A TWO-DIMENSIONAL HYBRID FEDERATED LEARNING FRAMEWORK FOR SECURE DATA COOPERATION OF MULTIPLE NETWORK SERVICE PROVIDERS

Huali Lu, Wenxiong Chen, Conghao Zhou, Huaqing Wu, Feng Lyu, and Xuemin (Sherman) Shen

ABSTRACT

In the era of artificial intelligence of things (AIoT), the substantial user data collected by network service providers (NSPs) holds great potential for enhanced service provisioning. However, the data isolation between NSPs limits the full utilization of data. In this article, we investigate collaborative computing among NSPs to fully unlock the potential of heterogeneous cellular data, with comprehensive discussions on its advantages, applications, and challenges. We propose a two-dimensional hybrid federated learning (*2DHFL*) framework to facilitate collaborative computing among multiple NSPs, with ingenious integration of horizontal and vertical federated learning techniques. The *2DHFL* framework can effectively address the challenges related to incomplete features and insufficient training data while maintaining data privacy. A case study on map matching task is presented with empirical data-driven experiment results to demonstrate the effectiveness of the *2DHFL* framework compared with state-of-the-art benchmark schemes.

INTRODUCTION

As 4G becomes ubiquitous, and the rollout of 5G accelerates, we are ushering in the era of the Artificial Intelligence of Things (AIoT), where cellular networks are the foundation for the Internet of Everything (IoE) [1, 2]. This transition is accomplished by a rapid expansion in the number of smart terminal devices and mobile users. According to Ericsson's report, there will be a substantial rise in cellular IoT connections, from 1.7 billion in 2020 to 5.9 billion in 2026. The large accumulation of IoT data enables the rapid development of AIoT. Currently, the scope of AIoT devices encompasses various domains, ranging from smart wearable devices and utility meters to infrastructure elements like smart manhole covers and on-board terminals. This technology integration spans across smart city initiatives, intelligent transportation systems, environmental monitoring, healthcare, and more. Consequently, major network service providers (NSPs) and Internet companies have amassed vast quantities of user data. As an example, in the first half of 2022, China's mobile Internet access traffic reached 124.1 billion GB, a year-on-year increase of 20.2 percent [3]. Leveraging big data technologies, cellular users' data support large-scale mobile perception and computing to advance the digital economic development, urban digital governance, residents' mobile behavior analysis, epidemics management, NSP resource allocation, and more.

Cellular networks offer significant advantages such as continuous availability, comprehensive coverage, high network penetration, and the support of device mobility. As a result, major NSPs can monitor and record all-weather, all-mission, all-user real-time communication processes between terminal devices and base stations (BSs), in terms of connection times, associated BSs, utilized applications (Apps), and consumed data traffic. These monitored data can accurately reflect users' mobile patterns, lifestyle patterns, and social interactions. While NSPs can utilize these monitored data to enhance the quality of service provisioning, each NSP must independently collect and store data owing to commercial and privacy constraints. As a result, the dataset for the collected data is fragmented across the city, which significantly limits comprehensive city-wide insights and t2Dhe support of large-scale perception activities. Recognizing these limitations, the industry has increasingly emphasized the cooperation among NSPs [4]. In China, major NSPs, including China Mobile, China Telecom, China Unicom, and China Broadnet, have underscored the necessity for intensified collaboration. Their goal is to unlock the full potential of communication big data and drive the future digital evolution of the communication sector. Within such a cooperative framework, NSPs can participate in diverse city perception activities at a large scale through collaborative computing among their data to enhance the overall urban experience.

There exist a few studies [5–9] utilizing cellular data for varied applications, such as coronavirus disease (COVID-19) tracing, map matching (MM), and pedestrian density prediction. For instance, Shen *et al.* [6] introduced a recurrent neural network-based model for map matching using cellular

Huali Lu and Feng Lyu are with Central South University, China;
Wenxiong Chen (corresponding author) is with Hunan Normal University, China; Huaqing Wu is with University of Calgary, Canada;
Conghao Zhou and Xuemin (Sherman) Shen are with University of Waterloo, Canada.

Digital Object Identifier: 10.1109/MWC.018.2300534

1536-1284/24/$25.00 © 2024 IEEE

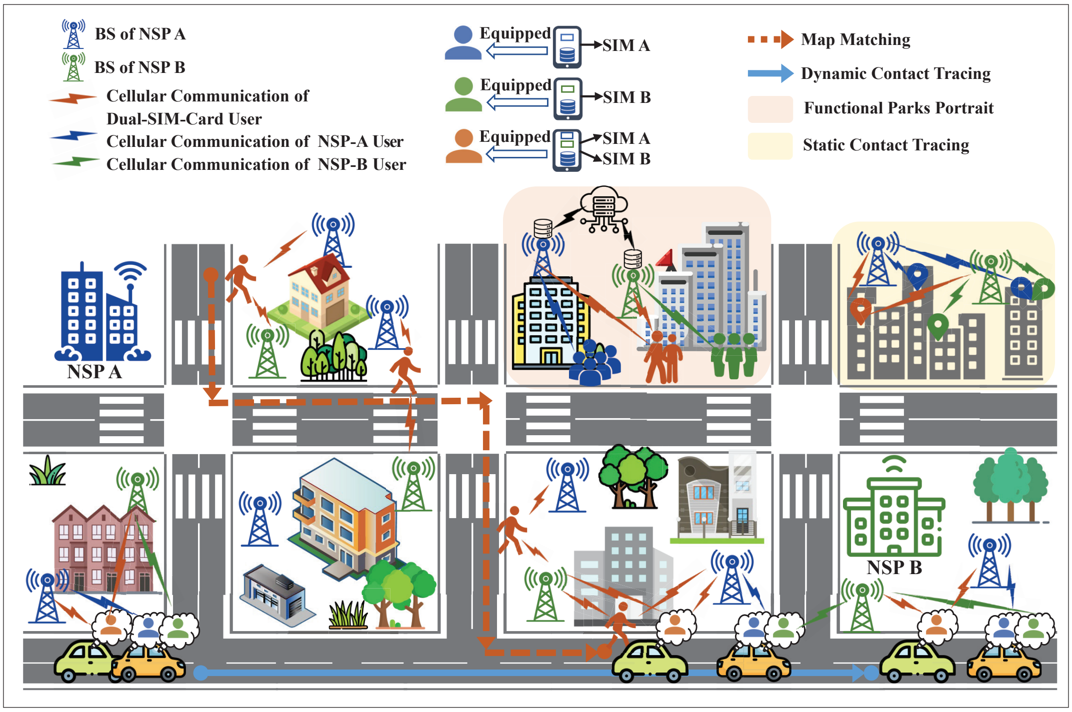


FIGURE 1. Illustration of the considered collaborative computing for multiple network service providers scenarios.

trajectories, and Huo *et al.* [7] estimated short-term pedestrian density with cellular data. There are also studies [8, 9] focusing on monitoring supply networks, assessing economic risks, and predicting cellular traffic, respectively. Despite these valuable contributions, most existing research has concentrated solely on harnessing cellular data from individual NSPs, rarely considering collaborative computing among multiple NSPs. This often results in suboptimal model accuracy and restricts the scope of practical applications. To bridge this research gap, we propose an innovative framework for collaborative computing that involves multiple NSPs to unlock the full potential of collective data intelligence.

Conducting collaborative computing among NSPs encounters substantial challenges, primarily in safeguarding data privacy. To address these challenges, we propose a privacy-preserving two-dimensional hybrid federated learning (*2DHFL*) framework, which ingeniously integrates horizontal and vertical federated learning (FL). This framework can effectively resolve issues related to incomplete features and insufficient training data while preserving data privacy. Specifically, considering the escalating prevalence of dual-SIM-card users (i.e., individuals who simultaneously use two subscriber identity module (SIM) cards from different NSPs and have their cellular usage data collected by these NSPs), a vertical federated learning (VFL) scheme is developed to enhance the learning model. For each pair of NSPs, a model can be trained using VFL to aggregate shareable features for their shared dual-SIM-card users into a comprehensive feature space. Concurrently, we employ horizontal federated learning (HFL) to adeptly aggregate these models from different NSP pairs to construct a high-quality global model. This hybrid approach ensures both effective model refinement and robust data privacy in the collaborative computing among NSPs.

The remainder of this article is organized as follows. We first introduce the advantages, applications, and challenges of collaborative computing among multiple NSPs. Then we present preliminaries on federated learning and the motivation for designing this novel framework. Afterwards, we elaborate on our proposed *2DHFL* framework. Finally, we carry out a case study to demonstrate the efficacy of our proposed framework, followed by the conclusion in the final section.

ADVANTAGES, APPLICATIONS AND CHALLENGES

In this section, we will first highlight the advantages of collaborative computing for multiple NSPs. Next, we will introduce several significant applications, as depicted in Fig. 1, followed by the challenges required to be addressed.

ADVANTAGES

Collaborative computing among NSPs introduces the following three advantages in comparison to single-NSP data-driven solution.

Breaking Data Barriers and Enhancing Data Integration: Due to privacy protection and security concerns, data collected by major NSPs should be stored and utilized separately, leading to distinct data islands. However, relying solely on one NSP's data proves insufficient for capturing comprehensive information. Collaborative computing for multiple NSPs overcomes the physical isolation of data from different NSPs while ensuring privacy protection, thereby allowing for the realization of multi-source data fusion to break data barriers and enhance data integration.

Expanding New Application Scenarios: Each single NSP has limited user coverage scope, resulting in constrained data sources. Consequently, lots of applications, such as city-wide App popularity analysis or nationwide contact tracing during an epidemic, cannot be effectively implemented. However, collaborative computing for multiple

NSPs facilitates the realization of such applications by breaking data barriers and facilitating data interaction, which can expand application scenarios to foster new business forms and patterns.

Enhancing Service Performance and Intelligence: Traditional applications relying on data collected from a single NSP often suffer from sparse and noisy cellular data, leading to unsatisfactory performance. For instance, the accuracy of single-NSP cellular data-based map matching generally requires further improvement due to the insufficient features resulted by the sparse and noisy cellular data. In contrast, collaborative computing for multiple NSPs allows for the effective gathering and utilization of complementary information through secured data interaction among multiple NSPs. This enhances model accuracy performance and intelligence, therefore facilitating high-quality service provision.

APPLICATIONS

Collaborative computing among multiple NSPs empowers a plethora of innovative applications, which can be primarily divided into two categories. In this subsection, we will explore exemplary applications within these categories, as shown in Fig. 1, using the collaboration between two NSPs as an example.

Performance-Enhancing Applications: These applications leverage collaborative computing among NSPs to improve their performance, addressing limitations encountered when using data from a single NSP.

Cellular Data-Based Map Matching: Map matching plays a crucial role in trajectory-based applications, where Global Positioning System (GPS) or cellular association traces are matched to road networks to identify accurate moving paths and positions of objects. Compared to GPS-based schemes which have limitations in scalability and user privacy concerns, cellular data-based schemes offer advantages in terms of low-cost data collection and privacy guarantee. However, the matching performance based on user trajectories from a single NSP may be suboptimal due to poor data quality, such as sparsity and location noise. Collaborative computing among NSPs can effectively address this issue by utilizing complementary features provided by different NSPs for the same path with information fusion.

Cellular Data-Based Functional Parks Portrait: In urban planning and development, identifying and characterizing urban functional parks is paramount. These areas serve as regional hubs for the aggregation of social resources and the efficient execution of specific urban functions, such as high-tech industrial parks, business districts, and medical park. To support the construction of smart cities, it is vital to understand various parameters within these regions, including population distribution, mobility, electricity consumption, traffic usage, and network load. However, the data provided by a single NSP may be incomplete, thereby challenging the capture of full characteristics of these functional parks. Collaborative computing among multiple NSPs can overcome this limitation by leveraging data from various sources to create more accurate portraits of functional parks.

Business-Expansion Applications: These applications leverage collaborative computing among NSPs to enable interactions and business expansion, which generally cannot be realized with a single NSP's data alone.

Cellular Data-Based Static Contact Tracing: Contact tracing is critical for identifying and controlling the exposure of individuals in various contexts, significantly contributing to the control of infectious disease, surveillance, and monitoring activities of sensitive individuals. Cellular data-based static contact tracing is effective due to the broad user coverage, allowing it to infer whether two individuals have come into contact by identifying if they were in the same space based on three-dimensional (3D) indoor positioning. Due to the diversity of SIM cards used by users, the cellular data of two individuals may be collected by different NSPs. Therefore, city-wide static contact tracing cannot be achieved based solely on a single NSP's data due to its limited user coverage. Collaborative computing among multiple NSPs has the potential to overcome limitations, enabling effective static contact tracing.

Cellular Data-Based Dynamic Contact Tracing: Dynamic contact tracing is equally important as the static contact tracing, since user contacts may occur in both static and dynamic environments, such as during travel. Traditional localization technology may not be suitable for determining contact between moving users due to high cost, low accuracy or device limitation, such as GPS positioning and Wireless Fidelity (WiFi) positioning. Dynamic contact tracing based on cellular data is a decent solution by matching users' moving trajectories and assessing spatio-temporal similarity between trajectory pairs to determine contact. Similar to static contact tracing, dynamic contact tracing also faces the challenge that a single NSP cannot provide complete information to deal with this task. Therefore, collaborative computing among multiple NSPs also enables the realization of this application.

In summary, collaborative computing among multiple NSPs opens up new possibilities for performance-enhancing and business-expansion applications. By leveraging the collective data and resources from multiple NSPs, these applications can overcome limitations in terms of data missing, small data scale, and incomplete data features. This contributes to the improvement of model accuracy, and the comprehensive exploring of city-wide user interaction behaviors to promote the construction of smart cities, leading to enhanced services and capabilities in various domains.

By leveraging the collective data and resources from multiple NSPs, these applications can overcome limitations in terms of data missing, small data scale, and incomplete data features. This contributes to the improvement of model accuracy, and the comprehensive exploring of city-wide user interaction behaviors to promote the construction of smart cities, leading to enhanced services and capabilities in various domains

CHALLENGES

Despite the significant potential of collaborative computing among multiple NSPs, the following challenges must be addressed to ensure successful implementation.

Data Isolation: Each NSP's independent data storage and management results in data isolation among NSPs. This hinders the sharing, integration, and utilization of data across NSPs, limiting the full potential of cellular data. Overcoming data isolation while enabling efficient data sharing and collaboration between NSPs is a critical challenge in multi-NSP collaborative computing.

Privacy and Security: User data collected by each NSP contains sensitive information that must be protected and kept confidential according to

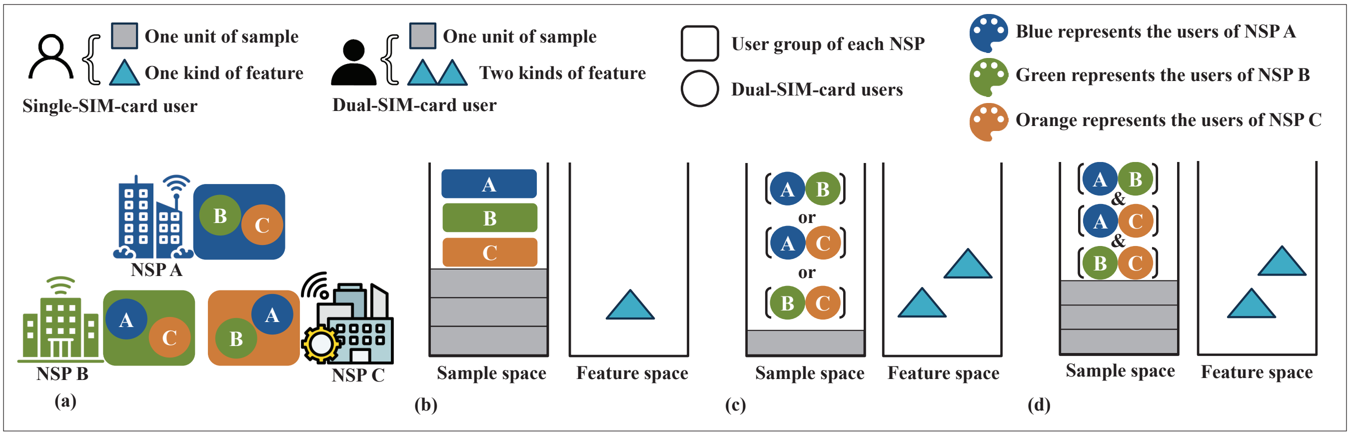


FIGURE 2. An illustration of the research motivation: a) Data distribution of multiple NSPs; b) HFL scenario and characteristic; c) VFL scenario and characteristic; d) 2DHFL scenario and characteristic.

privacy regulations. How to foster effective collaboration using collected user data from different NSPs while preserving data security and user privacy is a paramount challenge.

Heterogeneity of Data: Different NSPs may utilize distinct strategies to deploy BSs and collect data, leading to differences in data formats, structures, and qualities. Therefore, ensuring effective sharing and aggregation of the heterogeneous data across multiple NSPs to guarantee compatibility and consistency is crucial for meaningful collaborative computing.

Data Quality and Reliability: Ensuring the quality and reliability of shared information is a significant challenge in collaborative computing among multiple NSPs. This is because data inconsistencies, biases, errors, and outliers can affect the accuracy and validity of collaborative models and analysis. Therefore, the implementation of data cleansing, preprocessing, and quality assurance techniques is necessary to maintain the integrity and reliability of the shared data.

Addressing these challenges requires multi-party collaborative efforts. The development of standardized protocols, privacy-preserving frameworks, and data governance mechanisms holds the key to unlocking the full potential of collaborative computing among multiple NSPs.

Limitation of Traditional Federated Learning

As discussed in previous sections, effective data sharing and aggregation while maintaining data privacy and security plays a vital role in enabling collaborative computing across NSPs. To achieve this, it is imperative to establish a robust framework for seamlessly interfacing and aggregating data features from diverse NSPs.

FL emerges as a pioneering machine learning paradigm crafted specifically for training models using decentralized and privacy-sensitive data sources [10–12]. This innovative paradigm facilitates collaborative learning, allowing distributed entities to collaboratively construct a global model without sharing raw data. Within the realm of FL, two specialized variants, HFL and VFL, have emerged. HFL is characterized as a feature-aligned FL framework, aligning data features among participants with different training samples, effective in scenarios where entities share similar data attributes or features from diverse sources. Conversely, VFL is a sample-aligned FL framework, featuring participants with overlapping training sample IDs but less overlap in data features, suitable for situations where entities possess complementary data attributes.

However, the conventional pure FL framework falls short in multi-NSP scenarios due to the inherent complexity of data distribution. The rise of dual-SIM-card users exacerbates these intricacies, as shown in Fig. 2a. In single-SIM-card user scenarios, different NSPs serve distinct user groups, and each user within that group is equipped with a single SIM card from that NSP. Thus, each NSP maintains dissimilar data samples while sharing similar features, naturally aligning with the HFL framework, as depicted in Fig. 2b. However, the emergence of dual-SIM-card users, who concurrently associate with two NSPs, complicates the situation as they bring rich information to both NSPs. In scenarios with dual-SIM-card users, each NSP pair aligns with the VFL framework, with distinct features but shared samples, as illustrated in Fig. 2c. A closer examination of the HFL and VFL frameworks reveals their drawbacks. Relying solely on the HFL framework, although the samples are sufficient, the features are singular, and the rich information brought by dual-SIM-card users is not effectively utilized. Using the VFL framework alone, despite its abundant features, leads to insufficient samples. Consequently, developing a novel FL framework is imperative to address the current dilemma faced by single HFL or VFL. Notably, some studies in the existing literature attempt to optimize HFL and VFL frameworks by proposing hierarchical or hybrid federated learning. However, these architectures are not well-suited for the collaborative computing scenario involving multiple NSPs, primarily for two reasons. Firstly, most of them only consider single-dimensional framework optimization. Secondly, even if a two-dimensional hybrid design is considered, it is often tailored for specific scenarios and lacks generalizability to other application contexts. Therefore, motivated by this, we propose a two-dimensional hybrid federated learning framework that leverages the advantages of both HFL and VFL structures with sufficient samples and abundant features, depicted in Fig. 2d.

The *2DHFL* Framework

Figure 3 shows the design overview of our proposed *2DHFL* framework, where multiple NSPs

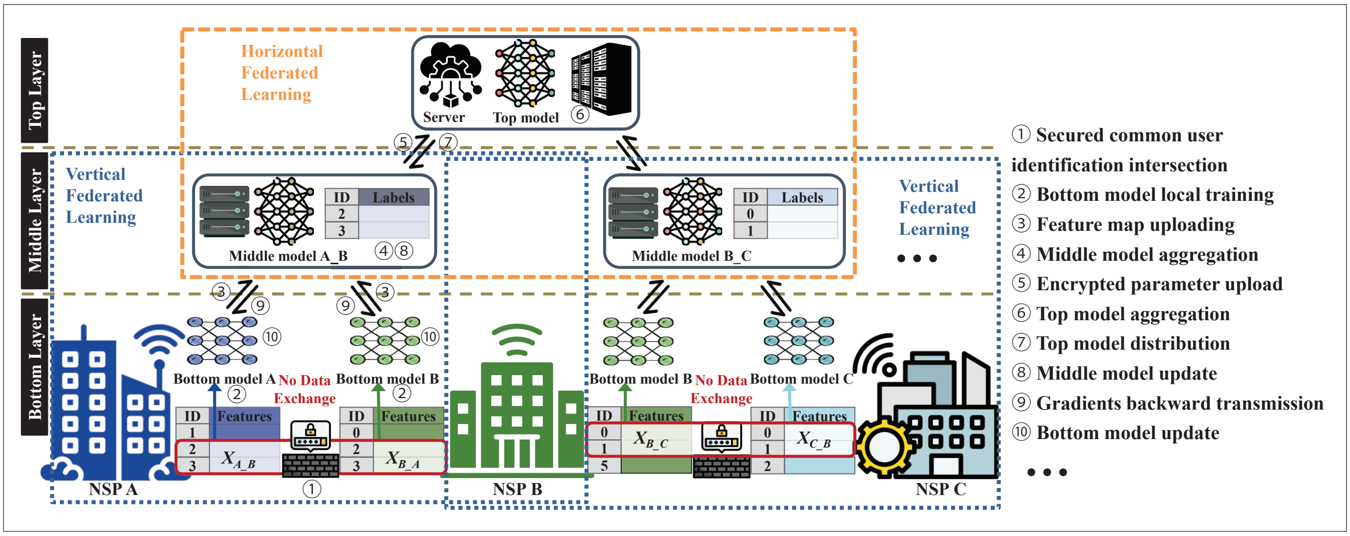


FIGURE 3. Design overview of *2DHFL* framework.

can collaboratively train a powerful global model in a privacy-preserving manner. Our *2DHFL* framework consists of three layers. The first layer is the bottom layer, constituted by heterogeneous NSPs. Each NSP serves as a distributed client node, and can locally train task models based on its own cellular data. The second layer is the middle layer composed of several edge servers, which act as trusted third parties, aggregating data features from their managed NSPs in the bottom layer to train an enhanced middle model. The third layer is the top layer consisting of a trusted cloud server with powerful computing capability, which aggregates the model parameters transmitted from the middle layer to train the powerful global model.

As shown in Fig. 3, the implementation of *2DHFL* consists of two parts. The first part is the VFL component aiming to enhance the model using all the available data from heterogeneous NSP pairs with keeping sensitive data safe and uncompromised. Different NSPs do not exchange their raw data; instead, only encrypted data information is collected and aggregated by a trusted third party to be used to train a middle model with a more complete feature space. The second part is the HFL component designed to address the issue of insufficient training data within each NSP pair and insecure data sharing across multiple NSP pairs. Specifically, the cloud server aggregates encrypted model parameters uploaded by middle servers to enhance the quality of the global model. As a result, the sufficient knowledge between multiple NSP pairs can be shared and aggregated without direct data exchange.

Next, we introduce the workflow of the *2DHFL* framework and detail implementation procedure as follows.

Secured Common User Identification Intersection: Given each NSP pair, the first step is to find the common identifiers served by both NSPs to align the training data samples. This involves finding users who own SIM cards from both NSPs, by a secure multi-party protocol, specifically the RSA-based blind signature scheme.

Bottom Model Local Training: Each NSP trains its bottom model locally to extract feature map based on its specific and non-shared data samples aligned with other NSPs.

Feature Map Uploading: Each NSP uploads the extracted local feature map to the middle server. Concurrently, for privacy protection, a homomorphic encryption algorithm based on Brakerski-Gentry-Vaikuntanathan (BGV) scheme is exploited to deal with privacy leakage problem that possibly occurs during the transmission.

Middle Model Aggregation: Each middle server aggregates encrypted feature maps offered by each NSP pair to train a middle model.

Encrypted Parameter Upload: Each middle server encrypts its current model parameters by a modified BGV-based somewhat homomorphic encryption algorithm and then uploads them to the global cloud server.

Top Model Aggregation: The cloud server aggregates the encrypted model parameters from each middle model to establish a global model.

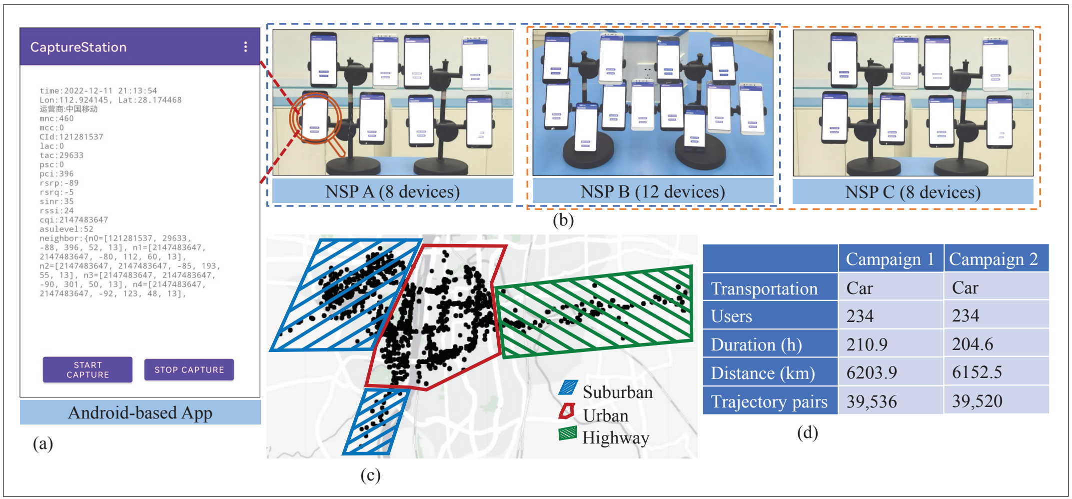
Top Model Distribution: The updated global model parameters are shared and distributed to middle servers to facilitate further personalized middle model training.

Middle Model Update: Each middle server updates its model parameters based on the received parameters of the global model.

Gradients Backward Transmission: Each middle server performs backward transmission, sending the gradients back to each NSP.

Bottom Model Update: Each NSP calculates the gradients for its bottom model parameters based on the local data and gradients received from the middle server, and then updates the bottom model using the gradients.

Interactive iterations between the local NSPs, the middle servers, and the global cloud server are performed in multiple rounds to ultimately optimize the global model. This collaborative computing framework eliminates the need for raw data exchange while aggregating the unique features of local data from each NSP. It is worth noting that our *2DHFL* framework is applicable to all scenarios involving two or more participating NSPs. In cases with two participating NSPs, *2DHFL* can be simplified into a pure VFL framework when there are dual-SIM-card users, or into a pure HFL framework when no dual-SIM-card users exist. Obviously, the accuracy of the final global model increases with the number of participating NSPs.



```
time:2022-12-11 21:13:54
Lon:112.924145, Lat:28.174468
运营商:中国移动
mnc:460
mcc:0
CId:121281537
lac:0
tac:29633
psc:0
pci:396
rsrp:-89
rsrq:-5
sinr:35
rssi:24
cqi:2147483647
asulevel:52
neighbor:{n0=[121281537, 29633,
-88, 396, 52, 13], n1=[2147483647,
2147483647, -80, 112, 60, 13],
n2=[2147483647, 2147483647, -85, 193,
55, 13], n3=[2147483647, 2147483647,
-90, 301, 50, 13], n4=[2147483647,
2147483647, -92, 123, 48, 13],
```

	Campaign 1	Campaign 2
Transportation	Car	Car
Users	234	234
Duration (h)	210.9	204.6
Distance (km)	6203.9	6152.5
Trajectory pairs	39,536	39,520

FIGURE 4. Illustration of data collection: a) (a) Data collection platform; b) Data collection devices; c) Data collection environment; d) Data collection dataset.

CASE STUDY

In this section, we carry out a case study to demonstrate the efficacy of the proposed *2DHFL* framework. Our study focuses on a map-matching service that relies on cellular association traces. When working with data from a single NSP, we often encounter limitations in terms of available features, resulting in suboptimal matching accuracy. In this specific case, we assess the performance enhancement achieved by employing collaborative computing involving three NSPs with dual-SIM-card users.

METHODOLOGY

Data Collection: Due to privacy concerns, direct access to NSPs' data is restricted. Therefore, we actively perform manual data collection to support our research, as shown in Fig. 4. Specifically, we design a real-time Android-based data collection App, depicted in Fig. 4a, and deploy 20 mobile phones[1] to capture sufficient research data. Subsequently, we conduct two extensive data collection campaigns, one for gathering cellular association traces of dual-SIM-card users between NSP A and NSP B, and the other for collecting cellular association traces of dual-SIM-card users between NSP B and NSP C. For simplicity, we assume no dual-SIM-card users existing between NSP A and NSP C. These campaigns share the same transportation mode and environments, encompassing urban, suburban, and highway settings, with varying BS densities and road conditions, as illustrated in Fig. 4c. Key statistics from two data collection campaigns are presented in Fig. 4d.

Model Design: In our considered three-NSP scenario with dual-SIM-card users, our *2DHFL* framework is implemented with one top server, two middle servers, and three local NSP participants. We have developed an innovative model for the map matching task, comprising a stacked bidirectional gated recurrent unit (SBi-GRU) encoder and an attentional GRU decoder, to learn the mapping function that projects cellular association traces to road segment-based trajectories. Specifically, each NSP implements the SBi-GRU encoder to extract hidden feature maps from its local cellular association traces and transmits these feature maps to its corresponding middle server. Subsequently, each middle server employs the attentional GRU decoder to decode combined feature maps into road segments and uploads its model parameter to top server. Finally, the top server establish its attentional GRU decoder to accomplish map matching based on the aggregated model parameters from the middle servers.

Baselines: For the comprehensive evaluation of the proposed *2DHFL*, we employ the following five baselines for performance comparison.

Hidden Markov Model (HMM) [13]: a traditional model-based method to solve the map matching problem.

Seq2Seq [14]: a widely-adopted data-driven model, commonly known as the Encoder-Decoder model, suitable for processing map matching problem.

DeepMM* [15]: an advanced deep learning approach specifically tailored for the map matching task, recognized for its notable effectiveness.

HFL: a feature-aligned FL framework where data features of participants are aligned, while training samples differ. In our scenario, NSP A and NSP C can adopt HFL framework.

VFL: a sample-aligned FL framework where participants have overlapping training sample IDs but less overlapping data features. In our scenario, the VFL framework can be adopted between pairs of NSPs sharing the same dual-SIM-card users, such as NSP A and NSP B, or NSP B and NSP C.

In summary, these five baselines are chosen from two perspectives. *HMM*, *Seq2Seq*, and *DeepMM** represent classical map-matching algorithms with single NSP data, assessing the effectiveness of collaborative computing among NSPs. *HFL* and *VFL* are two typical single FL frameworks, providing a benchmark to evaluate the superiority of our proposed *2DHFL* over traditional single FL frameworks.

Performance Metrics: To evaluate the effectiveness of these map-matching algorithms, we employ the following four performance metrics:

- *Precision:* the ratio of the total length of correctly matched routes to the total length of all routes.
- *Recall:* the ratio of the total length of the correctly matched routes to the total length of the routes in the ground truth.
- *F1-score:* a weighted average value of Precision and Recall, providing a balanced assessment of model performance.

---

[1] These 20 devices simulate the data collection of dual-SIM-card users, with 8 devices associated with NSP A and NSP C, and the remaining 12 devices associated with NSP B, as shown in Fig. 4b.

- *Time:* the average inference time, offering insights into the computational efficiency of the algorithms.

Experiment Setup: To ensure a realistic estimation of generalization error, we adopt a tripwise data splitting approach. Specifically, for each pair of NSPs with dual-SIM users, 80 percent of their respective private data is allocated for model training, with the remaining 20 percent reserved for model testing. Notably, the cellular association traces in the testing set are entirely distinct from those in the training set.

PERFORMANCE COMPARISON

Overall Performance: Figures 5a–d display box plots illustrating the distribution of scores for each algorithm across the metrics of Precision, Recall, F1-score, and Time, respectively. Across all critical metrics, *2DHFL* consistently outperforms baseline methods, demonstrating its superior efficacy. Figure 5e shows the average performance scores for all metrics, revealing three significant findings. Firstly, *2DHFL* consistently exhibits exceptional performance across all accuracy metrics, showcasing substantial improvements compared to baselines designed for single NSP data, as well as compared to single FL frameworks. For instance, in terms of the F1-score metric, *HMM*, *Seq2Seq*, *DeepMM**, *HFL*, and *VFL* achieve average scores of approximately 0.38, 0.46, 0.67, 0.54, and 0.86, while *2DHFL* attains an average score of 0.93, representing remarkable performance enhancements of 144.7 percent, 102.2 percent, 38.8 percent, 72.2 percent, and 8.1 percent, respectively. Secondly, collaborative computing among NSPs proves effective, and single *VFL* outperforms single *HFL*. Thirdly, regrading computational efficiency, *2DHFL* significantly outperforms HMM, albeit with slight performance gaps compared to the other four baselines.

Visualization: To enhance the clarity of our proposed *2DHFL*'s effectiveness, we use visualizations to depict the map matching performance of the six models. In Fig. 6, the ground truth of the route is displayed alongside the matched results of the three approaches designed for single-NSP data, the matched results of single HFL and VFL using two-NSP data for joint inference, and the matched result of our *2DHFL* using three-NSP data for joint inference. These visual representations emphasize that *2DHFL* consistently identifies the correct route with higher matching accuracy compared to those produced by the five baseline methods. Moreover, this visual evidence serves to confirm the effectiveness of our proposed multi-NSP collaborative computing framework.

CONCLUSION

In this article, we have discussed about the advantages, real-world applications, and the inherent challenges within the domain of collaborative computing among multiple NSPs. We have proposed a novel *2DHFL* framework for effective multi-NSP collaborative computing, while ensuring the stringent preservation of data privacy and security. This innovative framework integrates HFL and VFL to address the challenges of incomplete feature sets and limited training data. We have provided an insightful depiction of the driving forces that led to the conceptualization of this novel framework, highlighting the motivation behind its development. Additionally, we have presented a high-level overview of the detailed steps to implement our proposed *2DHFL* framework. To illustrate its effectiveness, we have presented a case study that demonstrates the real-world applicability of the *2DHFL* framework. Looking forward, our future research endeavors will delve into the optimization of communication protocols to enhance efficiency and address potential privacy vulnerabilities, further elevating the overall performance and data security of our framework.

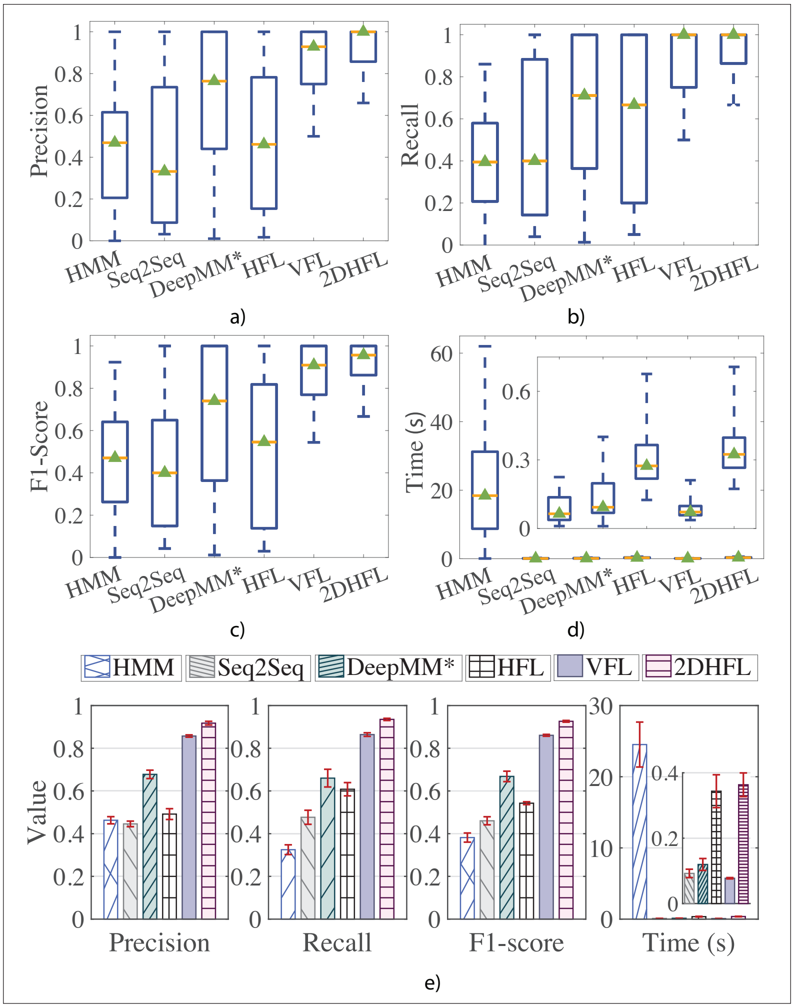


FIGURE 5. Overall performance comparison: a) Performance of precision; b) Performance of recall; c) Performance of F1-score; d) Performance of time; e) Average performance.

ACKNOWLEDGMENT

This research was supported in part by the National Key Research and Development (K&D) Program of China (2022YFF0604504), the National Natural Science Foundation of China (Grants No. 62320106006), the 111 Project (Grant No. B18059), the Central South University Innovafion-Driven Research Program (Grant No. 2023CXQD029), the Hunan Provincial Natural Science Foundation (Grant No. 2024JJ5268), the Hunan Province Education Science Planning Project (Grant No. XJK23ACJ004) and Natural Sciences and Engineering Research Council of Canada (NSERC) (Grant No. RGPIN-2023-03759).

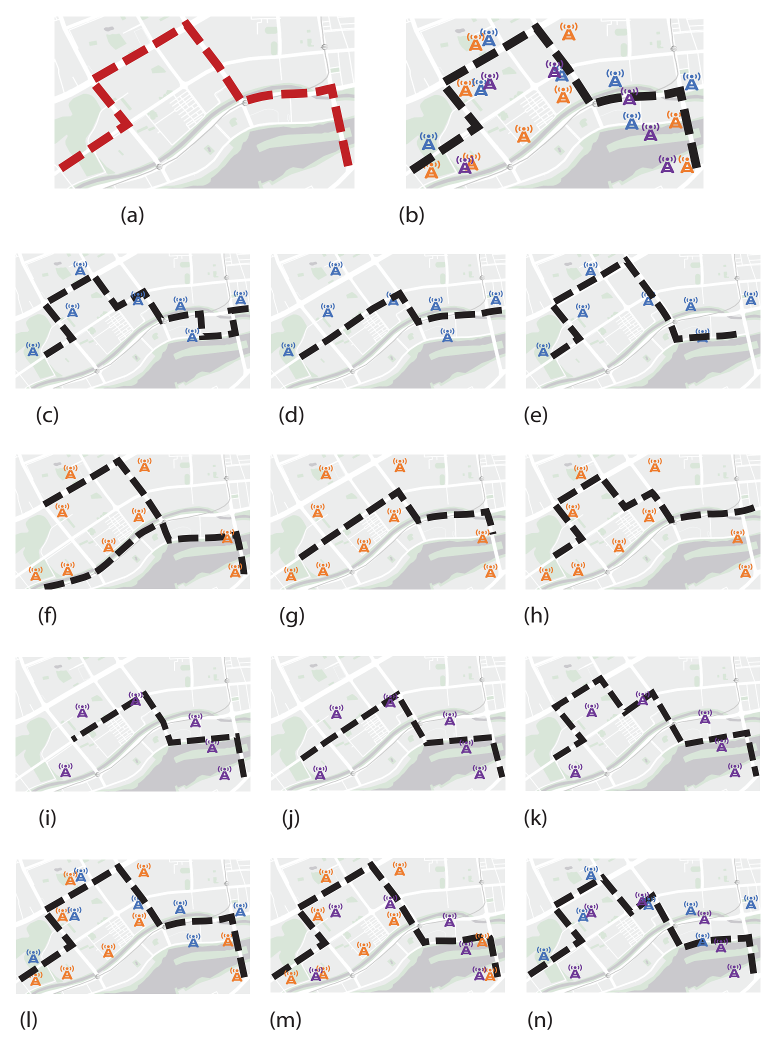


FIGURE 6. Visualization of map matching: a) Ground truth; b) 2DHFL + NSP A & B & C; c) HMM + NSP A; d) Seq2Seq + NSP A; e) DeepMM* + NSP A; f) HMM + NSP B; g) Seq2Seq + NSP B; h) DeepMM* + NSP B; i) HMM + NSP C; j) Seq2Seq + NSP C; k) DeepMM* + NSP C; l) VFL + NSP A & B; m) VFL + NSP B & C; n) HFL + NSP A & C.

REFERENCES

[1] W. Shi *et al.*, "Intelligent Reflection Enabling Technologies for Integrated and Green Internet-of-Everything Beyond 5G: Communication, Sensing, and Security," *IEEE Wirel. Commun.*, vol. 30, no. 2, 2023, pp. 147–54.

[2] S. Duan *et al.*, "Distributed Artificial Intelligence Empowered by End-Edge-Cloud Computing: A Survey," *IEEE Commun. Surv. Tutor.*, vol. 25, no. 1, 2022, pp. 591–624.

[3] Cnnic, "The 50th Statistical Report on China's Internet Development," China Internet Network Information Center, 2022.

[4] L. Xie *et al.*, "Collaborative Sensing in Perceptive Mobile Networks: Opportunities and Challenges," *IEEE Wirel. Commun.*, vol. 30, no. 1, 2023, pp. 16–23.

[5] K. H. Grantz *et al.*, "The Use of Mobile Phone Data to Inform Analysis of COVID-19 Pandemic Epidemiology," *Nat. Commun.*, vol. 11, no. 1, 2020, p. 4961.

[6] Z. Shen *et al.*, "DMM: Fast Map Matching for Cellular Data," *Proc. ACM MobiCom*, 2020.

[7] J. Huo *et al.*, "Short-Term Estimation and Prediction of Pedestrian Density in Urban Hot Spots Based on Mobile Phone Data," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, 2021, pp. 10,827–38.

[8] T. Reisch *et al.*, "Monitoring Supply Networks From Mobile Phone Data for Estimating the Systemic Risk of an Economy," *Scientific Reports*, vol. 12, no. 1, 2022, p. 13347.

[9] Z. Wang *et al.*, "Spatial-Temporal Cellular Traffic Prediction for 5G and Beyond: A Graph Neural Networks-Based Approach," *IEEE Trans Ind. Informat.*, 2022.

[10] Y. Deng *et al.*, "AUCTION: Automated and Quality-Aware Client Selection Framework for Efficient Federated Learning," *IEEE Trans. Parallel Distrib. Syst.*, vol. 33, no. 8, 2021, pp. 1996–2009.

[11] Z. Guo *et al.*, "Deep Federated Learning Enhanced Secure POI Microservices for Cyber-Physical Systems," *IEEE Wirel. Commun.*, vol. 29, no. 2, 2022, pp. 22–29.

[12] Y. Deng *et al.*, "Improving Federated Learning With Quality-Aware User Incentive and Auto-Weighted Model Aggregation," *IEEE Trans. Parallel Distrib. Syst.*, vol. 33, no. 12, 2022, pp. 4515–29.

[13] P. Newson and J. Krumm, "Hidden Markov Map Matching Through Noise and Sparseness," *Proc. ACM SIGSPATIAL*, 2009.

[14] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to Sequence Learning with Neural Networks," *Proc. NeurIPS*, 2014.

[15] J. Feng *et al.*, "DeepMM: Deep Learning Based Map Matching with Data Augmentation," *IEEE Trans. Mob. Comput.*, 2020.

BIOGRAPHIES

HUALI LU [S'20] received her B.Sc. and M.Sc. degrees in College of Computer Science and Electronic Engineering from Hunan University, Changsha, China, in 2017 and 2020. She is currently working toward the Ph.D. degree in School of Computer Science and Engineering, Central South University, Changsha, China. Her researches mainly focus on spatial-temporal data mining, compact data collection, trajectory similarity computing, and data synthesis.

WENXIONG CHEN received B.Sc. in communication engineering from Hunan Normal University, China and M.Sc. in MTI from Hunan University China, in 2007 and 2011, respectively, and Ph.D. in Management Science and Engineering in 2021 from Central South University China. He has been an associate professor at Hunan Normal University China since 2022. His research interest is artificial intelligence, big data, Internet of Things, mobile computing.

CONGHAO ZHOU [S'19, M'22] received the Ph.D. degree in Electrical and Computer Engineering from the University of Waterloo, Waterloo, ON, Canada, in 2022. He received the B.Eng. degree from Northeastern University, Shenyang, China, in 2017, and the M.Sc. degree from the University of Illinois Chicago, Chicago, IL, USA, in 2018. He is currently a Postdoctoral Fellow at University of Waterloo, Waterloo, ON, Canada. His research interests include space-air-ground integrated networks, network slicing, and machine learning for wireless networks.

HUAQING WU [S'15, M'21] received the Ph.D. degree from the University of Waterloo, Ontario, Canada, in 2021, and the B.E. and M.E. degrees from Beijing University of Posts and Telecommunications, Beijing, China, in 2014 and 2017, respectively. She is currently an assistant professor with the Department of Electrical and Software Engineering, University of Calgary, Alberta, Canada. Her current research interests include B5G/6G, space-air-ground integrated networks, Internet of vehicles, edge computing/ caching, and artificial intelligence (AI) for future networking.

FENG LYU [M'18, SM'22] received the Ph.D. degree from the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China, in 2018, and the BS degree in software engineering from Central South University, Changsha, China, in 2013. He is currently a professor with the School of Computer Science and Engineering, Central South University, Changsha, China. His research interests include vehicular networks, beyond 5G networks, big data measurement and application design, and edge computing.

XUEMIN (SHERMAN) SHEN [M'97, SM'02, F'09] received the Ph.D. degree in electrical engineering from Rutgers University, New Brunswick, NJ, USA, in 1990. He is a University Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research focuses on network resource management, wireless network security, Internet of Things, 5G and beyond, and vehicular networks.