Privacy-Preserving Federated Active Learning for Data Completion in Sparse Crowdsensing

Zhifei Wang[†], Wenbin Liu[*†], En Wang[*†], and Jie Wu[§,‡]
[†]College of Computer Science and Technology, Jilin University, P.R. China
[§]China Telecom Cloud Computing Research Institute, Beijing
[‡]Department of Computer and Information Sciences, Temple University, Philadelphia, USA
Email: zfwang22@mails.jlu.edu.cn,liuwb16@mails.jlu.edu.cn, wangen@jlu.edu.cn, jiewu@temple.edu

***Abstract*—Mobile CrowdSensing (MCS) relies on sparse data completion to infer missing information. However, centralized approaches require uploading raw data to a central server, which entails significant privacy risks. Moreover, under limited sampling budgets, random or fixed sampling is often ineffective at improving completion accuracy. Therefore, we proposes a privacy-preserving framework for sparse data completion based on Federated Learning (FL), which avoids raw data transmission while maintaining inference accuracy. We instantiate the framework with three models: Federated Deep Matrix Factorization (FDMF), Federated Recurrent Neural Network–enhanced Deep Matrix Factorization (FRDMF), and Federated Active Learning–enhanced Deep Matrix Factorization with Differential Privacy (FRDMF-DP-AL). To mitigate privacy risks in model updates, we add Laplace noise to the uploaded parameters to achieve ϵ-differential privacy. Experimental results indicate that the framework delivers a strong privacy–utility–efficiency trade-off, supporting city-scale MCS deployments.**

***Index Terms*—Mobile CrowdSensing (MCS), sparse data completion, federated learning, privacy preservation, active learning.**

I. Introduction

Mobile CrowdSensing (MCS) has emerged as an important paradigm for large-scale data collection, enabling applications such as air-quality monitoring and traffic prediction by aggregating measurements from participating mobile users [1]–[3]. However, due to limited sampling budgets and device limitations, MCS often yields sparse observations. In sparse MCS (SMCS) [4], missing data can be imputed using centralized completion methods (e.g., matrix factorization [5]), but two critical challenges persist: (1) privacy risks [6], because uploading raw data to a central server can expose users' locations and mobility trajectories [7]; and (2) a lack of targeted sampling strategies [8], under which random or fixed sampling often fails to prioritize informative (high-value) locations, yielding only limited improvements in completion accuracy.

Federated Learning (FL) [9] enables collaborative model training while keeping data local, but its application to SMCS still poses several challenges. First, traditional completion methods such as Deep Matrix Factorization (DMF) rely on global spatiotemporal correlations [8], whereas FL trains on client-local data; directly applying these methods often fails due to insufficient cross-client correlations [10]. Second,the limited local data per client exacerbates generalization errors (e.g., overfitting or underfitting), particularly in sparse sensing scenarios. Under constrained sampling budgets, random or fixed samplingoften induces sampling bias [11], redundant sampling [12], and failure to account for predictive uncertainty [13], leading to insufficient use of sensing resources [14] and further degrading completion accuracy. Third, transmitting gradients still entails privacy risks—adversaries can reconstruct training data through gradient inversion [15].

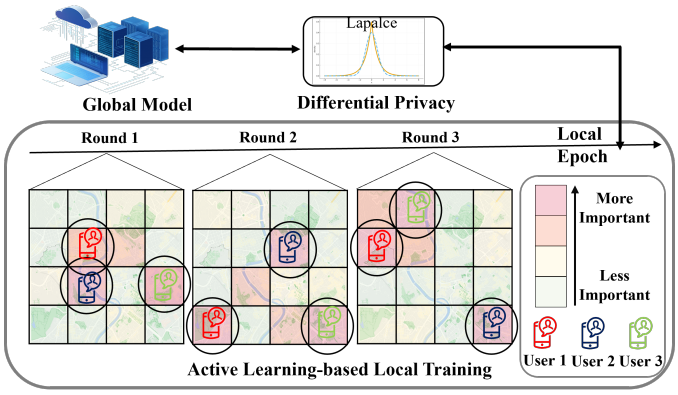


Fig. 1. Privacy-Preserving Federated Active Learning for Data Completion in Sparse Crowdsensing.

To address these challenges, we propose a federated active learning framework for data completion in SMCS (see Fig. 1). First, because FL trains on client-local data and cannot directly exploit global spatiotemporal correlations, we deploy a lightweight recurrent neural network (RNN) on each client to capture local temporal dependencies. We then integrate information globally via federated aggregation. Second, to overcome the accuracy bottleneck under tight sampling budgets, we incorporate active learning into client-side training: at the end of each local epoch, we select a small set of high-value locations based on predictive uncertainty, thereby reducing completion error at the same budget. Finally, to mitigate privacy risks in gradient sharing, we ensure ϵ-differential privacy by clipping gradients and applying the Laplace noise before upload, thereby reducing susceptibility to gradient-

Corresponding Authors: Wenbin Liu and En Wang.

inversion attacks.

Our main contributions are summarized as follows:

- We propose a federated learning framework for sparse-data completion that performs client-local spatiotemporal modeling, avoids exchanging raw data, and reduces reliance on global spatiotemporal correlations.
- We integrate a lightweight RNN into the DMF backbone to capture temporal dependencies, which improves accuracy and robustness to sparsity.
- We ensure ϵ-differential privacy by clipping client updates and injecting noise before upload, which mitigates gradient-inversion attacks; we then quantitatively analyze the privacy–utility trade-off.
- We introduce a budget-constrained active-learning scheme that prioritizes sensing of high-value samples, thereby improving sampling efficiency and accuracy without increasing communication cost or privacy risk.
- We conduct extensive experiments on four real-world datasets. The results demonstrate that, under limited sampling budgets and time constraints, our framework reduces completion error, preserves privacy, and scales efficiently.

We present an SMCS framework that integrates federated and active learning, achieves accurate data completion and strong privacy protection, and explicitly addresses limited sampling budgets.

II. Related Work

A. Data Completion

Early studies on SMCS mainly focused on centralized data completion algorithms. For example, Zhu et al. [16] proposed a compressive sensing-based framework to infer missing data from sparse observations. Subsequently, Wang et al. [17] introduced DMF, which leverages neural networks to learn latent spatiotemporal correlations. Although these methods can achieve high accuracy, they require aggregating raw data on a central server, which raises the risk of exposing sensitive user information (e.g., location trajectories in traffic datasets [18]). Several surveys [1], [19], [20] further point out that existing SMCS literature primarily emphasizes completion accuracy while neglecting privacy protection—an essential gap that this work aims to address.

B. Federated Learning and Privacy

Federated Learning (FL) enables collaborative model training without data sharing, making it highly suitable for privacy-sensitive scenarios [21], [22]. However, its application in data completion has not yet been thoroughly explored. Traditional FL frameworks (e.g., FedAvg) mainly focus on classification tasks and assume that data across different clients are independently and identically distributed (IID). In Sparse MCS, however, this assumption does not hold because sensing regions are spatially fragmented. Recent attempts to combine FL with matrix completion [10] still rely on the assumption of global correlations, which limits their practicality. It is worth noting

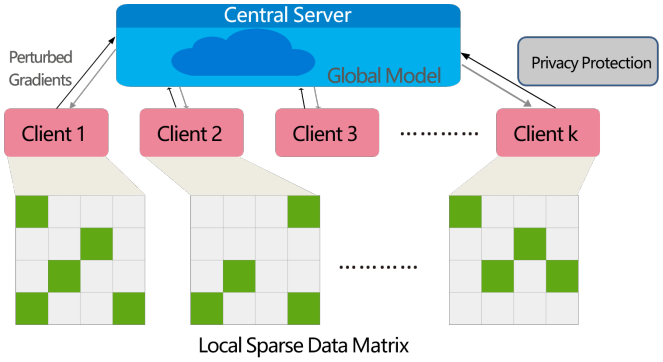


Fig. 2. Federated Learning Framework.

that RNNs have demonstrated strong potential in capturing local temporal dependencies for traffic prediction [23], yet their integration into an FL-based completion framework has not been investigated—this represents one of the key innovations of our work.

C. Active Learning

Active Learning (AL) improves performance under limited budgets by actively selecting valuable samples [24]. Its core idea is to guide sampling through model feedback, making observations more effective in enhancing the model. Early approaches adopted uncertainty sampling, prioritizing the selection of samples with the highest prediction uncertainty [25], but they tended to concentrate on local regions. Subsequently, methods based on information gain and diversity constraints introduced representativeness considerations to alleviate excessive bias [26]. With the development of deep learning, deep active learning approaches have further combined deep feature representations with Bayesian estimation, demonstrating stronger sample selection capability and adaptability in complex tasks [27]. However, these methods rely on centralized data and incur high costs, making them difficult to directly apply to distributed and budget-constrained SMCS scenarios.

D. Privacy-Preserving Techniques

Privacy preservation in distributed systems is typically achieved through cryptographic methods (e.g., homomorphic encryption [28]) or noise perturbation. Although Secure Multi-Party Computation (SMC) [29] can provide strong security guarantees, its computational overhead is often prohibitive for resource-constrained mobile devices. In contrast, Local Differential Privacy (LDP) [30] offers a lightweight alternative by perturbing client updates before transmission to protect privacy. For example, Dwork et al. [31] proved that Laplace noise injection can satisfy ϵ-differential privacy and quantify its utility loss. Building on this, we design an enhanced FL framework with LDP, specifically tailored for sparse data completion, which achieves a balance between privacy and accuracy through adaptive noise scaling.

III. System Model and Problem Definition

A. System Model

A SMCS system consists of a server and K distributed clients (mobile devices). The target sensing area is divided into N subregions, where each client k can collect only partial observations from certain subregions within a given time period. Let $Y^{(k)} \in \mathbb{R}^{N\times T}$ denote the ground-truth data matrix of client k, where each row corresponds to a subregion and each column corresponds to a time step. Due to sparse sampling, the observed data matrix $Y'^{(k)}$ is masked by a binary indicator matrix $C^{(k)} \in 0, 1^{N\times T}$:

$$Y'^{(k)} = Y^{(k)} \odot C^{(k)}, \tag{1}$$

where $\odot$ denotes the Hadamard product (element-wise multiplication). The goal of the server is to collaboratively train a global model $f_G(\cdot)$ across clients such that the completed data $\hat{Y}^{(k)} = f_G(Y'^{(k)})$ can approximate the ground-truth data $Y^{(k)}$, while ensuring privacy preservation.

B. Problem Definition

Under the FL framework illustrated in Fig. 2, we model the data completion problem as a distributed optimization task with privacy constraints, consisting of three key components:

- **Local training:** Each client k trains a local model $f_k(\cdot)$ using its sparse dataset $Y'^{(k)}, C^{(k)}$. The model includes a lightweight RNN to capture temporal dependencies and a DMF decoder for spatial completion.
- **Privacy constraint:** The gradient update $\tilde{g}_k$ uploaded by client k must satisfy ϵ-differential privacy through gradient clipping and Laplace noise injection.
- **Active learning constraint:** After each training round, the client evaluates the value of unobserved points and selects the most informative ones within a budget B for sensing, thereby gradually expanding its local observation set and improving completion accuracy.
- **Global objective:** The server aggregates local gradients to update the global model θ_G, aiming to minimize the overall completion error under both privacy and active learning budget constraints:

$$\min_{\theta_G}\Bigg[\frac{1}{K}\sum_{k=1}^{K}\left(\frac{1}{2|Y'^{(k)}|}\left\|(Y^{(k)} - f_G(Y'^{(k)}))\odot C^{(k)}\right\|_F^2\right) + \lambda\cdot\underbrace{\frac{1}{K}\sum_{k=1}^{K}\|\Delta\theta_k - \Delta\tilde{\theta}_k\|_1}_{\mathcal{L}_{\text{priv}}(\epsilon)}\Bigg], \tag{2}$$

where $\Delta\tilde{\theta}_k = \Delta\theta_k + \text{Lap}(2C/\epsilon)$, $\Delta\theta_k = \theta_k - \theta_G$, and $|\cdot|_F$ denotes the Frobenius norm.

IV. Model Design

A. Fundamentals of Deep Matrix Factorization

The DMF framework reconstructs the complete data matrix $Y \in \mathbb{R}^{N\times T}$ from sparse observations $Y' = Y \odot C$ using a time-aware loss function:

$$\mathcal{L}_{\text{DMF}} = \frac{1}{2|Y'|}\sum_{t=1}^{T}\|(Y_t - f_\theta(Z_t))\odot C_t\|_F^2, \tag{3}$$

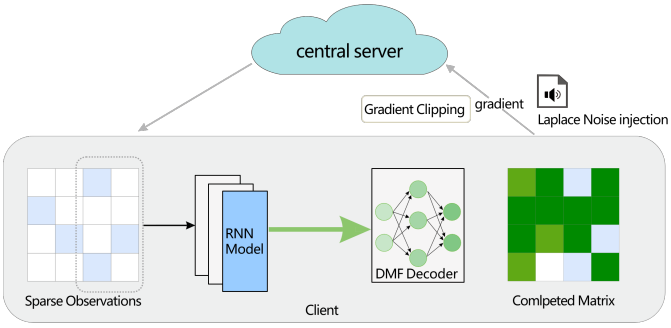


Fig. 3. Federated RNN-enhanced Deep Matrix Factorization with Differential Privacy.

where $Z_t \in \mathbb{R}^{N\times d}$ is the low-rank representation at time step t, and $f_\theta(\cdot)$ is a neural network parameterized by θ.

B. Federated RNN-enhanced DMF

FRDMF encodes temporal sequences on the client side using an RNN and decodes spatial correlations on the server side using a DMF; all parameters (including h_0) are synchronized and updated through federated aggregation. The local RNN on client k employs a sliding window of length $\tau = 24$:

$$h_{t'} = \text{RNN}\left(Y'^{(k)}_{t'-\tau:t'}, h_{t'-1}\right), \quad \tau = 24. \tag{4}$$

The global DMF decoder on the server maps the embeddings to the completed results:

$$\hat{Y}_t^{(k)} = f_{\theta_G}\left(W_z h_t + b_z\right), \tag{5}$$

where W_z, b_z, and θ_G are global parameters updated via federated aggregation.

As illustrated in Fig. 3, FRDMF captures temporal dependencies through the RNN and decodes spatial correlations through the DMF, thereby achieving sparse matrix completion.

C. Differential Privacy Guarantee

To mitigate privacy risks such as gradient inversion attacks, we incorporate gradient clipping and Laplace noise injection into the training and aggregation processes of FRDMF to satisfy ϵ-differential privacy. Specifically, the client gradient is clipped using the ℓ_2 norm as follows:

$$\bar{g}_k = g_k \cdot \min\left(1, C/\|g_k\|_2\right), \quad \|\bar{g}_k\|_2 \le C. \tag{6}$$

The first-order sensitivity satisfies $\Delta Q \le 2C$ (derived from clipping and the triangle inequality). Then, Laplace noise is injected according to the privacy budget:

$$\tilde{g}_k = \bar{g}_k + \text{Lap}\left(0, 2C/\epsilon\right). \tag{7}$$

This ensures that each communication round satisfies ϵ-differential privacy. The mechanism protects privacy by limiting the influence of any single data point and weakening the correlation between gradients and raw data. The privacy–utility trade-off is empirically validated in Section V, where the theoretical error bound is approximately $2C\sqrt{2}/\epsilon$.

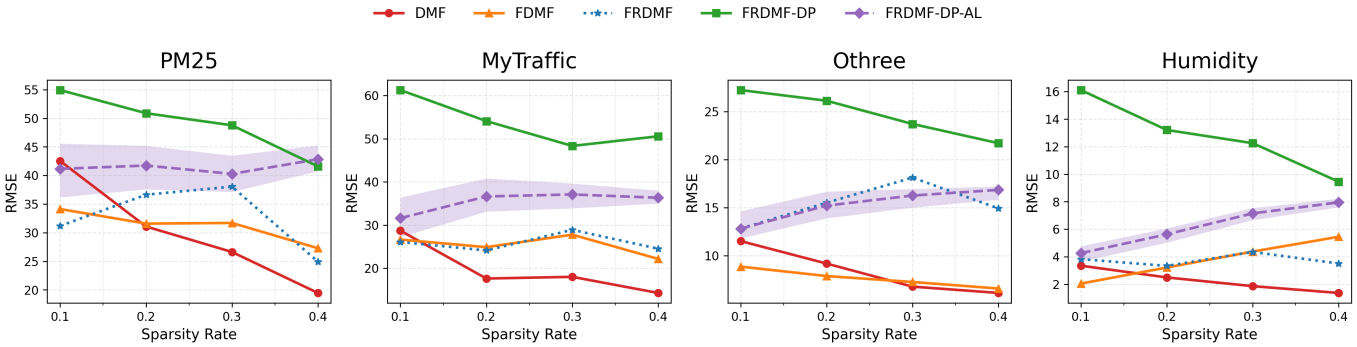


Fig. 4. Impact of Sparsity Rate on Model Robustness.

TABLE I
PERFORMANCE COMPARISON OF DIFFERENT MODELS.

Model	RMSE				Time Cost (s)				ASR (%)
	MyTraffic	PM25	Othree	Humidity	MyTraffic	PM25	Othree	Humidity	
DMF	23.47	37.35	8.20	1.86	2.13	2.07	2.07	2.89	100*
FDMF	23.69	29.09	7.28	3.22	1.20	1.09	1.28	1.72	24
RNN-DMF	103.92	94.63	78.26	28.19	183.80	35.40	148.93	37.52	100*
FRDMF	24.44	32.48	16.31	4.35	952.50	171.14	670.93	55.01	31
FRDMF-DP	50.72	48.81	21.76	12.25	410.86	79.23	267.17	55.84	19
FRDMF-DP-AL	**37.10**	**40.27**	**16.26**	**7.15**	**822.09**	**112.86**	**607.93**	**206.22**	**19**

D. Active FL-enhanced Deep Matrix Factorization

Random or fixed sampling often leads to low utilization under limited budgets. Therefore, we extend FRDMF with an active learning module, constructing the Federated Active Learning-enhanced Deep Matrix Factorization (FRDMF-DP-AL, DP stands for differential privacy and AL for active learning). At the end of each local epoch, Monte Carlo (MC) Dropout is used to evaluate the uncertainty of unobserved points, and the Top-B most informative samples are selected for sensing within the budget B, without altering the communication or privacy process.

1) Input and output definition: Let x_t denote the input feature at time step t, such as partial observations Y_t' or its low-rank representation Z_t; the corresponding y_t represents the ground-truth target value (i.e., the observation vector of the complete matrix at time t). Accordingly, the model prediction $\hat{y}_t$ is the completion result generated from x_t.

2) Uncertainty estimation: During the inference phase, M forward passes are performed for the same input, yielding a set of predictions:

$$\left\{\hat{y}_t^{(m)}\right\}_{m=1}^{M}, \quad \hat{y}_t^{(m)} = f_\theta(x_t; \text{Dropout}), \tag{8}$$

where Dropout remains activated during inference, thus introducing randomness. The predictive mean and variance are given by:

$$\bar{y}_t = \frac{1}{M}\sum_{m=1}^{M} \hat{y}_t^{(m)}, \sigma_t^2 = \frac{1}{M}\sum_{m=1}^{M}\left(\hat{y}_t^{(m)} - \bar{y}_t\right)^2. \tag{9}$$

The variance σ_t^2 serves as the uncertainty metric, which measures the stability of predictions at that position.

3) Sample selection strategy: At the end of each local training round, client k computes the uncertainty of all unobserved points, ranks them in descending order of σ_t^2, and selects the top B points as the candidate set for sensing:

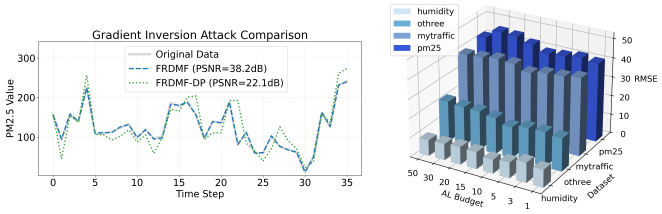
$$\mathcal{Q}^{(k)} = \arg\max_{|S|=B} \sum\nolimits_{t\in S} \sigma_t^2, \tag{10}$$

where B is the active learning budget. The selected $\mathcal{Q}^{(k)}$ will be added as new observations in the next training round, thereby gradually improving model accuracy.

4) Federated aggregation and update: The active learning process is embedded into the standard FL workflow:

- **Local training:** Each client trains the FRDMF model using the updated observation set.
- **Uncertainty estimation:** MC Dropout is applied to compute variances for unobserved points.
- **Sample selection:** Re-sense $\mathcal{Q}^{(k)}$ and expand the local dataset.
- **Gradient upload:** Upload updated gradients to the server under differential privacy protection.
- **Global aggregation:** The server updates the global parameters θ_G and distributes them back to the clients.

5) Theoretical Guarantee: Active learning admits formal guarantees in classical theory [32]: under classification and realizable conditions (e.g., low noise and a bounded disagreement coefficient), uncertainty-/disagreement-driven querying can reduce the label complexity from the passive-learning baseline $\Theta\big(d/\varepsilon_{\text{err}}\big)$ to approximately $O\big(\theta\, d\, \log(1/\varepsilon_{\text{err}})\big)$ (where d denotes the hypothesis-class complexity and θ the disagreement coefficient), i.e., the dependence on the target error ε_{err} improves from inverse to



(a) Gradient Inversion Attack Comparison. (b) Performance evaluations on AL Budget

Fig. 5. Privacy Robustness and Active-Learning Efficiency.

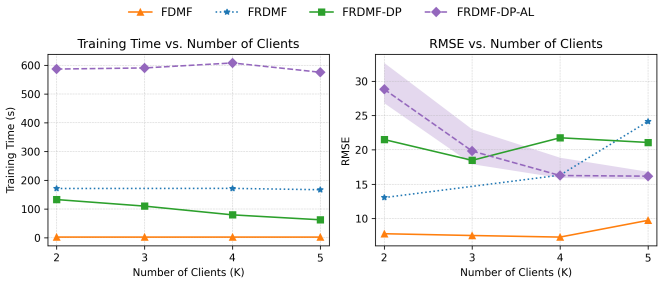


Fig. 6. Training Time and RMSE vs. Number of Clients.

logarithmic. In more general settings, there are still upper bounds showing improvements over random sampling, together with corresponding lower bounds. Our FRDMF-DP-AL adopts an "uncertainty-first" selection strategy consistent with these disagreement/uncertainty-driven theories, thereby providing a citable theoretical underpinning for the empirical finding that it outperforms random sampling under the same budget.

V. Experimental Validation

A. Datasets and Settings

We evaluate our proposed method using four real-world datasets: PM25 and Othree (36 monitoring stations in Beijing, hourly PM2.5/ozone data from 2014–2016, reflecting the dynamic variations of urban air quality), Humidity(humidity data from the EPFL campus sampled at 0.5-hour intervals, representing a limited spatial-coverage sampling scenario), and MyTraffic (taxi speed data from 50 main roads at 10-minute intervals, representing spatiotemporal traffic patterns). All data are normalized to the range of $[0, 1]$, and sparse sensing scenarios are simulated using sparsity ratios $0.1, 0.2, 0.3, 0.4$. Each dataset is divided into training (70%), validation (15%), and test (15%) sets while maintaining temporal continuity. The experimental setup includes $K = 2, 3, 4, 5$ clients, 30 local training epochs per round, a batch size of 32, and 30 communication rounds. The default active learning budget is set to 10 (see Fig. 5b).

B. Baselines and Result Analysis

We compare our proposed framework against the following baseline models: centralized DMF [8], RNN-DMF, federated DMF (FDMF), federated RNN-enhanced DMF (FRDMF), FRDMF with differential privacy (FRDMF-DP), and FRDMF-DP with active learning (FRDMF-DP-AL). These models are designed to evaluate the feasibility of sparse data completion from centralized, federated, and privacy-preserving perspectives. The model architecture consists of an RNN with a hidden dimension of 64 and a DMF latent space of dimension 32, optimized using the Adam optimizer with learning rate $\eta = 0.001$, $\beta_1 = 0.9$, and $\beta_2 = 0.999$.

1) Overall Performance and Privacy Protection: Under the benchmark setting (sparsity rate $= 0.3$, $K = 4$ clients, AL budget $= 10$), FRDMF-DP achieves the best performance in terms of privacy metrics. However, compared with FRDMF without differential privacy, its average RMSE increases by 65.7%, indicating a significant drop in accuracy. Since active learning (AL) only affects the sampling process during training without modifying the privacy mechanism, FRDMF-DP-AL provides similar privacy protection as FRDMF-DP, while achieving improved accuracy (Table I): on the Othree dataset, RMSE decreases from 21.76 to 16.26 (-25.3%), and ASR drops from 31% to 19%, demonstrating the effectiveness of privacy preservation. Notably, the centralized DMF achieves a lower RMSE (23.47) on the MyTraffic dataset, but its ASR reaches 100%, exposing the privacy risks of centralized schemes. In contrast, federated training inherently avoids centralized raw data storage, and the addition of differential privacy noise injection significantly reduces attack strength (Fig. 5a: PSNR 38.2 dB $\rightarrow$ 22.1 dB).

2) Sparse Adaptability of Federated Learning: As shown in Fig. 4, the federated models exhibit strong robustness to sparse data under different sparsity rates. Taking PM25 as an example, when the sparsity rate increases from 0.1 to 0.4, the RMSE of FRDMF-DP changes from $54.6 \rightarrow 41.57$ (-23.9%), while the fluctuation of FRDMF-DP-AL is only 4.08% ($41.15 \rightarrow 42.83$). In contrast, the RMSE of the centralized DMF varies by 54.5% ($42.52 \rightarrow 19.38$). This difference arises because the federated models can capture localized spatiotemporal features through client-side modeling rather than relying solely on global correlations. Moreover, temporal modeling with the RNN is crucial for sparse adaptability: on the Othree dataset, removing the RNN leads to a 46.3% increase in RMSE (see Table II).

TABLE II
Ablation Experiment: The Necessity of RNN, DP and AL for Model Performance (ε=18, Sparsity=0.3).

Model	RMSE				ASR (%)			
	MyTraffic	PM2.5	Othree	Humidity	MyTraffic	PM2.5	Othree	Humidity
RNN-DMF	103.92	94.63	78.26	28.19	100	99	100	97
FDMF-DP	25.16	33.12	12.87	12.41	25	26	12	19
FRDMF	24.44	32.48	16.31	4.35	32	35	31	33
FRDMF-DP	50.72	48.81	21.76	12.25	26	28	19	22
FRDMF-DP-AL	37.10	40.27	16.26	7.15	26	28	19	22

3) Adjustment Effects of Privacy Budget: The privacy budget ϵ directly affects both model accuracy and privacy performance. When $\epsilon = 18$, FRDMF-DP achieves a favorable balance on the Othree dataset (RMSE $= 21.76$, ASR $= 13.3\%$). Compared with $\epsilon = 6$, the RMSE decreases significantly while the increase in ASR remains limited, indicating that a larger ϵ value can achieve effective privacy protection without noticeably compromising model accuracy. In terms of

scalability (Fig. 6), when the number of clients increases from $K = 2$ to $K = 5$, the RMSE fluctuation remains below 2%, and the total training time is reduced by approximately 53%, demonstrating that the proposed framework is well-suited for large-scale distributed scenarios.

VI. Conclusion

This paper proposes a federated active learning framework for SMCS data completion, which integrates localized temporal modeling, global matrix factorization, and ϵ-differential privacy under the condition that raw data are never shared. Three federated models are developed: FDMF, FRDMF, and FRDMF-DP-AL. Experiments on four real-world datasets demonstrate that when $\epsilon = 18$, differential privacy reduces the PSNR of gradient inversion attacks from 38.2 dB to 22.1 dB with only limited accuracy degradation; under the same privacy mechanism, active learning further reduces the error (e.g., on Othree, RMSE $21.76 \rightarrow 16.26$) without increasing privacy risk (ASR remains comparable to FRDMF-DP).

Overall, the proposed framework achieves a verifiable balance among privacy, utility, and sampling efficiency: federated collaboration eliminates raw data sharing, the lightweight RNN enhances stability under sparse scenarios, differential privacy effectively suppresses gradient inversion attacks, and uncertainty-driven active sampling significantly improves completion accuracy within a given budget.

Acknowledgment

This work is supported in part by National Natural Science Foundation of China under Grant Nos. 92567204, 62272193, and 62472194, and Jilin Science and Technology Research Project 20260101016JJ.

References

[1] R. Nasser, R. Mizouni, S. Singh, and H. Otrok, "Systematic survey on artificial intelligence based mobile crowd sensing and sourcing solutions: Applications and security challenges," *Ad Hoc Networks*, vol. 164, p. 103634, 2024.

[2] 路文浩, 赵勇, 季雅泰, 张琪, 许凯, and 朱正秋, "面向应急响应群智感知的异构群体协作任务分配方法," 物联网学报, vol. 9, no. 4, 2025.

[3] H. Aly, A. Basalamah, and M. Youssef, "Automatic rich map semantics identification through smartphone-based crowd-sensing," *IEEE Transactions on Mobile Computing*, vol. 16, no. 10, pp. 2712–2725, 2016.

[4] L. Wang, D. Zhang, Y. Wang, C. Chen, X. Han, and A. M'hamed, "Sparse mobile crowdsensing: challenges and opportunities," *IEEE Communications Magazine*, vol. 54, no. 7, pp. 161–167, 2016.

[5] J. Fan and J. Cheng, "Matrix completion by deep matrix factorization," *Neural Networks*, vol. 98, pp. 34–41, 2018.

[6] E. De Cristofaro and G. Tsudik, "Practical private set intersection protocols with linear complexity," in *International Conference on Financial Cryptography and Data Security*, pp. 143–159, 2010.

[7] W. Liu, Y. Yang, E. Wang, and J. Wu, "Fine-grained urban prediction via sparse mobile crowdsensing," in *IEEE International Conference on Mobile Ad Hoc and Sensor Systems*, pp. 265–273, 2020.

[8] K. Xie, X. Li, X. Wang, G. Xie, J. Wen, and D. Zhang, "Active sparse mobile crowd sensing based on matrix completion," in *Proceedings of the International Conference on Management of Data*, pp. 195–210, 2019.

[9] R. Zhao, P. Zhi, X. Yang, Z. Zhang, G. Liu, C. Di, and Q. Zhou, "Client to server: Heterogeneous distribution knowledge transfer for federated learning," *Tsinghua Science and Technology*, vol. 30, no. 1, p. 112–123, 2025.

[10] M. Cheung, "Feddmf: Privacy-preserving user attribute prediction using deep matrix factorization," *arXiv preprint arXiv:2312.15420*, 2023.

[11] K. Manohar, B. W. Brunton, J. N. Kutz, and S. L. Brunton, "Data-driven sparse sensor placement for reconstruction: Demonstrating the benefits of exploiting known patterns," *IEEE Control Systems Magazine*, vol. 38, no. 3, pp. 63–86, 2018.

[12] C. Yang, K. Liang, X. Zhang, and X. Geng, "Sensor placement algorithm for structural health monitoring with redundancy elimination model based on sub-clustering strategy," *Mechanical Systems and Signal Processing*, vol. 124, pp. 369–387, 2019.

[13] A. Singh, R. Nowak, and P. Ramanathan, "Active learning for adaptive mobile sensing networks," in *Proceedings of the 5th International Conference on Information Processing in Sensor Networks*, pp. 60–68, 2006.

[14] N. Karnik, M. G. Abdo, C. E. Estrada-Perez, J. S. Yoo, J. J. Cogliati, R. S. Skifton, P. Calderoni, S. L. Brunton, and K. Manohar, "Constrained optimization of sensor placement for nuclear digital twins," *IEEE Sensors Journal*, vol. 24, no. 9, pp. 15501–15516, 2024.

[15] J. Geiping, H. Bauermeister, H. Dröge, and M. Moeller, "Inverting gradients-how easy is it to break privacy in federated learning?," *Advances in Neural Information Processing Systems*, vol. 33, pp. 16937–16947, 2020.

[16] Y. Zhu, Z. Li, H. Zhu, M. Li, and Q. Zhang, "A compressive sensing approach to urban traffic estimation with probe vehicles," *IEEE Transactions on Mobile Computing*, vol. 12, no. 11, pp. 2289–2302, 2012.

[17] E. Wang, M. Zhang, W. Liu, H. Xiong, B. Yang, Y. Yang, and J. Wu, "Outlier-concerned data completion exploiting intra-and inter-data correlations in sparse crowdsensing," *IEEE/ACM Transactions on Networking*, vol. 31, no. 2, pp. 648–663, 2022.

[18] X. Meng, S. Sun, X. Zhang, Q. Leng, and J. Fang, "A survey on trajectory representation learning methods," *Frontiers of Computer Science*, vol. 19, no. 12, p. 1912379, 2025.

[19] Z. Wang, J. Hu, R. Lv, J. Wei, Q. Wang, D. Yang, and H. Qi, "Personalized privacy-preserving task allocation for mobile crowdsensing," *IEEE Transactions on Mobile Computing*, vol. 18, no. 6, pp. 1330–1341, 2018.

[20] I. J. Vergara-Laurens, L. G. Jaimes, and M. A. Labrador, "Privacy-preserving mechanisms for crowdsensing: Survey and research challenges," *IEEE Internet of Things Journal*, vol. 4, no. 4, pp. 855–869, 2016.

[21] M. Wang, L. Zhou, X. Huang, and W. Zheng, "Towards federated learning driving technology for privacy-preserving micro-expression recognition," *Tsinghua Science and Technology*, vol. 30, no. 5, pp. 2169–2183, 2025.

[22] S. Wei, Y. Tong, Z. Zhou, Y. Xu, J. Gao, T. Wei, T. He, and W. Lv, "Federated reasoning llms: a survey," *Frontiers of Computer Science*, vol. 19, no. 12, p. 1912613, 2025.

[23] N. Ramakrishnan and T. Soni, "Network traffic prediction using recurrent neural networks," in *IEEE International Conference on Machine Learning and Applications*, pp. 187–193, 2018.

[24] D. Li, Z. Wang, Y. Chen, R. Jiang, W. Ding, and M. Okumura, "A survey on deep active learning: Recent advances and new frontiers," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 36, no. 4, pp. 5879–5899, 2024.

[25] A. Saran, S. Yousefi, A. Krishnamurthy, J. Langford, and J. T. Ash, "Streaming active learning with deep neural networks," in *International Conference on Machine Learning*, pp. 30005–30021, 2023.

[26] M. Bailey, S. Moayedpour, R. Li, A. Corrochano-Navarro, A. Kötter, L. Kogler-Anele, S. Riahi, C. Grebner, G. Hessler, H. Matter, *et al.*, "Deep batch active learning for drug discovery," *bioRxiv*, pp. 2023–07, 2023.

[27] S. Mohamadi and H. Amindavar, "Deep bayesian active learning, a brief survey on recent advances," *arXiv preprint arXiv:2012.08044*, 2020.

[28] X. Yi, R. Paulet, E. Bertino, X. Yi, R. Paulet, and E. Bertino, *Homomorphic Encryption*. Springer, 2014.

[29] O. Goldreich, "Secure multi-party computation," *Manuscript. Preliminary Version*, vol. 78, no. 110, pp. 1–108, 1998.

[30] S. Truex, L. Liu, K.-H. Chow, M. E. Gursoy, and W. Wei, "Ldp-fed: Federated learning with local differential privacy," in *Proceedings of the third ACM International Workshop on Edge Systems, Analytics and Networking*, pp. 61–66, 2020.

[31] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of Cryptography: Third Theory of Cryptography Conference*, pp. 265–284, 2006.

[32] S. Hanneke, "Theory of disagreement-based active learning," *Foundations and Trends in Machine Learning*, vol. 7, no. 2-3, pp. 131–309, 2014.