- **Data analytics methods:** Choosing suitable techniques. - **Real-time analysis:** Performing meaningful analysis.
Application & Decision-Making	- Apply insights to specific problems.	- Actionable insights: Ensuring effective decision-making. - Use case adaptation: Adapting insights.
Feedback & Improvement	- Gather participant feedback for enhancement.	- **Feedback mechanisms:** Collecting constructive feedback. - **Iterative improvement:** Refining the process.
Continuous Monitoring & Evolution	- Continuously monitor and evolve the process.	- **Process optimization**: Adapting to changes. - **Scalability:** Handling growth in participants and data.

Task Coordination Algorithms: Algorithms can coordinate data collection tasks to maximize coverage while considering participants' availability, task priorities, and historical patterns.

– **Dynamic Heatmaps:** Generate dynamic heatmaps that visualize areas with high participant density or data contributions, aiding in identifying suitable sensing areas.

– **Incentive Mechanisms:** Design effective incentive structures that encourage participants to contribute data from specific areas or at specific times.

– **Geo-Fencing and Triggers:** Implement geo-fencing mechanisms that trigger data collection when participants enter predefined areas of interest.\

Table 4. A summary of pertinent research on the choice of sensing area

Sl. No	Ref. No	Research Framework	Aim	Approach/ Method	Dataset
1.	[15]	Multiple sensing costs for SMCS	To maximize the information gleaned from the gathered data while minimizing the overall cost of the sample	Greedy approach and Pareto optimization selection	Parking, Flow, Traffic, Humidity
2.	[16]	Sensing area diversity and malevolent participants	To preserve the quality of sensed data while minimizing sensing costs	Optimization selection	Taxispeed
3.	[17]	Online task distribution with quality assurance	To guarantee the accuracy of sensed data and to lower the necessary number of assigned tasks	Query-By-Committee	Temperature, Air quality
4.	[18]	Cost-efficient distribution of work	To guarantee the accuracy of sensed data and to lower the necessary number of assigned tasks	Query-By-Committee	Temperature, Humidity, Air quality, Taxispeed
5.	[19]	Matrix completion as the foundation for SMCS	To lower the cost of sensing and guarantee the accuracy of missing data	Sampling scheduling algorithm	Traffic, Air quality
6.	[20]	Multimodal Urban Sensing	To deal with the choice of sensing region for various task situations	Reinforcement Learning	Temperature, Humidity, Air quality
7.	[21]	Task user recruitment	To improve the inferred data's accuracy	Reinforcement Learning	Temperature, Humidity, Air quality, Taxispeed
8.	[22]	Selection of sensing area	To guarantee the accuracy of sensed data	Reinforcement Learning	Temperature, Humidity, Air quality
9.	[23]	Selection of sensing area	To lower the price of data sensing while raising the calibre of data sensed	Deep Reinforcement Learning	Temperature, Humidity, Air quality
10.	[24]	Time-critical metropolitan environment	To maintain current selection model sensing area	Reinforcement Learning	Air quality
11.	[25]	Guaranteed quality gathering of sparse data	To cut down on unnecessary information and choose the right user base	Mobile Edge Computing	Air quality, Taxispeed
12.	[26]	Participants not interested to undertake the sensing task in critical areas	To lower the price of data sensing	Subarea Division Learning	Temperature, Humidity, Air quality

– **Feedback Loop:** Continuously gather feedback from participants about their preferences and contribute to decisions regarding sensing area selection.

– **Hybrid Approaches:** Combine multiple techniques, such as predictive modeling and context-aware algorithms, to enhance the accuracy of sensing area selection.

– **Privacy-Preserving Techniques:** Employ techniques that respect participants' privacy while still providing accurate location-based data.

Table 5. A summary of pertinent research on inferring missing data

Sl. No.	Ref. No	Research Framework	Aim	Approach/ Method	Dataset
13.	[27]	Traffic conditions on the roads	To infer the missing data	Compressed Sensing	Traffic
14.	[28]	Real Wireless Sensor Network architecture	To utilize wireless sensor networks to gather and compress large, dispersed signals.	Compressed Sensing	Self-collected data
15.	[29]	Metropolitan Noise Map	To reconstruct the noise map from randomly selected and insufficient samples	Compressed Sensing	Self-collected data
16.	[30]	MCS architecture	To reduce the excess load on the users	Compressed Sensing	Noise complaints
17.	[31]	Monitoring the state of the urban air quality	To create a low-cost sensing incentive system that targets users and works well	Compressed Sensing	Air quality
18.	[32]	Monitoring the state of the urban air quality	To create a low-cost sensing incentive system that targets users and works well	Compressed Sensing	Air quality
19.	[33]	Configuring Signal mappings	To keep sensing costs and signal quality in balance	Bayesian Compressed Sensing	Signal map
20.	[19]	Extensive MCS deployment	To retrieve the unsampled data and lower the sensing data acquisition cost	Matrix completion	Traffic, Air quality
21.	[34]	Wireless Sensor Networks	To deduce the missing data	kNN algorithm	Temperature, Humidity, Voltage
22.	[35]	Environmental Crowd Sensing	To deduce the missing data	kNN algorithm	Temperature, Air quality
23.	[36]	The Internet of things	To deduce the missing data in IoT	Gaussian mixture model	Temperature
24.	[37]	Use high-bandwidth technology to infer missing data	To deduce the missing data	Random forest model algorithm	SPECT, Promoter, Lymphography
25.	[38]	Signal strength maps	To improvise maps of signal strength	Random forest model algorithm	Signal map

– **Participant Engagement Strategies:** Involve participants in the decision-making process, allowing them to express their preferences for sensing areas and influencing data collection efforts.

6.2 Missing Data Interference

Missing data inference in Sparse Mobile Crowdsensing (SMCS) refers to the process of estimating and predicting data values for the periods or locations where data is not collected due to participants' sporadic contributions. Given the intermittent nature of data collection in SMCS, missing data inference becomes crucial for maintaining data continuity and ensuring the usability of collected data. Table 5 presents an outline of the relevant research work on missing data inference. The challenges in missing data interference are given below:

– **Irregular Sampling:** SMCS data might not follow a regular time interval or spatial grid, making traditional interpolation methods less applicable.

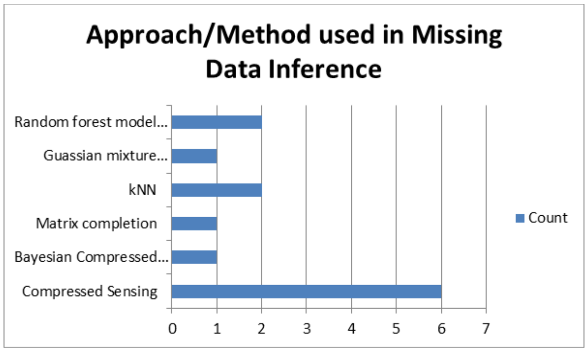

Fig. 3. Comparison on Approach/Method used in Missing Data Inference

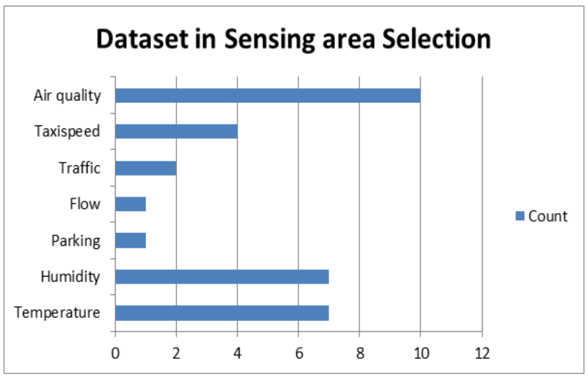

Fig. 4. Comparison on Datasets used in Sensing Area Selection

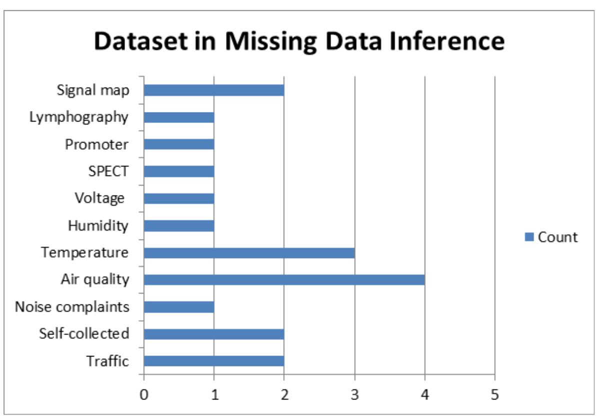

Fig. 5. Comparison on Datasets used in Missing Data Inference

– **Uncertainty:** Predicting missing data involves dealing with uncertainty due to the sporadic nature of contributions and the potential for various external factors affecting data availability.

– **Diverse Data Types:** SMCS data can include various sensor readings (e.g., location, temperature, images), each requiring specific techniques for accurate inference.

Approaches to Missing Data Inference:

– **Temporal and Spatial Interpolation:** These methods estimate missing data points based on neighboring observed data points. However, due to irregular data sampling, specialized interpolation techniques that account for variable time intervals or spatial distances might be needed.

– **Time-Series Forecasting**: Predictive modeling using time-series forecasting techniques (such as ARIMA, LSTM, or Prophet) can help predict future data points based on historical patterns. These models capture trends and seasonality, providing accurate predictions even with missing data.

– **Contextual Information:** Utilizing contextual information like user behaviors, location history, and environmental factors can enhance the accuracy of missing data predictions. For example, a user's previous patterns might help predict their future contributions.

– **Machine Learning:** Data-driven machine learning models can learn patterns from available data and use them to infer missing values. These models can handle complex relationships and interactions among different data attributes.

– **Hybrid Approaches:** Combining multiple techniques, such as leveraging both temporal patterns and contextual information, can yield more accurate predictions, especially in scenarios with sparse data.

– **Probabilistic Modeling:** Bayesian methods or probabilistic models can capture uncertainty in missing data inference, providing not only point estimates but also confidence intervals.

Figure 2 presents a graphical representation comparing the approaches/methods used in sensing area selection. Among the relevant work presented in the sensing area selection problem, reinforcement learning method has been used more compared to the other methods. Figure 3 shows that compressed sensing method has been used more missing data inference problem compared to the other methods, among the related papers presented.

Figure 4 shows the graphical comparison on the dataset used by the researchers in the relevant papers in sensing area selection problem. Figure 5 shows the graphical comparison on the dataset used by the researchers in the relevant papers in missing data inference problem.

Some of the open research problems in the domain of SMCS are:

– Developing more accurate and efficient techniques for filling in missing data gaps in SMCS, especially when data is sporadic, remains a significant challenge.
– Finding innovative ways to protect participants' privacy while collecting and sharing data in SMCS remains a top concern. Balancing the need for data accuracy with privacy is an ongoing challenge.
– Designing algorithms and strategies for scheduling data collection tasks to maximize data coverage while considering participant availability, energy constraints, and task priorities is a complex optimization problem.
– Designing effective and fair incentive mechanisms that motivate participants to contribute data, even when contributions are sporadic, is an open area of research. This involves finding the right balance between rewards and user engagement.
– Investigating how to seamlessly integrate data from diverse sensing modalities, including mobile devices, traditional sensors, and remote sensing technologies, is an open research problem.

7 Conclusion

Sparse Mobile Crowdsensing (SMCS) is a variant of Mobile Crowdsensing where data contributions are sporadic and irregular, often due to participant availability or specific task requirements. Unlike traditional MCS, which assumes continuous data collection and sufficient participant availability, SMCS focuses on balancing the cost and quality of sensed data, addressing real-world challenges such as data delays, sensor malfunctions, and noise. Ensuring data quality in SMCS is crucial for meaningful insights, with recent research leveraging deep learning techniques to overcome these challenges.

References

1. **Zhan, Y., Xia, Y., Zhang, J. (2008).** Quality-Aware Incentive Mechanism Based on Payoff Maximization for Mobile Crowdsensing. Ad Hoc Networks, Vol. 72, pp. 44–55.
2. **Lin, Y., Cai, Z., Wang, X., Hao, F., L. Wang, Sai, A.M.V.V. (2021).** Multi-Round Incentive Mechanism for Cold Start-Enabled Mobile Crowdsensing. IEEE Transactions on Vehicular Technology, Vol. 70, No. 1, pp. 993–1007. doi:10.1109/TVT.2021.3050339.
3. **Dutta, P., Aoki, P.M., Kumar, N. (2009).** Common Sense: Participatory Urban Sensing Using a Network of Handheld Air Quality Monitors. Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems, pp 349–350. doi: 10.1145/1644038.1644095.
4. **Eriksson, J., Girod, L., Hull, B., Newton, R., Madden, S., Balakrishnan, H. (2008).** The Pothole Patrol: Using Mobile Sensor Network Road Surface Monitoring. Proceedings 6th International Conference Mobile Systems Applications Services - MobiSys'08, Vol. 6, pp. 29–39. doi: 10.1145/137 8600.1378605.
5. **Bengtsson, L., Lu, X., Thorson, A., Garfield, R., von Schreeb, J. (2011).** Improved Response Disasters Outbreaks Tracking Population Movements Mobile Phone Network Data: Post-Earthquake Geospatial Study Haiti. PLoS Medicine, Vol. 8, No. 8, pp. 100–1083. doi: 10.1371/journal. pmed.1001083.
6. **Mathur, S., Jin, T., Kasturirangan, N. (2010).** ParkNet: Drive-By Sensing Road-Side Parking Statistics. Proceedings 8th International Conference Mobile Systems Applications Services - MobiSys '10, Vol. 8, pp. 123–136. doi: 10.1145/1814433.1814448.

ISSN 2007-9737

7. **Wang, N., Zhao, X., Wang, L., Zou, Z. (2019).** Novel System Rapid Investigation Damage Detection Cultural Heritage Conservation Based Deep Learning. Journal of Infrastructure Systems, Vol. 25, No. 3. doi: 10.1061/(ASCE)IS.1943-555X. 0000499.
8. **Marakkalage, S.H., Sarica, S., Lau, B.P.L., Viswanath, S.K., Balasubramaniam, T., Yuen, C., Yuen, B., Luo, J., Nayak, R. (2018).** Understanding Lifestyle Older Population: Mobile Crowdsensing Approach. IEEE Transactions on Computational Social Systems, Vol. 6, No. 1, pp. 82–95. doi: 10.1109/TCSS.2018.2883691.
9. **Sun, Y., Ding, W., Shu, L., Li, K., Zhang, Y., Zhou, Z., Han, G. (2021).** Enabling Mobile Crowd Sensing Data Collection Smart Agriculture: Vision. IEEE Systems Journal, Vol. 16, No. 1, pp. 132–143. doi: 10.1109/JSYST.2021.3104107.
10. **Leonardi, C., Cappellotto, A., Caraviello, M., Lepri, B., Antonelli, F. (2014).** SecondNose: Air Quality Mobile Crowdsensing System. Proceedings 8th Nordic Conference Human-Computer Interaction: Fun Fast Foundational, Vol. 8, pp. 1051–1054.
11. **Gao, R., Zhao, M., Ye, T. (2014).** Jigsaw: Indoor Floor Plan Reconstruction Mobile Crowdsensing. Proceedings 20th Annual International Conference Mobile Computing Networking, Vol. 20, pp. 249–260.
12. **Goel, U., Mongia, K., Gupta, Q., Rajput, H., Jha, V. (2022).** Sparse Mobile Crowdsensing: Components Frameworks. Proceedings 2022 IEEE World AI IoT Congress, pp. 773–778. doi: 10.1109/aiiot54504.2022.9817212.
13. **Zhang, X., Yang, Z., Sun, W., Liu, Y., Tang, S., Xing, K., Mao, X. (2015).** Incentives Mobile Crowd Sensing: Survey. IEEE Communications Surveys & Tutorials, Vol. 18, No. 1, pp. 54–67. doi: 10.1109/COMST. 2015.2415528.
14. **Lai, C., Zhang, X. (2020).** Duration-Sensitive Task Allocation Mobile Crowd Sensing. IEEE Systems Journal, Vol. 14, No. 3, pp. 4430–4441. doi: 10.1109/JSYST.2020.2967847.
15. **Zhu, Z., Chen, B., Liu, W., Zhao, Y., Liu, Z., Zhao, Z. (2021).** 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.
16. **Sun, P., Wang, Z., Wu, L., Shao, H., Qi, H., Wang, Z. (2021).** Trustworthy Cost-Effective Cell Selection Sparse Mobile Crowdsensing Systems. IEEE Transactions on Vehicular Technology, Vol. 70, No. 6, pp. 6108–6121. doi: 10.1109/TVT.2021.3077112.
17. **Wang, L. (2015).** CCS-TA: Quality-Guaranteed Online Task Allocation Compressive Crowdsensing. Proceedings 2015 ACM International Joint Conference Pervasive Ubiquitous Computing, pp. 683–694. doi: 10.1145/2750858.2807513.
18. **Wang, L. (2017).** SPACE-TA: Cost-Effective Task Allocation Exploiting Intradata Interdata Correlations Sparse Crowdsensing. ACM Transactions on Intelligent Systems Technology, Vol. 9, No. 2, pp. 1–28. doi: 10.1145/3131671.
19. **Xie, K., Li, X., Wang, X., Xie, G., Wen, J., Zhang, D. (2019).** Active Sparse Mobile Crowd Sensing Based Matrix Completion. Proceedings International Conference Management Data, pp. 195–210. doi: 10.1145/3299869.3319856.
20. **Liu, W., Yang, Y., Wang, E., Wang, L., Zeghlache, D., Zhang, D. (2019).** Multi-Dimensional Urban Sensing Sparse Mobile Crowdsensing. IEEE Access, Vol. 7, pp. 82066–82079.
21. **Liu, W., Wang, L., Wang, E., Yang, Y., Zeghlache, D., Zhang, D. (2019).** Reinforcement Learning-Based Cell Selection Sparse Mobile Crowdsensing. Computer Networks, Vol. 161, pp. 102–114.
22. **Liu, W., Yang, Y., Wang, E., Wu, J. (2020).** User Recruitment Enhancing Data Inference Accuracy Sparse Mobile Crowdsensing. IEEE Internet of Things Journal, Vol. 7, No. 3, pp. 1802–1814.
23. **Wang, L., Liu, W., Zhang, D., Wang, Y., Wang, E., Yang, Y. (2018).** Cell Selection Deep Reinforcement Learning Sparse Mobile Crowdsensing. Proceedings 38th International

Conference Distributed Computing Systems, Vol. 38, pp. 1543–1546.

24. **Han, L., Yu, Z., Wang, L., Yu, Z., Guo, B. (2021).** Keeping Cell Selection Model Up-To-Date Adapt Time-Dependent Environment Sparse Mobile Crowdsensing. IEEE Internet of Things Journal, Vol. 8, No. 18, pp. 13914–13925.
25. **Xia, X., Zhou, Y., Li, J., Yu, R. (2019).** Quality-Aware Sparse Data Collection MEC-Enhanced Mobile Crowdsensing Systems. IEEE Transactions on Computational Social Systems, Vol. 6, No. 5, pp. 1051–1062.
26. **Wei, X., Li, Z., Liu, Y., Gao, S., Yue, H. (2021).** SDLSC-TA: Subarea Division Learning Based Task Allocation Sparse Mobile Crowdsensing. IEEE Transactions on Emerging Topics in Computing, Vol. 9, No. 3, pp. 1344–1358.
27. **Zhu, Y., Li, Z., Zhu, H., Li, M., Zhang, Q. (2013).** Compressive Sensing Approach Urban Traffic Estimation Probe Vehicles. IEEE Transactions on Mobile Computing, Vol. 12, No. 11, pp. 2289–2302.
28. **Quer, G., Masiero, R., Pillonetto, G., Rossi, M., Zorzi, M. (2012).** Sensing Compression Recovery WSNs: Sparse Signal Modeling Monitoring Framework. IEEE Transactions on Wireless Communications, Vol. 11, No. 10, pp. 3447–3461.
29. **Rana, R.K., Chou, C.T., Kanhere, S.S., Bulusu, N., Hu, W. (2010).** Ear-Phone: End-To-End Participatory Urban Noise Mapping System. Proceedings 9th ACM/IEEE International Conference Information Processing Sensor Networks, Vol. 9, pp. 105–116.
30. **Xu, L., Hao, X., Lane, N.D., Liu, X., Moscibroda, T. (2015).** More with Less: Lowering User Burden Mobile Crowdsourcing Compressive Sensing. Proceedings ACM International Joint Conference Pervasive Ubiquitous Computing, pp. 659–670.
31. **Liu, T., Zhu, Y., Yang, Y., Ye, F. (2019).** ALC2: Active Learning Meets Compressive Crowdsensing Urban Air Pollution Monitoring. IEEE Internet of Things Journal, Vol. 6, No. 6, pp. 9427–9438.
32. **Liu, T., Zhu, Y., Yang, Y., Ye, F. (2016).** Incentive Design Air Pollution Monitoring Compressive Crowdsensing. Proceedings IEEE Global Communications Conference, pp. 1–6.
33. **He, S., Shin, K.G. (2018).** Steering Crowdsourced Signal Map Construction Bayesian Compressive Sensing. Proceedings IEEE Conference Computer Communications, pp. 1016–1024.
34. **Pan, L., Li, J. (2010).** K-Nearest Neighbor Based Missing Data Estimation Algorithm Wireless Sensor Networks. Wireless Sensor Networks, Vol. 2, No. 2, pp. 115–122.
35. **Marchang, N., Tripathi, R. (2020).** KNN-ST: Exploiting Spatio-Temporal Correlation Missing Data Inference Environmental Crowd Sensing. IEEE Sensors Journal, Vol. 21, No. 3, pp. 3429–3436.
36. **Yan, X., Xiong, W., Hu, L., Wang, F., Zhao, K. (2015).** Missing Value Imputation Based Gaussian Mixture Model Internet Things. Mathematical Problems in Engineering, Vol. 8, pp. 548–605.
37. **Stekhoven, D.J., Bühlmann, P. (2012).** MissForest–Non-Parametric Missing Value Imputation Mixed-Type Data. Bioinformatics, Vol. 28, No. 1, pp. 112–118.
38. **Alimpertis, E., Markopoulou, A., Butts, C., Psounis, K. (2019).** City-Wide Signal Strength Maps: Prediction Random Forests. Proceedings World Wide Web Conference, pp. 2536–2542.

Article received on 14/12/2024; accepted on 28/08/2025.
**Corresponding author is Wangjam Niranjan Singh.*