The Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-24)

HiFi-Gas: Hierarchical Federated Learning Incentive Mechanism Enhanced Gas Usage Estimation

Hao Sun[1*], **Xiaoli Tang**[2*], **Chengyi Yang**[1*], **Zhenpeng Yu**[1], **Xiuli Wang**[1], **Qijie Ding**[1],
Zengxiang Li[1], **Han Yu**[2]
[1]ENN Group, Beijing, China
[2]School of Computer Science and Engineering, Nanyang Technological University, Singapore
lizengxiang@enn.cn, han.yu@ntu.edu.sg

Abstract

Gas usage estimation plays a critical role in various aspects of the power generation and delivery business, including budgeting, resource planning, and environmental preservation. Federated Learning (FL) has demonstrated its potential in enhancing the accuracy and reliability of gas usage estimation by enabling distributedly owned data to be leveraged, while ensuring privacy and confidentiality. However, to effectively motivate stakeholders to contribute their high-quality local data and computational resources for this purpose, incentive mechanism design is key. In this paper, we report our experience designing and deploying the Hierarchical FL Incentive mechanism for the Gas usage estimation (`HiFi-Gas`) system. It is designed to cater to the unique structure of gas companies and their affiliated heating stations. `HiFi-Gas` provides effective incentivization in a hierarchical federated learning framework that consists of a horizontal federated learning (HFL) component for effective collaboration among gas companies and multiple vertical federated learning (VFL) components for the gas company and its affiliated heating stations. To motivate active participation and ensure fairness among gas companies and heating stations, we incorporate a multi-dimensional contribution-aware reward distribution function that considers both data quality and model contributions. Since its deployment in the ENN Group in December 2022, `HiFi-Gas` has successfully provided incentives for gas companies and heating stations to actively participate in FL training, resulting in more than 12% higher average gas usage estimation accuracy and substantial gas procurement cost savings. This implementation marks the first successful deployment of a hierarchical FL incentive approach in the energy industry.

Introduction

Gas usage estimation (Liu et al. 2021; Sharma et al. 2021) is vital for the safe and efficient operation of electric grids. It provides essential information for budgeting, resource planning and managing environmental considerations. Its impact can be felt by gas companies, utility service providers and even homeowners. Accurate estimation allows individuals and organizations to anticipate energy costs, make informed decisions on energy efficiency and identify conservation opportunities, aligning with sustainability goals and carbon footprint reduction (Le Quéré et al. 2020).

Traditional approaches for gas usage estimation based on historical data and statistical models face limitations due to data sparsity and subjective input parameters (Soldo 2012; Yu et al. 2014; Tamba et al. 2018). To improve accuracy and reliability, energy companies are turning to data-driven techniques, such as machine learning (ML). ML algorithms can analyze vast data volumes, identify patterns and forecast gas consumption more precisely, leveraging historical consumption data, weather patterns and other factors.

Collaborative model training techniques, such as federated learning (FL) (Yang et al. 2019, 2020; Goebel et al. 2023), have emerged to enhance ML solutions by leveraging distributedly owned data resources from multiple organizations (Warnat-Herresthal et al. 2021; Chen et al. 2023). FL (Liu et al. 2020, 2022b) enables collaborative training of models across distributed data sources without centralizing sensitive data, ensuring privacy and confidentiality. In this way, energy companies can leverage data from various sources, including gas companies, heating stations and consumers, to build more accurate and robust gas usage estimation models.

Despite their advantages, FL-based gas usage estimation methods face significant challenges that hinder their widespread adoption in practice (Zeng et al. 2021; Tu et al. 2022; Cheng, Li, and Liu 2022; Khan et al. 2020; Zhan et al. 2021). Firstly, it is challenging to effectively motivate self-interested data owners (a.k.a., FL clients) to contribute their computation and communication resources to FL model training. Economic compensation mechanisms are necessary to incentivize them to participate meaningfully. Secondly, ensuring the integrity and quality of FL client contributions is vital. As unreliable clients can negatively impact the performance of the global FL model through malicious behavior or low-quality updates. To overcome these challenges, FL incentive mechanism design has emerged as an important research field to identify the optimal payment and organizational structures within federated networks, with the goal of achieving desired operational objectives (Khan et al. 2020; Zhan et al. 2021).

To address these challenges, we propose the Hierarchical FL Incentive mechanism for Gas usage estimation (`HiFi-Gas`). The key features and benefits of `HiFi-Gas`

---

[*]These authors contributed equally.

are as follows:

- **Customized Collaboration:** `HiFi-Gas` is built upon a hierarchical FL framework that comprises a horizontal federated learning (HFL) component and multiple vertical federated learning (VFL) components. The HFL component enables ENN gas companies to leverage external entities' historical gas supply data and weather information for accurate gas usage estimation. The VFL components facilitate collaboration among each gas company and their affiliated heating stations within their coverage, taking into account the vertical partitioning of data among them.
- **Active Participation and Fairness:** `HiFi-Gas` incorporates a multi-dimensional reward distribution function that considers both data quality and model contributions. This approach promotes active participation and ensures fairness among gas companies and heating stations, incentivizing them to actively contribute their resources and expertise to the FL process.
- **Enhanced Gas Usage Estimation with Privacy Preservation:** `HiFi-Gas` effectively motivates participants to join FL-based gas usage estimation by providing them with monetary compensation. This incentivization mechanism facilitates the improvement of gas usage estimation model accuracy, while safeguarding the privacy and confidentiality of participants' data.

By leveraging the hierarchical FL framework, incorporating a fair incentive and preserving data privacy, `HiFi-Gas` offers a promising solution to enhance collaborative gas usage estimation in practice.

The `HiFi-Gas` approach has been successfully deployed in ENN Group[1] since December 2022 in a province in China through collaboration among the ENN Group and the Trustworthy Federated Ubiquitous Learning (TrustFUL) Lab[2], Nanyang Technological University (NTU), Singapore. Since its deployment, `HiFi-Gas` has successfully provided incentives for gas companies and heating stations to actively participate in FL training, resulting in more than 12% higher average gas usage estimation accuracy and substantial gas procurement cost savings. To the best of our knowledge, it is the first successfully deployed hierarchical federated learning incentive approach in the energy industry.

Application Description

ENN Group, established in 1989, is a leading player in the natural gas and green energy industry with diversified operations in distribution, trade, transportation and storage, production, and intelligent engineering. The company is dedicated to creating modern energy systems that enhance people's quality of life. With its robust industrial ecosystem, ENN is actively developing an industrial digital intelligence platform to empower various stakeholders, including over 25 million household customers and more than 200,000 enterprise customers across 20 provinces in China.

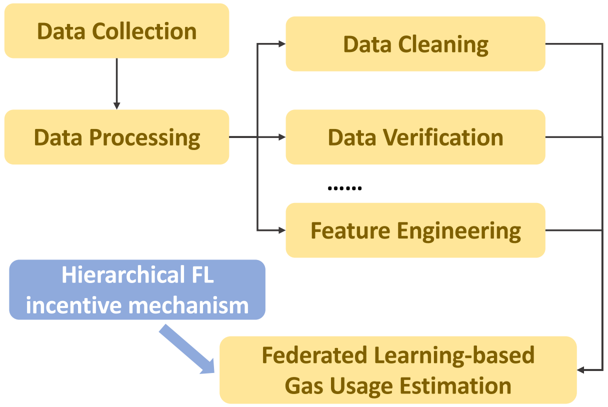


Figure 1: An overview of the FL-based gas usage estimation process in the ENN Group.

Due to the higher likelihood of rising natural gas prices in winter, sub-companies of the ENN Group need to procure natural gas in advance, which requires them to make predictions regarding the amount of gas required for a specified period in advance. However, as the volume of data owned by each sub-company is sparse and potentially biased, it is difficult for them to make accurate predictions on their own. Therefore, an FL-based platform has been built to involve all the sub-companies collaboratively to train the estimation model, while protecting their data privacy. In this section, we offer an overview of the ENN gas usage estimation platform which is built based on FL.

Figure 1 illustrates the ENN gas usage estimation model overall workflow. It is an integral part of its industrial digital intelligence platform. Initially, gas usage data are collected and stored locally in data silos. To prepare the data for analysis and model training, standard techniques (e.g., Fourier transform, wavelet transform) are applied. Subsequently, the gas estimation model is trained using the FL paradigm, in collaboration with gas companies and heating stations. The FL model training subsystem consists of two components: 1) the ENN FL Server, and 2) the ENN FL Client (Chen et al. 2023) as shown in Figure 2. The incentive model deployed in this subsystem aims to ensure the efficient functioning of the FL ecosystem.

ENN FL Server

The ENN FL Server is a critical component comprising the FL model aggregation server and utility modules supporting system administrators. It offers a variety of FL model training and aggregation approaches, including FedAvg (McMahan et al. 2017) and FedOBD (Chen et al. 2023), which can be integrated into the system. System administrators can easily configure the FL model training process using dedicated user interfaces. After completing the configuration steps, the information is transmitted to the FL model aggregation server via a local area network (LAN).

Designed with a modular approach, the FL model aggregation server can support different FL model aggregation

---

[1] https://www.enn.cn/

[2] https://trustful.federated-learning.org/

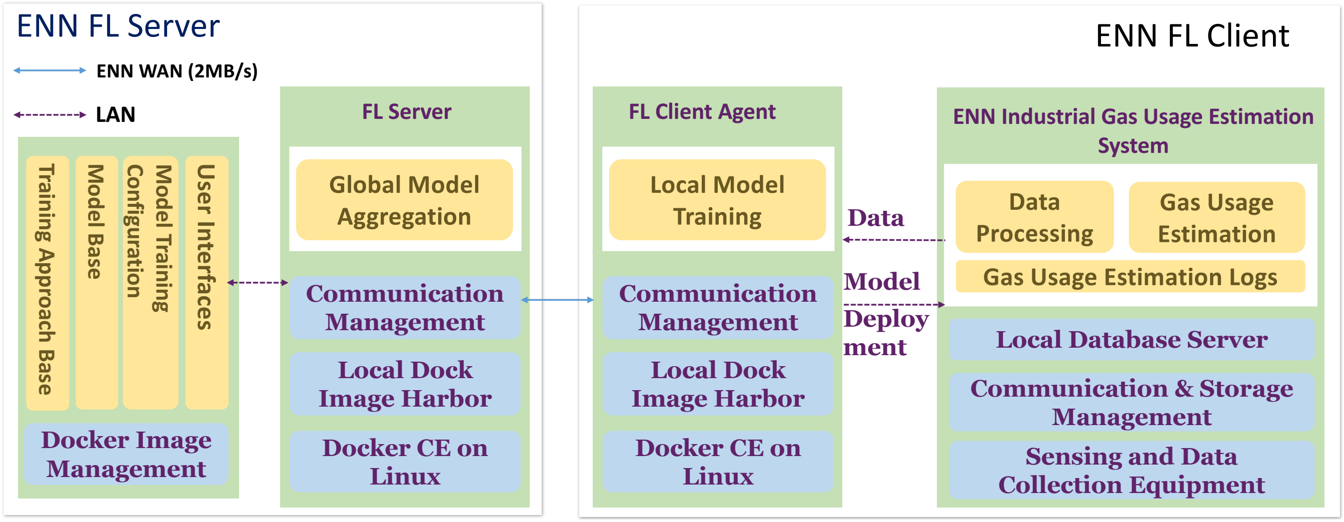


Figure 2: An overview of the ENN FL model training platform.

methods. It takes local model updates from FL clients as inputs and generates a global FL model as output. In addition, the server determines if further rounds of FL training are necessary to enhance global model accuracy. Efficient communication with clients is facilitated by the Communication Management module, which can accommodate special requirements like transmission compression using stochastic quantization (Alistarh et al. 2017).

Figures 3 and 4 depict the user interfaces (UIs) of the ENN FL model training platform, which visualize the FL training process for Chinese-speaking system administrators. We have annotated key features for clarity. Figure 3 displays the screen that allows administrators to configure the FL training process for a specific model by defining critical parameters, such as the mode of federated learning and the chosen model training and aggregation approach. Once FL training begins, the training activities are visualized in Figure 4, enabling administrators to monitor progress. The example shows one FL server and two FL clients, each denoted by their respective activities listed in the relevant boxes. Users can easily navigate through the training progress and access detailed information for each record. The left-hand side panel features an FL training flowchart, illustrating the overall progress. Upon completion of training, administrators can access a summary of the performance of the resulting FL model. The ENN FL platform stores detailed FL model training activities and performance evaluation results, facilitating future reviews and audits.

ENN FL Client

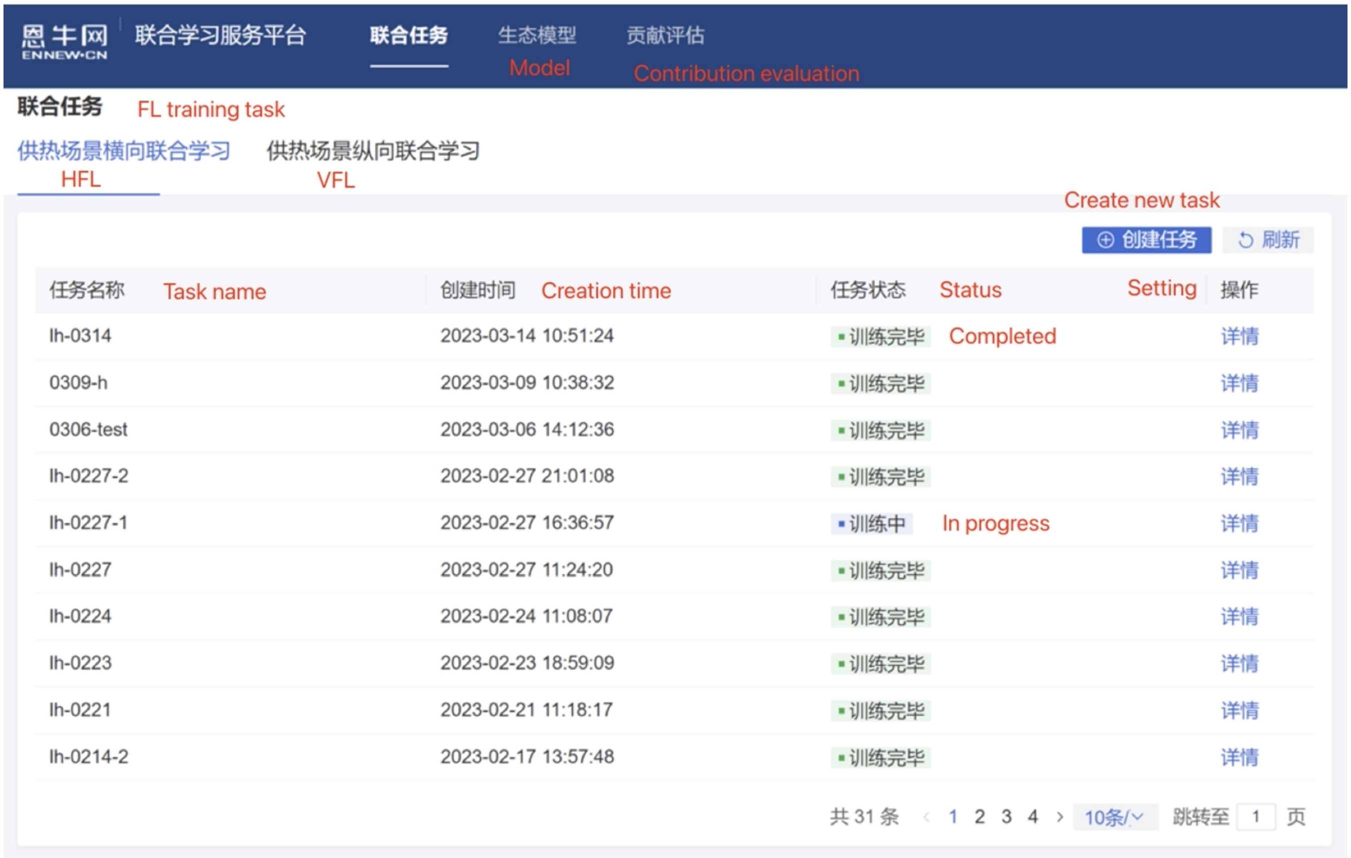
ENN FL clients are commonly deployed in various industrial facilities, including factories, power plants, and coal chemical plants. These facilities use sensors to monitor equipment and collect relevant data, which is aggregated and stored in the appropriate format by the Communication & Storage Management module in the Local Database Server. The ENN gas usage estimation system conducts data and signal processing on this infrastructure and employs a Deep Neural Network (DNN)-based model to predict gas demand. To utilize historical gas usage information and weather data from different facilities, the DNN-based model is trained through FL. Each ENN FL Client is equipped with an FL Client Agent module, facilitating the facility's data participation in the FL process.

Use of AI Technology

In this section, we provide details of the AI Engine of the ENN FL model training subsystem for gas usage estimation built on the `HiFi-Gas` approach. It operates within a hierarchical federated learning ecosystem and aims to distribute rewards fairly to participants, thereby motivating them to actively engage in FL training and improve the performance of the resulting gas usage estimation model.

Two types of entities are involved in the ENN gas supply chain: 1) gas companies, and 2) heating stations. Gas companies acquire natural gas from external sources and distribute it to heating stations affiliated to them. However, both types of entities face challenges in making accurate predictions based solely on their own local data. Gas companies face data sparsity issues, making it difficult to train precise gas demand prediction models. Heating stations heavily rely on heating strategists to manually create daily heating plans and estimate gas usage based on weather forecasts. However, the subjectivity involved in formulating strategies manually and the lack of access to high-precision weather forecast data negatively impact the effectiveness of such plans.

To address these challenges, our solution employs two FL ecosystems: 1) an HFL system among gas companies (Figure 5), and 2) multiple VFL systems between each gas company and the affiliated heating stations (Figure 6). In the client-server HFL system, there are generally n^H FL clients suitable for FL-based gas usage estimation model training. Each client i possesses a local dataset $D_i^H = (\mathbf{x}_j^H, y_j^H)_{j=1}^{|D_i^H|}$,



任务名称 Task name	创建时间 Creation time	任务状态 Status	Setting 操作
lh-0314	2023-03-14 10:51:24	训练完毕 Completed	详情
0309-h	2023-03-09 10:38:32	训练完毕	详情
0306-test	2023-03-06 14:12:36	训练完毕	详情
lh-0227-2	2023-02-27 21:01:08	训练完毕	详情
lh-0227-1	2023-02-27 16:36:57	训练中 In progress	详情
lh-0227	2023-02-27 11:24:20	训练完毕	详情
lh-0224	2023-02-24 11:08:07	训练完毕	详情
lh-0223	2023-02-23 18:59:09	训练完毕	详情
lh-0221	2023-02-21 11:18:17	训练完毕	详情
lh-0214-2	2023-02-17 13:57:48	训练完毕	详情

Figure 3: An example of the user interface for configuring FL training in the ENN FL platform.

where $\mathbf{x}_j^H$ represents the j-th local training sample, and y_j^H denotes the corresponding ground truth label of $\mathbf{x}_j^H$. $|D_i^H|$ denotes the total number of data samples in D_i^H. The goal of the HFL system is to solve the following optimization problem under the aforementioned setting:

$$\min_{\theta^H} \sum_{i=1}^{n^H} \frac{|D_i^H|}{|D^H|} \mathcal{L}_i^H(\theta^H; D_i^H), \qquad (1)$$

where θ^H represents the model parameters. $|D^H|$ denotes the total number of samples, calculated as the sum of the sample sizes of all clients (i.e., $|D^H| = \sum_{i=1}^{n^H} |D_i^H|$). $\mathcal{L}_i^H(\theta^H; D_i^H)$ represents the local loss of a specific client i, defined as the average loss over its local dataset, given by $\mathcal{L}i^H(\theta^H; D_i^H) = \frac{1}{|D_i^H|} \sum_{j=1}^{|D_i^H|} l(\theta^H; \mathbf{x}_j^H, y_j^H)$. Here, $l(\theta^H; \mathbf{x}_j^H, y_j^H)$ denotes the loss function for the model with parameters θ^H when predicting label y_j^H for input $\mathbf{x}_j^H$.

Unlike HFL, which partitions data by samples, VFL partitions data by features. Let n^V denote the number of participants in VFL model training. The dataset $D^V = \{\mathbf{x}_j^V, y_j^V\}_{j=1}^{|D^V|}$ is partitioned across the n^V participants, each of which is associated with a unique set of features. For example, consider participant i, who maintains the i-th block of features $\mathbf{x}_{j,i}$ for the j-th sample $\mathbf{j} = [x_{j,1}^T, \cdots, x_{j,n^V}^T]^T$. Each participant i conducts training on the model parameter θ^V utilizing its local raw features $\mathbf{x}_{j,i}$, with the goal of minimizing the loss function:

$$\min_{\theta^V} \frac{1}{n^V} \sum_{j=1}^{n^V} \mathcal{L}^V(\theta^V; D_j^V). \qquad (2)$$

The `HiFi-Gas` Workflow

Data Quality Evaluation The first step is for each participant to preprocess its local data before initiating FL model training. Specifically, in the case of HFL, historical gas usage and weather conditions significantly impact current gas usage. Thus, each participant can evaluate its data quality by analyzing the correlation between historical gas usage data and weather data with the actual gas usage data. Let $\mathbf{X}_i$ denote the historical gas usage and weather data from participant i for a continuous period of T days, and let $\mathbf{Y}_i$ denote the corresponding actual gas usage data. The correlation between $\mathbf{X}_i$ and $\mathbf{Y}_i$ can be formulated as:

$$\text{corr\_score}_i^H = \frac{Cov(\mathbf{X}_i, \mathbf{Y}_i)}{\sqrt{Var(\mathbf{X}_i) Var(\mathbf{Y}_i)}}, \qquad (3)$$

where the supscript H means that the equation is applicable for the HFL setting. Cov is the covariance function (Rice 2006) and Var is the variance function (Breiman 2001).

月度评估详情 监控周期：2023-01-15 至2023-02-14

生态圈数据 Data

参与方 Participant	供热站数量	数据量
横向 [redacted] 城燃	4	
横向 [redacted] 城燃	46	

精度统计 Accuracy

激励机制贡献评估 Contribution evaluation

阶段	横向 [redacted] 城燃	横向 [redacted] 城燃
数据质量评估价值分配占比 Data quantity evaluation	5.48%	94.52%
模型贡献评估价值分配占比 Model contribution evaluation	33.59%	66.41%

Figure 4: An example of the user interface for visualizing the FL training process in the ENN FL platform.

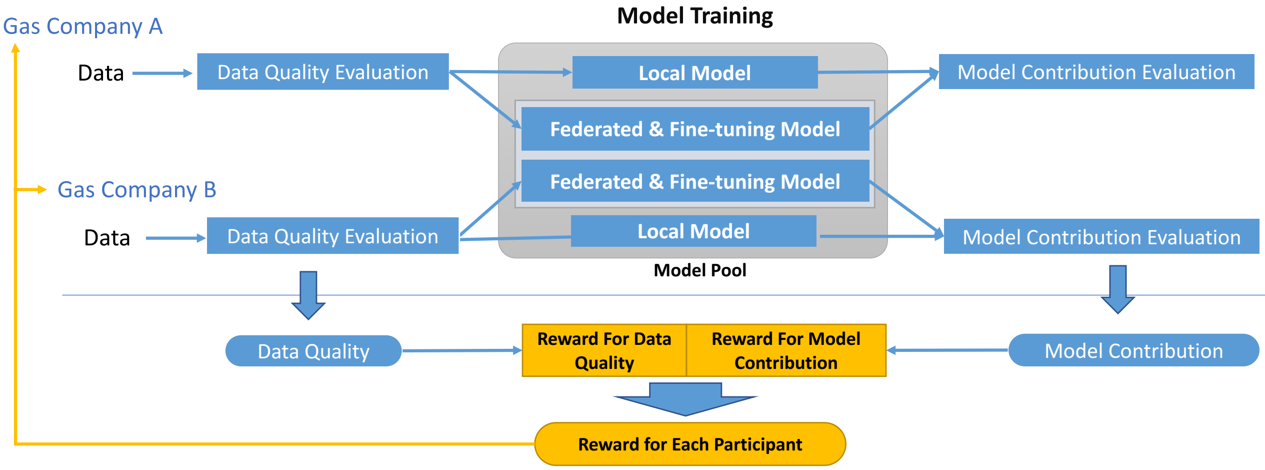


Figure 5: The `HiFi-Gas` incentive model for HFL scenarios.

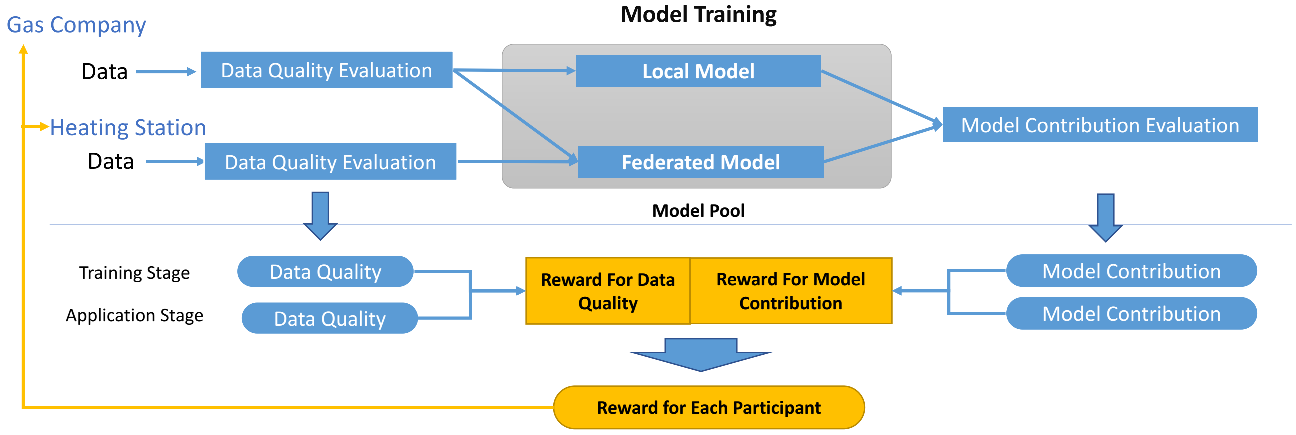


Figure 6: The `HiFi-Gas` incentive model for VFL scenarios.

The data quantity weightage score for participant i can be computed as:

$$\text{quant\_score}_i^H = \frac{|D_i^H|}{|D^H|}, \tag{4}$$

where $|D_i^H|$ and $|D^H|$ are the total number of samples in participant i and in the HFL ecosystem, respectively. Then, the final data quality evaluation result for participant i under the HFL setting is defined as:

$$\text{quality}_i^H = \text{corr\_score}_i^H \times \text{quant\_score}_i^H, \tag{5}$$

whereas under the VFL setting, the data quality of participant i is defined as:

$$\text{quality}_i^V = \text{corr\_score}_i^V. \tag{6}$$

Model Contribution Evaluation The second step is to assess participants' contributions from a model perspective. Each FL participant trains a local model using its own data as depicted in Figures 5 and 6. Subsequently, each of them employs its local model to predict gas usage for a consecutive period of T days. By comparing these predictions with the actual gas usage, the adapted Symmetric Mean Absolute Percentage Error (SMAPE) (Hyndman and Koehler 2006) can be calculated as:

$$\text{SMAPE}_i^{local} = \frac{1}{T}\sum_{t=1}^{T} \frac{|y_{i,t}^{local} - \hat{y}_{i,t}|}{(|y_{i,t}^{local}| + |\hat{y}_{i,t}|)}, \tag{7}$$

where $\hat{y}_{i,t}$ represents the actual gas usage at time t of participant i. $y_{i,t}^{local}$ represents the prediction generated by the local model of participant i at the same time t. T is the number of time periods.

Similarly, the prediction error of the global FL model adapted for gas usage estimation for participant i is:

$$\text{SMAPE\_new}_i^{global} = \frac{1}{T}\sum_{t=1}^{T} \frac{|y_{i,t}^{global} - \hat{y}_{i,t}|}{(|y_{i,t}^{global}| + |\hat{y}_{i,t}|)}, \tag{8}$$

where $y_{i,t}^{global}$ is the predicted gas usage of i at time t generated by the global model.

Then, the contribution of participant j to the global FL model can be obtained as:

$$\text{contribution}_j = \sum_{i \neq j} \frac{\text{SMAPE\_new}_i^{local} - \text{SMAPE\_new}_i^{global}}{n-1}. \tag{9}$$

n is the number of participants in FL model training.

Reward Distribution Let R_{data} and R_{model} denote the total reward available for data quality and model contribution, respectively. The data quality reward and model contribution reward for each participant i can be computed as:

$$\text{r}_i^{quality} = R_{data} \times \frac{\text{quality}_i}{\sum_{j=1}^{n} \text{quality}_j}, \tag{10}$$

$$\text{r}_i^{contribution} = R_{model} \times \frac{\text{contribution}_i}{\sum_{j=1}^{n} \text{contribution}_j}. \tag{11}$$

Application Development and Deployment

The `HiFi-Gas` was developed by teams from the TrustFUL Lab and the Digital Research Institute of the ENN Group, utilizing the PyTorch framework (Paszke et al. 2017) as the foundation. Prior to the implementation of the AI Engine, a comprehensive evaluation was conducted comparing it to four cutting-edge contribution evaluation approaches:

1. **GTG-Shapely** (GTG-S) (Liu et al. 2022a): It reconstructs FL models from gradient updates for SV calculation, avoiding repeated training. It uses guided Monte Carlo sampling, along with truncation techniques, to efficiently reduce model reconstructions and evaluations.
2. **WT-Shapley** (WT-S) (Yang et al. 2022a): This method enhances the GTG-Shapley approach by incorporating weight calculations during truncation operations.
3. **WTDP-Shapely** (WTDP-S) (Yang et al. 2022b): This method enhances WT-Shapley by employing Dynamic Programming instead of random sampling for accurate utility value estimation during SV calculation.
4. **MC**: In this approach, the contribution of each participant is computed by considering its impact on the performance of all other participants.

In the offline experimental evaluation to make a decision on which approach to adopt for deployment, we utilize the ETTh1 dataset, one of the hourly Electricity Transformer Temperature (ETT) datasets (Zhou et al. 2021). The ETTh1 dataset encompasses temperature measurements from diverse electric power transformers and contains a total of 17,420 chronological records. During the offline experiments, we designate the most recent 5% of records as the standardized validation set. We employ a total of 5 clients in our experiments. Each client's model architecture includes 2 LSTM layers followed by 1 linear layer. The input sequence length and prediction sequence length are both set to 96.

To comprehensively evaluate the efficiency and performance of `HiFi-Gas` in comparison to four baselines across diverse FL scenarios, the following two distinct scenarios were designed:

- **Different Data Quantities without Noise**: Participants contribute data in varying quantities, ordered chronologically. The data proportions for each participant are in the ratios of 1:2:3:4:5.
- **Equal Data Quantity with Noise**: All participants contribute the same amount of data, but their individual data includes diverse proportions of Gaussian noise. The original values are used as the mean (μ) of the distribution, while the standard deviations (σ) are set to 0.1, 0.2, 0.3, 0.4, and 0.5 for the respective participants.

The performance of the comparison approaches is assessed using the following evaluation metrics:

- **Time**: The efficiency of each approach is assessed based on the total time taken to calculate the contributions.
- **Cosine Distance**: We represent the SV result calculated by the MR method (Song, Tong, and Wei 2019) as a vector $\phi^* = \langle \phi_1^*, \cdots, \phi_n^* \rangle$, and the estimation result obtained by any other approach as $\phi = \langle \phi_1, \cdots, \phi_n \rangle$. Then, the Cosine Distance is formulated as:

$$1 - \cos(\phi^*, \phi). \tag{12}$$

- **Euclidean Distance**: similar to the Cosine Distance, the Euclidean Distance is computed as:

$$\sqrt{\sum_{i=1}^{n} (\phi_i^*, \phi_i)^2}. \tag{13}$$

- **Maximum Difference**: it is computed as:

$$\max_{i=1}^{n} |\phi_i^*, \phi_i|. \tag{14}$$

The smaller the values of these metrics, the better the performance of an approach.

The performance results of the baselines and `HiFi-Gas` are presented in Table **??** for the different data quantities without noise FL setting, and in Table **??** for the equal data quantity with noise FL setting. It is evident that the Shapley value-based variants, such as GTG-Shapley, WT-Shapley, and WTDP-Shapley, outperform the simpler marginal contribution-based methods, including MC and `HiFi-Gas`, in terms of cosine distance, Euclidean distance, and maximum difference. However, MC and `HiFi-Gas` exhibit significantly superior performance when it comes to the time required for contribution evaluation. In comparison to the top-performing Shapley value variants, MC and `HiFi-Gas` demonstrate a reduction in time by over 738%, rendering them more suitable for practical deployment in industry settings. In the context of time efficiency, MC and `HiFi-Gas` exhibit similar performance. Nonetheless, in terms of cosine distance, Euclidean distance, and maximum difference, `HiFi-Gas` outperforms MC significantly. After jointly considering the overall performance of all approaches in terms of both efficiency and the accuracy of contribution evaluation, the decision was made to adopt `HiFi-Gas` for deployment.

	Time (sec)	Cosine Distance	Euclidean Distance	Maximum Difference
GTG-S	43.66	5.53e-3	4.75e-2	3.58e-2
WT-S	43.74	4.48e-3	4.28e-2	3.13e-2
WTDP-S	43.53	4.08e-6	1.30e-3	8.97e-4
MC	5.76	1.31e-1	2.47e-1	1.85e-1
`HiFi-Gas`	5.76	2.69e-2	1.05e-1	7.43e-2

Table 1: Performance comparison under different data quantities without noise.

	Time (sec)	Cosine Distance	Euclidean Distance	Maximum Difference
GTG-S	156.27	2.03e-3	4.77e-2	3.26e-2
WT-S	156.30	1.30e-3	3.85e-2	2.67e-2
WTDP-S	53.03	9.44e-4	2.78e-2	1.89e-2
MC	5.76	1.70e-1	3.83e-1	2.86e-1
`HiFi-Gas`	5.76	9.74e-2	2.58e-1	2.12e-2

Table 2: Performance comparison under equal data quantity with noise.

Application Use and Payoff

Starting from December 2022, `HiFi-Gas` has been integrated into the ENN Intelligent Gas Usage Estimation Platform for deployment. It has been adopted by two major gas companies in the ENN Group located in a province in China. For ease of reference and respect for the industry partner's request not to disclose the names of these companies, we refer to them as Company A and Company B in the rest of this paper. Company A has 4 affiliated heating stations, while Company B has 47 affiliated heating stations.

As highlighted previously, the ultimate objective of `HiFi-Gas` is to enhance the precision of gas usage estimation models for both gas companies and the affiliated heating stations within the FL ecosystem. This section provides an overview of the impact of `HiFi-Gas` on accuracy enhancements for Company A and Company B since its deployment.

Table **??** illustrates the remarkable enhancements in the accuracy of the resulting FL gas usage estimation model collaboratively trained by companies A and B under `HiFi-Gas`. Specifically, the average gas usage estimation accuracy for Company A and its affiliated heating stations has been improved by 14.67%, with model accuracy increasing from 81.1% before the deployment of `HiFi-Gas` to 93.0% afterwards. Similarly, for Company B and its affiliated heating stations, the average gas usage estimation accuracy has been improved by 10.31%, with model accuracy increasing from 78.6% before the deployment of `HiFi-Gas` to 86.7% afterwards.

`HiFi-Gas` is designed to encourage responsible participation in the federated model training through the sharing of high-quality local data. Notably, the average daily data quantity is reduced under `HiFi-Gas` by 17.61%. This is because the `HiFi-Gas` incentive mechanism effectively compels the participating entities to refine their local datasets more meticulously in order to remove noisy and erroneous data prior to engaging in federated model training. Consequently, these entities can elevate their training accuracy, leading to a substantial reduction in gas procurement costs. This outcome underscores the efficacy of the `HiFi-Gas` incentive mechanism.

Maintenance

The AI Engine adopts a modular architectural design, ensuring a clear separation of concerns and segregation of responsibilities. Throughout the deployment phase, there has been no AI maintenance activity required, even with changes in personnel authorization privileges and operational configurations within the system.

Lessons Learned During Deployment

During the deployment of `HiFi-Gas` for gas usage estimation, several valuable lessons were learned, providing insights into the effective implementation of FL incentive mechanisms. Here are some key takeaways:

1. **Clear Communication and Stakeholder Engagement**: Effective communication and active engagement with all stakeholders, including gas companies and heating stations, are crucial. Regular engagement fosters collaboration, feedback and continued improvement throughout the deployment process.
2. **Alignment of Incentives with Business Objectives**: The incentive mechanism should align with stakeholders' business objectives related to gas usage estimation. Structuring incentives to motivate accurate data contribution ensures their active participation.
3. **Data Quality Assurance**: Ensuring data quality from gas companies and heating stations is essential. Implementing robust data validation processes and conducting regular audits improve the accuracy and reliability of FL-based gas usage estimation.
4. **Continuous Monitoring and Evaluation**: Regularly monitoring and evaluating the efficacy of the incentive mechanism helps identify areas for improvement and ensures its effectiveness over time.
5. **Flexibility and Adaptability**: Designing the incentive mechanism with flexibility and adaptability allows it to accommodate changes in industry practices and regulations.

	Acc. w/o `HiFi-Gas`	Acc. with `HiFi-Gas`	%Δ Acc.
Company A	81.1%	**93.0%**	+14.67%
Company B	78.6%	**86.7%**	+10.31%

Table 3: Deployment results for `HiFi-Gas` for gas companies A and B.

Considering these lessons learned, the deployment of `HiFi-Gas` in the industry can lead to improved gas usage estimation accuracy, reliability and efficiency, promoting collaboration among stakeholders and enhancing the decision-making process.

Conclusions and Future Work

In this paper, we present our experience in designing and deploying a reward-based system for incentivizing federated learning participants to engage in the FL model training process with the goal to enhance the accuracy of gas usage estimation. We propose `HiFi-Gas`, a hierarchical incentive mechanism for federated learning that incorporates a reward distribution module tailored to the specific structure of gas companies and their associated heating stations. Since its deployment in the ENN Group AI Engine in a province in China, `HiFi-Gas` has significantly improved gas usage forecasting accuracy for two gas companies located in different cities. It has effectively motivated gas companies and heating stations to actively contribute to FL training and provide high-quality data, resulting in increased revenue for these entities compared to previously adopted practices.

Moving forward, our future work will involve a comprehensive evaluation of the `HiFi-Gas` using larger industrial datasets from various perspectives. In addition, we aim to enhance the resilience of the incentive mechanism against potential malicious participants (Lyu et al. 2022) and further advance its fairness (Shi, Yu, and Leung 2023; Yu et al. 2016, 2017) when distributing incentive benefits.

Acknowledgments

This research/project is supported by the National Research Foundation Singapore and DSO National Laboratories under the AI Singapore Programme (AISG Award No: AISG2-RP-2020-019); and the RIE 2020 Advanced Manufacturing and Engineering (AME) Programmatic Fund (No. A20G8b0102), Singapore.

References

Alistarh, D.; Grubic, D.; Li, J.; Tomioka, R.; and Vojnovic, M. 2017. QSGD: Communication-efficient SGD via gradient quantization and encoding. In *NeurIPS*.

Breiman, L. 2001. Random forests. *Machine learning*, 45: 5–32.

Chen, Y.; Chen, Z.; Guo, S.; Zhao, Y.; Liu, Z.; Wu, P.; Yang, C.; Li, Z.; and Yu, H. 2023. Efficient Training of Largescale Industrial Fault Diagnostic Models through Federated Opportunistic Block Dropout. In *IAAI*, 15485–15493.

Cheng, X.; Li, C.; and Liu, X. 2022. A Review of Federated Learning in Energy Systems. *I&CPS Asia*, 2089–2095.

Goebel, R.; Yu, H.; Faltings, B.; Fan, L.; and Xiong, Z., eds. 2023. *Trustworthy Federated Learning*, volume 13448. Springer, Cham.

Hyndman, R. J.; and Koehler, A. B. 2006. Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4): 679–688.

Khan, L. U.; Pandey, S. R.; Tran, N. H.; Saad, W.; Han, Z.; Nguyen, M. N. H.; and Hong, C. S. 2020. Federated Learning for Edge Networks: Resource Optimization and Incentive Mechanism. *IEEE Communications Magazine*, 58(10): 88–93.

Le Quéré, C.; Jackson, R. B.; Jones, M. W.; Smith, A. J.; Abernethy, S.; Andrew, R. M.; De-Gol, A. J.; Willis, D. R.; Shan, Y.; Canadell, J. G.; et al. 2020. Temporary reduction in daily global CO2 emissions during the COVID-19 forced confinement. *Nature climate change*, 10(7): 647–653.

Liu, J.; Wang, S.; Wei, N.; Chen, X.; Xie, H.; and Wang, J. 2021. Natural gas consumption forecasting: A discussion on forecasting history and future challenges. *Journal of Natural Gas Science and Engineering*, 90: 103930.

Liu, Y.; Huang, A.; Luo, Y.; Huang, H.; Liu, Y.; Chen, Y.; Feng, L.; Chen, T.; Yu, H.; and Yang, Q. 2020. FedVision: An Online Visual Object Detection Platform Powered by Federated Learning. In *IAAI*, 13172–13179.

Liu, Z.; Chen, Y.; Yu, H.; Liu, Y.; and Cui, L. 2022a. Gtg-shapley: Efficient and accurate participant contribution evaluation in federated learning. *TIST*, 13(4): 1–21.

Liu, Z.; Chen, Y.; Zhao, Y.; Yu, H.; Liu, Y.; Bao, R.; Jiang, J.; Nie, Z.; Xu, Q.; and Yang, Q. 2022b. Contribution-Aware Federated Learning for Smart Healthcare. In *IAAI*, 12396–12404.

Lyu, L.; Yu, H.; Ma, X.; Chen, C.; Sun, L.; Zhao, J.; Yang, Q.; and Philip, S. Y. 2022. Privacy and robustness in federated learning: Attacks and defenses. *IEEE transactions on neural networks and learning systems*.

McMahan, B.; et al. 2017. Communication-efficient learning of deep networks from decentralized data. In *AISTATS*, 1273–1282.

Paszke, A.; Gross, S.; Chintala, S.; Chanan, G.; Yang, E.; DeVito, Z.; Lin, Z.; Desmaison, A.; Antiga, L.; and Lerer, A. 2017. Automatic differentiation in pytorch. In *NIPS 2017 Autodiff Workshop*.

Rice, J. A. 2006. *Mathematical statistics and data analysis*. Cengage Learning.

Sharma, V.; Cali, Ü.; Sardana, B.; Kuzlu, M.; Banga, D.; and Pipattanasomporn, M. 2021. Data-driven short-term natural gas demand forecasting with machine learning techniques. *Journal of Petroleum Science and Engineering*, 206: 108979.

Shi, Y.; Yu, H.; and Leung, C. 2023. Towards fairness-aware federated learning. *TNNLS*.

Soldo, B. 2012. Forecasting natural gas consumption. *Applied energy*, 92: 26–37.

Song, T.; Tong, Y.; and Wei, S. 2019. Profit allocation for federated learning. In *2019 IEEE International Conference on Big Data (Big Data)*, 2577–2586. IEEE.

Tamba, J. G.; Essiane, S. N.; Sapnken, E. F.; Koffi, F. D.; Nsouandélé, J. L.; Soldo, B.; and Njomo, D. 2018. Forecasting natural gas: A literature survey. *IJEEP*, 8(3): 216.

Tu, X.; Zhu, K.; Luong, N. C.; Niyato, D.; Zhang, Y.; and Li, J. 2022. Incentive mechanisms for federated learning: From economic and game theoretic perspective. *IEEE Transactions on Cognitive Communications and Networking*.

Warnat-Herresthal, S.; Schultze, H.; Shastry, K. L.; Manamohan, S.; Mukherjee, S.; Garg, V.; Sarveswara, R.; Händler, K.; Pickkers, P.; Aziz, N. A.; et al. 2021. Swarm learning for decentralized and confidential clinical machine learning. *Nature*, 594(7862): 265–270.

Yang, C.; Liu, J.; Sun, H.; Li, T.; and Li, Z. 2022a. WT-Shapley: Efficient and Effective Incentive Mechanism in Federated Learning for Intelligent Safety Inspection. In *FL-AAAI*.

Yang, C.; Liu, J.; Sun, H.; Li, T.; and Li, Z. 2022b. WTDP-Shapley: Efficient and Effective Incentive Mechanism in Federated Learning for Intelligent Safety Inspection. *TBD*.

Yang, Q.; Liu, Y.; Chen, T.; and Tong, Y. 2019. Federated Machine Learning: Concept and Applications. *ACM Transactions on Intelligent Systems and Technology*, 10(2): 12:1–12:19.

Yang, Q.; Liu, Y.; Cheng, Y.; Kang, Y.; Chen, T.; and Yu, H., eds. 2020. *Federated Learning*. Springer, Cham.

Yu, H.; Miao, C.; Chen, Y.; Fauvel, S.; Li, X.; and Lesser, V. R. 2017. Algorithmic management for improving collective productivity in crowdsourcing. *Scientific Reports*, 7(12541).

Yu, H.; Miao, C.; Leung, C.; Chen, Y.; Fauvel, S.; Lesser, V. R.; and Yang, Q. 2016. Mitigating herding in hierarchical crowdsourcing networks. *Scientific Reports*, 6(4).

Yu, H.; Shen, Z.; Miao, C.; An, B.; and Leung, C. 2014. Filtering trust opinions through reinforcement learning. *Decision Support Systems*, 66: 102–113.

Zeng, R.; Zeng, C.; Wang, X.; Li, B.; and Chu, X. 2021. A comprehensive survey of incentive mechanism for federated learning. *arXiv preprint arXiv:2106.15406*.

Zhan, Y.; Zhang, J.; Hong, Z.; Wu, L.; Li, P.; and Guo, S. 2021. A survey of incentive mechanism design for federated learning. *TETC*, 10(2): 1035–1044.

Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; and Zhang, W. 2021. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *AAAI*, 11106–11115.