Federated Deep Contrastive Learning for Mid-Term Natural Gas Demand Forecasting

Dalin Qin[1a], Guobing Liu[1b], Zengxiang Li[2b], Weicheng Guan[b], Shubao Zhao[b], Yi Wang[2a]

[a] *Department of Electrical and Electronic Engineering, The University of Hong Kong, 999077, Hong Kong SAR, China*
[b] *Digital Technology Research Institute, ENN Group, Beijing, 100102, China*

Abstract

Accurate mid-term gas demand forecasting plays a crucial role for gas companies and policymakers to achieve reliable gas supply plans, supply contracts management, and efficient operation to meet the increasing gas demand. However, mid-term gas demand forecasting faces the problems of data paucity caused by the low frequency of collecting monthly data and heterogeneous consumption patterns of various usage categories. This paper proposes a novel **Fed**erated **Con**trastive pretraining - **L**ocal **C**lustered **F**inetuning paradigm (FedCon-LCF) by integrating federated learning, deep contrastive learning, and clustering approaches. The proposed method can utilize data from multiple gas companies to overcome data paucity issues in a privacy-preserving way, and high-performance forecasting can be achieved by local clustered regression considering the heterogeneous patterns. An improved hierarchical contrastive loss and multi-scale regression loss are integrated to develop the **F**orecasting-**O**riented **C**ontrastive **L**earning model (FOCL), which can effectively extract information and generate fine-grained representations of time series for accurate forecasting. The proposed method is evaluated on a dataset collected from 11 gas companies in 11 different Chinese cities with a total of 17648 clients over 10 usage categories. The proposed method outperforms the benchmark LSTM model with an average improvement of 25.30% in MSE and 16.52% in MAE for 3-month-ahead, 6-month-ahead, 9-month-ahead, and 12-month-ahead gas demand forecasting.

Keywords: Gas demand forecasting, federated learning, contrastive learning, privacy-preserving, heterogeneous consumption patterns

---

1. Introduction

Natural gas is crucial in supplying global energy demands for power generation, residential and commercial heating, industrial manufacturing, and other

[1] These authors contribute equally.
[2] Corresponding authors: lizengxiang@enn.cn, yiwang@eee.hku.hk

uses. The global natural gas demand grew 5.3% and made up 24% of all primary energy in 2021 [1]. An increasing demand shows in emerging economies as they grow and industrialize and underpinned by the continued coal-to-gas switching practice [2]. Accurate gas demand forecasting is thus becoming increasingly crucial for energy companies and policymakers to achieve reliable natural gas supply plans, supply contracts management, efficient operation, and so on.

Natural gas demand forecasting has been studied for a long time. [3] and [4] provided comprehensive reviews of natural gas demand forecasting from 1949 to 2015, covering forecasting methods, different forecasting horizons, data resources, and applied fields. The history of natural gas demand forecasting has been summed up in a more recent review [5], which divides it into the initial stage, conventional stage, AI stage, and all-around stage. The forecasting of gas demand can be classified into three categories, depending on the forecasting horizons: short-term forecasting (for hours to days), mid-term forecasting (for months to a year), and long-term forecasting (several years ahead). The field of short-term forecasting is the one that is growing the fastest, with research in this area increasing from 25% to 75% during the forecasting history. With the improvement of computing capabilities and the availability of larger daily gas consumption datasets, many effective methods have been proposed for short-term gas demand forecasting. [6] forecasted the residential natural gas demand up to 60 hours into the future and compared the performance of linear regression, kernel machine, and artificial neural network (RNN) methods. They found the recurrent neural network with the most accurate results. [7] proposed a hybrid method that integrates wavelet transform technique, multilayer bi-direction long short-term memory neural network (LSTM), and genetic algorithm for hourly gas demand forecasting. [8] proposed an improved singular spectrum analysis approach and combined it with the LSTM model for the daily gas demand forecasts of London, Hong Kong, Melbourne, and Karditsa. [9] proposed to utilize empirical mode decomposition, correlation fusion, and ensemble learning techniques for day-ahead city natural gas demand forecasting. They selected the base-learners LSTM, support vector regression, Bayes regression, and gradient boosting, and used lightGBM as the meta-learner. [10] investigated the combination of functional autoregressive and convolutional neural network models for high-resolution day-ahead gas demand forecasting in German.

Compared to short-term gas demand forecasting, the research on mid-term gas demand forecasting is far more limited, which only accounts for 8% of all forecasting-related works [5]. Mid-term gas demand forecasting is more challenging than short-term forecasting because of increased uncertainty, seasonal variations, and long-term trends. The existing works on mid-term gas demand forecasting mainly focus on machine-learning-based methods. [11] compared the effectiveness of extreme learning machine methods, artificial neural networks, and genetic programming for residential monthly gas demand forecasting in Iran. [12] proposed to combine support vector machine models with discrete wavelet transform methods and achieved better performance than artificial neural networks and genetic programming. [13] showed that the support vector regression model with polynomial cubic kernel function can accomplish effective monthly gas demand forecasting. [14] proposed an approach that integrates wavelet transform, sparse autoencoder, and LSTM to predict the difference between monthly gas supply and demand in the U.S.

Despite the above works, effective methods that utilize advanced deep learning-based techniques for mid-term gas demand forecasting are still lacking. Recently, a novel self-supervised paradigm known as contrastive learning is proposed, and it offers the advantages of (1) greater exploitation of data and (2) the capability to incorporate deep learning models to efficiently extract information from time series. Different from the conventional end-to-end forecasting methods that directly model the relationship between input covariates and the forecasting targets, some works propose to first utilize a contrastive learning method for learning time series representations, followed by a simple finetuning step to achieve accurate forecasting. For example, [15] developed a hierarchical contrasting approach to enable the deep learning-based encoder to learn different semantic levels of time series representations and showed state-of-the-art forecasting performance. Therefore, we are inspired by this work to take advantage of contrastive learning and design an effective approach for mid-term gas demand forecasting.

Different from short-term gas demand forecasting where a large number of daily gas consumption data can be collected, mid-term gas demand forecasting faces severe data paucity problems subject to low collecting frequency. This results in the historical monthly gas consumption dataset being very limited and insufficient to construct a powerful deep learning-based model. One possible way to address this issue is to collect monthly gas data from multiple gas companies to collaboratively train the high-performance deep learning model. However, such gas consumption data may reveal consumers' operation and business confidential information, thus gas companies may be unwilling to release their owned data in practice. The federated learning paradigm proposed in [16] provides solutions to this problem, which enables several clients to jointly train a global model without sharing the local raw data. It has already shown promising applications in the energy sectors for electricity forecasting. [17] investigated the combination of federated learning with differential privacy and secure aggregation techniques to achieve privacy-preserving residential electricity load short-term forecasting. [18] proposed a dynamical clustering federated learning approach to accomplish few-shot energy forecasting for buildings with heterogeneous data. [19] proposed to integrate federated learning with deep reinforcement learning for ultra short-term wind power forecasting. Motivated by these works, we would like to leverage the strength of federated learning to address the severe data paucity problem in mid-term gas demand forecasting.

Another significant challenge to be addressed in mid-term gas demand forecasting is the varied patterns of gas usage. [8] evaluated their proposed forecasting method for four cities, and it was found that the forecasting accuracy is correlated with climatic zone-specific patterns of gas consumption. [9] also pointed out that it is a challenge for a single model to achieve accurate forecasting on ten cities with diverse consumption patterns.

In order to achieve effective mid-term gas demand forecasting with the challenges of data paucity and heterogeneous consumption patterns, we propose a novel method by integrating federated learning, contrastive learning, and clustering approach. To sum up, the contributions of this paper are threefold:

1. **New paradigm:** Propose a novel **Fed**erated **Con**trastive pretraining - **L**ocal **C**lustered **F**inetuning paradigm (FedCon-LCF) for mid-term gas demand forecasting. The proposed approach can effectively address the severe data scarcity problem in mid-term forecasting tasks by utilizing multiple data resources in a privacy-preserving manner for representation extraction. It achieves high-accuracy forecasting by local clustered regression that takes the heterogeneous consumption patterns into consideration.

2. **New method:** Develop a **F**orecasting-**O**riented **C**ontrastive **L**earning model (FOCL) by integrating an improved hierarchical contrastive loss with multi-scale regression loss for model training. Different from traditional end-to-end forecasting models, the proposed model can generate fine-grained representations of any granularity for the temporal data, which enables better regression for mid-term gas demand forecasting.

3. **New findings:** Conduct extensive experiments on monthly gas demand data that has been collected from real-world industrial companies. The proposed FedCon-LCF method is found to make better use of data resources of multiple gas companies than the conventional end-to-end forecasting methods and exhibits more dominant advantages for longer-term forecasting. The proposed FedCon-LCF method can achieve comparable performance to centralized methods and show robustness for different forecasting ranges under the situation of heterogeneous consumption patterns. The experimental results show that the proposed FedCon-LCF method outperforms the benchmark local LSTM with an average improvement of 25.30% in MSE and 16.52% in MAE on nine companies including 413 clients over 5 different gas usage categories.

The rest of the paper is structured as follows. Section 2 defines the problems to be solved for mid-term gas demand forecasting. Section 3 elaborates on the methodologies, including the overall framework and implementation details. Section 4 reports experimental results and analysis. Section 5 draws the conclusions.

2. Problem Statement

The problem can be shown as Fig. 1. Consider there are several companies in different cities, which provide gas supply to the clients they serve and collect data from these clients. These gas companies need to conduct accurate mid-term gas demand forecasting in order to design better supply and delivery plans. For example, one gas company owns clients' gas consumption time series

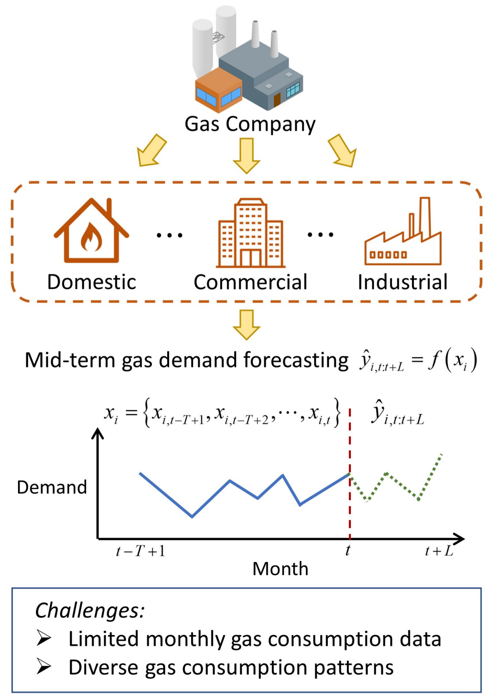


Figure 1: Problems that the gas company faces in mid-term gas demand forecasting.

data, where each time series $x_i = \{x_{i,t-T+1}, x_{i,t-T+2}, \cdots, x_{i,t}\}$ contains T historical monthly gas consumption records. The company aims to construct the forecasting model f to achieve multi-step forecasting $\hat{y}_{i,t:t+L}$ of gas demand in the next L months $\hat{y}_{i,t:t+L} = f(x_i)$.

However, there are mainly two problems in achieving the goals. One is that each company owns very limited data, considering that one client only generates 12 monthly data points in a year and data is collected from recent years subjected to the implementation of the measurement devices. The advanced deep learning-based forecasting models face high over-fitting risks and even fail to be trained by these scarce data. In addition, because of the competitive relationship among companies as well as the data privacy protection regulations, the raw data from different sites can not be directly gathered. The other problem is that clients of different gas usage types, including domestic usage, industrial usage, commercial usage, and so on, may have diverse gas consumption patterns and call for forecasting methods considering such heterogeneity. In conclusion, it is necessary to design effective forecasting methods that consider data shortage and complex consumption pattern problems.

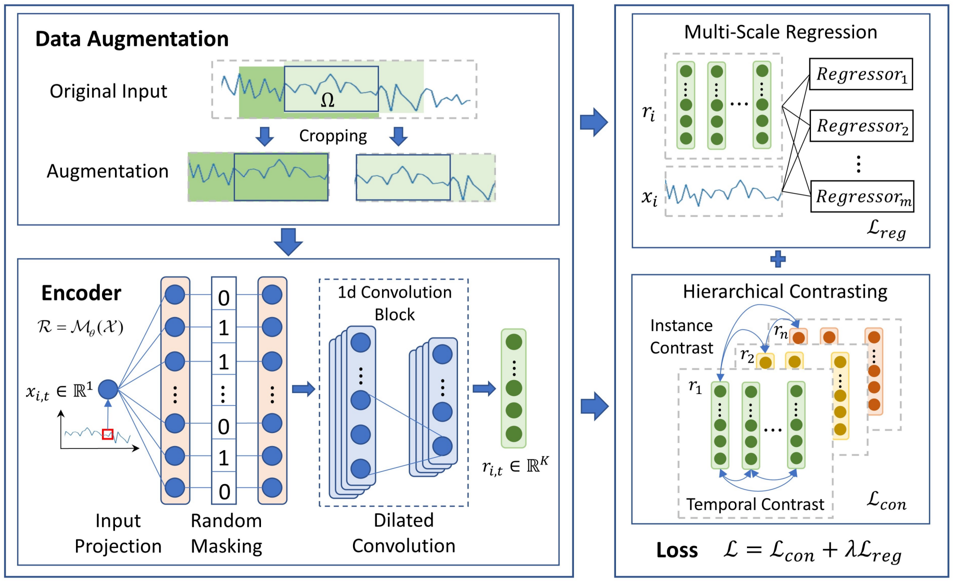


Figure 2: The structure of the proposed FOCL model, including data augmentation method, encoder structure, and loss function.

3. Proposed Method

This section will first introduce the forecasting-oriented contrastive learning algorithm. Then, the local clustered regression method will be described. Finally, the overall algorithm for the proposed paradigm will be demonstrated.

3.1. Forecasting-Oriented Contrastive Learning

A crucial issue for accurate forecasting is how to extract information from time series and create useful representations. [15] proposed a contrastive learning-based universal framework to learn time series representations, which shows great effectiveness for the downstream tasks. Inspired by this, we propose a FOCL model and demonstrate its structure as Fig. 2. The goal of the FOCL model is to learn the mapping function $\mathcal{R} = \mathcal{M}_\theta(\mathcal{X})$, where $\mathcal{M}_\theta$ denotes the function parameterized by θ. The input time series dataset $\mathcal{X} = \{x_1, x_2, \ldots, x_N\}$ contains total N time series, and each time series $x_i \in \mathbb{R}^{T\times 1}$ has T timestamps $x_{i,t}$, where each timestamps $x_{i,t}$ is 1 dimensional. The output $\mathcal{R} = \{r_1, r_2, \ldots, r_N\}$ is the extracted high-dimensional representation of the corresponding time series, where each timestamps $r_{i,t}$ has K dimensions. By the FOCL model, the original input time series can be mapped to the informative representations to improve the performance of downstream forecasting tasks.

The details of the proposed FOCL, including the data augmentation method, encoder architecture, and loss functions are introduced as follows:

3.1.1. Data Augmentation

The proposed FOCL model is a contrastive learning paradigm, where data augmentation is a crucial step that enables the model to learn from positive and negative pairs of augmented data. Specifically, we utilize random cropping as the augmentation method. Two augmented samples are created from the input time series x_i by randomly sampling two overlapping sub-series, with the timestamps in the overlapped area denoted as set Ω. When two augmented versions of the original time series are input, a well-trained FOCL model should keep representation consistent on the overlapped set Ω.

3.1.2. Encoder Structure

The mapping from the original texts to the representations is accomplished by the encoder, we adopt the structure in [15] which is consist of three components, namely input projection, random masking, and dilated convolution. Firstly, the input is mapped to high-dimensional latent space by the projection layer. Then, the latent vector will be masked by a random binary mask on each timestamp. Such timestamp masking procedure will improve the robustness of the model by keeping the representation consistency between the masked and unmasked samples. Finally, the dilated convolution module, which contains six one-dimensional convolution blocks, extracts information from the latent vectors and maps to the output representations.

3.1.3. Loss Functions

Different from the original TS2Vec approach [15], we design a forecasting-oriented contrastive loss function, which is the combination of improved hierarchical contrastive loss $\mathcal{L}_{con}$ and multi-scale regression loss $\mathcal{L}_{reg}$ balanced by a hyperparameter λ:

$$\mathcal{L} = (1-\lambda)\mathcal{L}_{con} + \lambda\mathcal{L}_{reg} \tag{1}$$

Improved Hierarchical Contrastive Loss: The hierarchical contrastive loss $\mathcal{L}_{con}$ is constructed from D hierarchies to guide the encoder generating comprehensive representations of time series, which can be formulated as:

$$\mathcal{L}_{con} = \frac{1}{D}\sum_{d}\mathcal{L}_{con}^{d} \tag{2}$$

where the total number of hierarchies is $D = \arg\max_d \left(2^d \le T\right)$; $\mathcal{L}_{con}^{d}$ denotes the contrasting loss in d-th hierarchy, which is consist of temporal contrasting loss ℓ_{temp}, and instance contrasting loss ℓ_{inst}. Denote r and r' as the representations from two augmentations of the original input. These two representations are contrasted with each other at hierarchical levels, where $r^d = \text{MaxPool1d}(r, \text{kernel size} = 2^d)$ denotes representations in d-th hierarchy. The $\mathcal{L}_{con}^{d}$ can be defined as:

$$\mathcal{L}_{con}^{d} = \frac{1}{NT}\sum_{i}\sum_{t}\left(\ell_{temp}^{(i,t)}\left(r^d, {r'}^d\right) + \ell_{inst}^{(i,t)}\left(r^d, {r'}^d\right)\right) \tag{3}$$

where $\ell^{(i,t)}$ denotes the loss computed for the t-th timestamp of the i-th input time series.

When computing the contrasting loss, the previous work ignores the semantic similarity of samples. However, the timestamps of gas demand time series may have correlations with each other because of temporal closeness, periodicity, and similar industry characteristics. Neglecting such similarity in constructing negative sample pairs may mislead representation learning [20]. Therefore, we propose to filter out similar representations from negative pairs by threshold σ. Specifically, we define a subset of the overlapped set Ω that $\Omega' := \{t'|t' \neq t, \alpha\left(r_{i,t}, r_{i,t'}\right) < \sigma\}$ when calculating $\ell_{temp}^{(i,t)}$, where $\alpha\left(\cdot,\cdot\right)$ denotes the cosine similarity. And set $\Psi := \{j|j \neq i, \alpha\left(r_{i,t}, r_{j,t}\right) < \sigma, 0 < j < B\}$, where B is the batch size. We consider the representations for the same timestamp as positive pairs and representations of different timestamps with similarity less than the threshold as negative pairs. Then the loss for temporal contrasting and instance contrasting can be defined as follows:

$$\ell_{temp}^{(i,t)} = -\log \frac{\exp\left(r_{i,t} \cdot r'_{i,t}\right)}{\exp\left(r_{i,t} \cdot r'_{i,t}\right) + \sum\limits_{t' \in \Omega'} \exp\left(r_{i,t} \cdot r_{i,t'}\right)} \tag{4}$$

$$\ell_{inst}^{(i,t)} = -\log \frac{\exp\left(r_{i,t} \cdot r'_{i,t}\right)}{\exp\left(r_{i,t} \cdot r'_{i,t}\right) + \sum\limits_{j \in \Psi} \exp\left(r_{i,t} \cdot r_{j,t}\right)} \tag{5}$$

The overall contrastive loss incorporates the above loss in a hierarchical manner, allowing the model to learn from both temporal dynamic properties and user-specific characteristics and generate fine-grained representations in arbitrary resolution.

Multi-Scale Regression Loss: As the contrastive loss enables the model to learn from the time series but regardless of the downstream tasks, we propose to integrate a multi-scale regression loss for the generation of forecasting-oriented representations. Consider a total of M regressors $f\left(\cdot\right)$ that map the representations to different forecasting ranges. The L_2-norm loss will be computed for each regressor, and the average loss is defined as the multi-scale regression loss:

$$\mathcal{L}_{reg} = \frac{1}{M} \sum_{m} \ell\left(\mathcal{X}|f_m\left(\mathcal{R}\right)\right) \tag{6}$$

In the task of mid-term gas demand forecasting, we calculate the multi-scale regression loss that involves 3-month-ahead, 6-month-ahead, 9-month-ahead, and 12-month-ahead forecasting. In this way, the model can generate representations that account for forecasting and improve the performance of the downstream tasks.

3.2. Local Clustered Regression

Based on the learned representations, the downstream forecasting tasks can be accomplished by linear regression models. However, the gas company can not construct models for every single client because each client has fewer than a hundred data points, which are insufficient to fit monthly forecasting models. In addition, the gas consumption patterns of different usage types are diverse. As a result, it is also inappropriate to construct one general regression model for all clients regardless of their heterogeneous behaviors. A compromise approach to solving this problem is to cluster the clients and treat each cluster separately. An intuitive way is to conduct clustering according to the raw time series profiles [21]. Recently, some studies show a more effective way to conduct clustering according to the extracted features or representations from the time series [22][23][24]. Inspired by these works, we suggest the company first construct client clusters according to the client-level representation similarity, then fit a regression model for each client cluster.

The company generates client-level representation for each serving client as $\mathrm{MaxPool1d}(r_i,\ \text{kenerl size} = |r_i|)$, and constructs clients cluster set $\mathbb{C}$ as follows:

$$\underset{\mathbb{C}}{\arg\min} \sum_{k=1}^{K} \sum_{r \in C^k} \left\| r - \bar{r}^k \right\|^2 \tag{7}$$

where C^k is the k-th cluster with client representation centroid $\bar{r}^k = \frac{1}{|C^k|} \sum_{r \in C^k} r$.

Then the clustered regression can be achieved by the ordinary least squares approach with k-th cluster representation set $\mathcal{R}(k)$ and forecasting target $\mathbf{Y}(k)$:

$$\widehat{\beta}_k = \left[\mathcal{R}_k^\top \mathcal{R}_k\right]^{-1} \mathcal{R}_k^\top \mathbf{Y}_k, \forall k \tag{8}$$

3.3. Full FedCon-LCF Algorithm

In this subsection, we introduce the full algorithm of the proposed FedCon-LCF paradigm that provides insights for mid-term gas demand forecasting, which can be seen in Fig. 3.

The majority of present federated forecasting methods directly model the relationship between covariates and the forecasting target, which may suffer performance degradation caused by the distribution heterogeneity. In contrast to these works, we propose a novel paradigm that first performs federated contrastive learning to jointly construct an encoder to extract informative representations from the time series, followed by local finetuning of the downstream forecasting through clustered regression that takes into account the heterogeneous consumption patterns. Our approach can provide four main advantages: (1) The federated scheme can effectively address the issue of data scarcity in mid-term gas demand forecasting and ensure privacy preservation by removing the requirement for each company to upload its raw data. (2) The collaborative construction of an encoder, as opposed to a forecasting model, is more resistant to the problem of heterogeneous data distribution, because the encoder focuses more on capturing the characteristics of the time series itself which will be less affected by the heterogeneous conditional probability distributions of covariates and targets from various companies. (3) The self-supervised contrastive learning approach has the potential to maximize the value of data by better utilizing

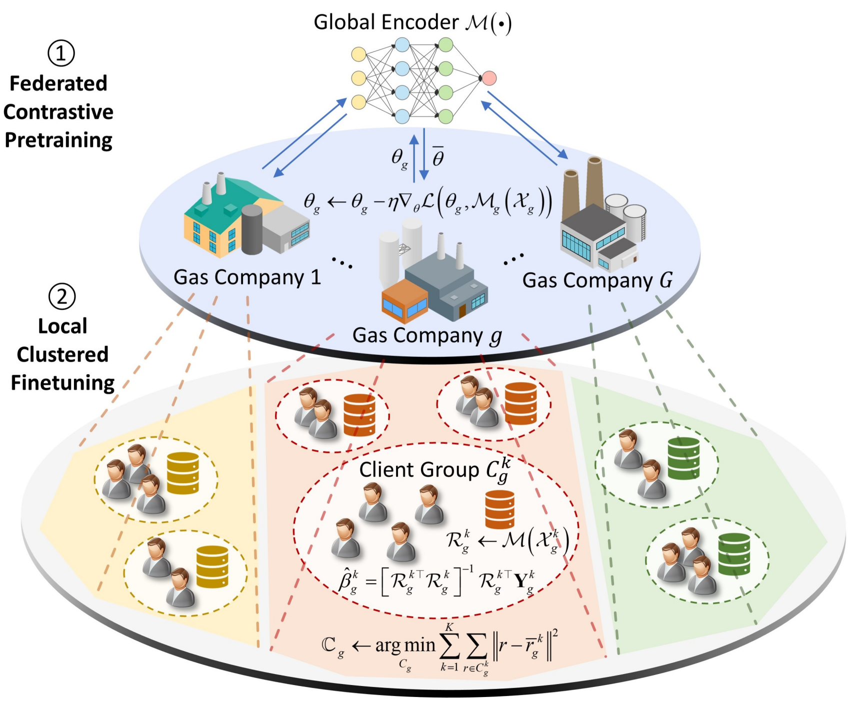


Figure 3: The proposed federated contrastive pretraining - local clustered finetuning paradigm.

it. The generated fine-grained representations lay the foundation for the improvement of downstream forecasting tasks. (4) The local clustered regression approach can handle the heterogeneous consumption patterns by achieving a good balance between building one model for all clients and one model for each individual client.

The implementation of the proposed FedCon-LCF model can be summarized as **Algorithm** 1. Assume there exist G gas companies, denoted as set $\mathbb{G}$, that serve clients from different industries and collect data from them. Define the federated global communication round as E_{global}, local model updating epoch as E_{local}, and learning rate η. Firstly, the federated contrastive pretraining is conducted among several gas companies. In each communication round, the gas company computes the hierarchical contrastive loss and multi-scale regression loss to update the encoder with local data and returns the updated local encoder to the trustful server. The server aggregates the model from all the gas companies to update the global encoder and redistributes it to companies. The federated training process will be stopped until the model converges. The global encoder is then capable of generating fine-grained representations with raw time series input. Based on the trained encoder, the gas company can transform the served clients' data into both client-level and timestamp-level representations. According to the similarity of the client-level representations,

Algorithm 1: FedCon-LCF

function *FedCon-LCF*(η, E_{global}, E_{local}, $\mathbb{G}$, $\mathcal{M}$)**:**
  Initialize: $\mathcal{M} \leftarrow \theta_0$;
  Federated Contrastive Pretraining:
  for *each global round* $e \in E_{global}$ **do**
    for *each gas company* $g \in \mathbb{G}$ **do**
      for *each local update epoch* $l \in E_{local}$ **do**
        Generate representation: $\mathcal{R}_g \leftarrow \mathcal{M}_g\left(\mathcal{X}_g\right)$;
        Compute integrated loss: $\mathcal{L} = \mathcal{L}_{con} + \lambda \mathcal{L}_{reg}$;
        Update local encoder $\mathcal{M}_g$: $\theta_g \leftarrow \theta_g - \eta \nabla_\theta \mathcal{L}\left(\theta_g, \mathcal{R}_g\right)$;
      Return θ_g to Server ;
    Model Aggregate: $\bar{\theta} = \frac{|\mathcal{X}_g|}{|\mathcal{X}|}\theta_g$;
    Parameters Distribute: $\mathbb{G} \leftarrow \bar{\theta}$;
  Local Clustered Finetuning:
  for *each gas company* $g \in \mathbb{G}$ **do**
    Generate client-level representation:
    $\mathrm{MaxPool1d}\left(\mathcal{M}\left(\mathcal{X}_g\right), \mathrm{kernel\_size} = |r|\right)$;
    Cluster based on representations:
    $$\mathbb{C}_g \leftarrow \underset{\mathbb{C}_g}{\arg\min} \sum_{k=1}^{K} \sum_{r \in C_g^k} \left\| r - \bar{r}_g^k \right\|^2;$$
    for *each client group* $C_g^k \in \mathbb{C}_g$ **do**
      $\widehat{\beta}_g^k = \left[{\mathcal{R}_g^k}^\top \mathcal{R}_g^k\right]^{-1} {\mathcal{R}_g^k}^\top \mathbf{Y}_g^k$

the clients can be divided into different clusters. The company can then utilize the timestamp-level representations to fit regression models of different ranges to achieve mid-term gas demand forecasting.

The majority of the computation time required to implement **Algorithm** 1 is taken up by the federated contrastive pretraining process, which is affected by the number of federated communication rounds E_{global}, the number of local training epochs E_{local}, and the size of the local training data from various companies. The federated training time is proportional to E_{global}, and the time of each round is proportional to the local model updating time. The company with the most training data that causes the longest local training time will determine the computation time for each round because the model aggregation in each round can only be carried out when all local models have finished updating. To this end, the time complexity of the proposed algorithm can be presented as $O(E_{global} \cdot (E_{local} \cdot N_{max}))$, where N_{max} denotes the largest number of companies' local data.

Table 1: The Statistics of the Gas Consumers

Categories	Usage Types	Total Clients Number
Domestic	Domestic usage Residential cooking stove Heating boilers Residential water heaters Residential heating furnace	10051
Industrial	Industrial usage Direct-fired absorption chiller Industrial and commercial steam Industrial manufacturing	5771
Commercial	Commercial usage	1826

4. Case Studies

In this section, we conduct mid-term gas demand forecasting experiments using a dataset collected from real-world gas companies to evaluate the effectiveness of the proposed method. Detailed experiment setups, experimental results, and relative analysis and discussion will be provided.

4.1. Experimental Setups

4.1.1. Data Description and Exploratory Analysis

The dataset used in this work is collected from 11 gas companies in 11 different Chinese cities and contains data from a total of 17648 clients of 3 categories with 10 usage types. 17235 of these clients have monthly gas data with a five-year range, and 413 of them have data with a 9-year range. Table. 1 displays the statistics of the data with three main categories and a total of ten usage types.

The 413 clients with 9-year range data are served by 9 different gas companies, and we choose the last two years of the whole nine years as the test set. The rest seven years' data of these 413 clients and all of 17235 clients' data is used for model training and validation. The representation encoder for the TS2Vec, FOCL, and FedCon-LCF approach is trained using 17235 clients' five years data, while the downstream regression model is fitted using 413 clients' first seven years' data.

4.1.2. Model Settings

We provide the details of the benchmark model and the proposed model as follows.

As the long short-term memory neural network (LSTM) shows great ability in capturing time-series characteristics [25], we choose it as the benchmark stands for the end-to-end forecasting model to compare with the proposed method. The local benchmark LSTM model is built with three LSTM layers for feature extraction, and one dense layer as the output layer, where the number of hidden units is set to 16. The selected number of hidden units and layers in the network is relatively small in order to prevent over-fitting to local's limited data. We construct the federated version of LSTM (FedLSTM) and train it with the FedAvg [16] protocol. We also include a recently developed clustering-based federated model called IFCA for comparison [26]. Besides, we compare our model with the original TS2Vec that employs the paradigm of first representation learning, followed by downstream forecasting, and we reconstruct the original TS2Vec model with the same settings as [15].

For the proposed FOCL model, we construct it with a dense layer for input projection, a random masking layer with 50% mask probability on each unit of timestamp, and six sequentially connected convolutional blocks with two 1-D convolutional layers and one residual block contained in each block. We examine in the experiment that the 10 hidden units for each layer in the model, and output representations with dimension 32 are effective for the downstream forecasting task. We set the threshold σ of 0.9 for the negative pairs filtering. The coefficient λ that balances the improved hierarchical contrastive loss and the multi-scale regression loss is set to the default 0.5. When conducting the full algorithm of FedCon-LCF, the local representation clustering is achieved by the K-means++ algorithm [27], and the number of clusters is decided based on the performance on the validation set. According to the experimental results, constructing two to three client clusters for each gas company is enough to account for the heterogeneous consumption patterns problem.

All the models are trained with the Adam optimizer with a learning rate of 0.001. In order to prevent the over-fitting problem during the training process, the dropout with 0.1 probability is adopted.

4.1.3. Evaluation Metrics

Mean absolute error (MAE) and mean squared error (MSE) are selected as metrics to evaluate the performance of the model forecasting:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \widehat{y}_i| \tag{9}$$

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n} (y_i - \widehat{y}_i)^2 \tag{10}$$

The model will be evaluated with these metrics for each gas company, as well as different gas demand forecasting ranges.

4.2. Results Analysis and Discussion

In this section, we compare the models for three different settings: local, federated, and centralized. The data are used in different ways for training these three types of models. With the local and federated settings, where privacy is maintained, the gas company will not disclose the local raw data. Each gas company only uses the serving clients' data for training the local model.

Table 2: Comparison of Model Performance for 3-Month-Ahead Forecasting

Company No.	Local Model						Federated Model						Centralized Model			
	LSTM		TS2Vec		FOCL		Fed LSTM		IFCA		FedCon-LCF		LSTM		FOCL	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
1	0.0288	0.1189	0.0262	0.113	0.0258	0.1134	0.028	0.1235	0.0263	0.1189	**0.0249**	**0.11**	0.0274	0.1173	0.0245	0.1093
2	0.0428	0.16	0.0349	0.1434	0.0339	0.1416	0.0426	0.1484	0.0423	0.1472	**0.0332**	**0.139**	0.0418	0.1485	0.0322	0.1373
3	0.0334	0.1344	0.0314	0.1285	**0.0305**	0.1258	0.0326	0.1326	0.0311	0.1288	0.0307	**0.1255**	0.0315	0.1256	0.0299	0.1243
4	0.0319	0.1265	0.0341	0.1359	0.0375	0.14	0.0309	**0.1232**	**0.0299**	0.1289	0.0315	0.1289	0.0303	0.1274	0.033	0.1324
5	0.0343	0.1335	0.028	0.1196	0.027	0.117	0.0325	0.1216	0.0331	0.1218	**0.0257**	**0.1155**	0.0323	0.1244	0.0253	0.1146
6	0.0397	0.1347	0.0305	0.1257	0.0304	0.1244	0.0394	0.1384	0.038	0.1354	**0.0301**	**0.1236**	0.0388	0.1355	0.0305	0.1251
7	0.0401	0.1423	0.0398	0.1435	0.0395	0.1403	0.0395	0.1395	0.039	0.139	**0.0376**	**0.1365**	0.0375	0.1399	0.0387	0.1415
8	0.0332	0.1352	0.0297	0.1212	0.0298	0.1194	0.0313	**0.116**	0.0308	0.1232	**0.0286**	0.1187	0.0284	0.1214	0.027	0.1179
9	0.04	0.1461	0.0325	0.1333	0.0315	**0.1298**	0.0376	0.1366	0.0363	0.1351	**0.0315**	0.1307	0.0363	0.1381	0.0308	0.1287
Average	*0.0343*	*0.134*	*0.0307*	*0.1263*	*0.0302*	*0.1247*	*0.0333*	*0.13*	*0.0341*	*0.1309*	*0.0294*	*0.1227*	*0.0322*	*0.1275*	*0.0289*	*0.1223*

Table 3: Comparison of Model Performance for 6-Month-Ahead Forecasting

Company No.	Local Model						Federated Model						Centralized Model			
	LSTM		TS2Vec		FOCL		Fed LSTM		IFCA		FedCon-LCF		LSTM		FOCL	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
1	0.0353	0.1418	0.028	0.1198	0.0278	0.1206	0.0359	0.1444	0.0328	0.3175	**0.0272**	**0.1174**	0.0347	0.135	0.0268	0.1172
2	0.0596	0.1975	0.0399	0.1573	**0.0368**	0.1515	0.0613	0.1973	0.058	0.1803	0.0375	**0.1508**	0.0548	0.1768	0.0353	0.1482
3	0.038	0.1475	0.0314	0.1316	0.031	0.1306	0.0386	0.1518	0.0372	0.1411	**0.0306**	**0.1289**	0.0379	0.1445	0.0297	0.127
4	0.0359	0.1434	0.0332	0.1375	0.0361	0.1406	0.0384	0.1484	0.0371	0.1402	**0.0326**	**0.1336**	0.0389	0.1433	0.0321	0.1338
5	0.0436	0.1631	0.029	0.124	0.0278	0.1207	0.0393	0.1535	0.0377	0.1488	**0.0263**	**0.1193**	0.04	0.1467	0.0259	0.1176
6	0.0655	0.2028	0.0345	0.1375	0.0346	0.1372	0.0504	0.1652	0.0468	0.1528	**0.0341**	**0.1367**	0.0526	0.1744	0.0343	0.1369
7	0.0641	0.2048	0.0384	0.1452	0.0391	0.1453	0.0509	0.168	0.0452	0.159	**0.0379**	**0.1423**	0.0502	0.1669	0.0383	0.1447
8	0.0388	0.1556	0.0287	0.1215	0.0291	**0.1206**	0.0362	0.1402	0.032	0.1329	**0.0282**	0.1211	0.0356	0.1354	0.0271	0.1197
9	0.0775	0.2302	0.038	0.1457	0.0369	**0.1418**	0.0601	0.1917	0.0426	0.1662	**0.0366**	0.1419	0.0529	0.1704	0.036	0.1409
Average	*0.0459*	*0.1654*	*0.0321*	*0.1321*	*0.0317*	*0.131*	*0.0429*	*0.1582*	*0.0410*	*0.1710*	*0.031*	*0.129*	*0.0413*	*0.1492*	*0.0303*	*0.1281*

Table 4: Comparison of Model Performance for 9-Month-Ahead Forecasting

Company No.	Local Model						Federated Model						Centralized Model			
	LSTM		TS2Vec		FOCL		Fed LSTM		IFCA		FedCon-LCF		LSTM		FOCL	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
1	0.0417	0.1575	0.0305	0.1269	0.0303	0.1273	0.0416	0.1572	0.0375	0.148	**0.0298**	**0.1245**	0.0403	0.1535	0.0295	0.1251
2	0.0554	0.1897	0.0437	0.1671	0.0413	0.1632	0.0695	0.2109	0.0603	0.1913	**0.0409**	**0.1607**	0.0519	0.1753	0.0399	0.1605
3	0.0369	0.1446	0.0333	0.1381	0.0322	0.1354	0.0374	0.1486	0.0351	0.1457	**0.0319**	**0.1338**	0.0367	0.1453	0.0312	0.1325
4	0.0388	0.1525	0.0345	0.1422	0.0361	0.1426	0.039	0.154	**0.031**	0.1452	0.0343	**0.1386**	0.0386	0.1501	0.0327	0.1372
5	0.0469	0.174	0.0321	0.1324	0.0294	0.1259	0.0401	0.1576	0.0367	0.1442	**0.0284**	**0.1251**	0.04	0.1523	0.028	0.1245
6	0.0486	0.1675	0.0389	0.1468	0.0394	0.1473	0.0461	0.1549	**0.0385**	**0.1457**	0.0388	0.1468	0.0446	0.1558	0.0386	0.1458
7	0.0486	0.1691	0.0407	0.1506	0.0411	0.1516	0.0478	0.1604	0.0417	0.1525	**0.0397**	**0.148**	0.0473	0.1613	0.0394	0.1486
8	0.0371	0.1455	0.0297	0.1245	0.0306	**0.1234**	0.0357	0.1418	0.0316	0.1266	**0.0299**	0.1251	0.0375	0.1464	0.0288	0.1226
9	0.06	0.1944	0.0423	0.155	0.0418	0.1528	0.053	0.1792	0.0424	0.1533	**0.0412**	**0.1516**	0.0543	0.1757	0.0403	0.1503
Average	*0.0439*	*0.1615*	*0.0347*	*0.1392*	*0.034*	*0.1374*	*0.0436*	*0.16*	*0.0394*	*0.1503*	*0.0334*	*0.1356*	*0.0418*	*0.1549*	*0.0327*	*0.1349*

Table 5: Comparison of Model Performance for 12-Month-Ahead Forecasting

Company No.	Local Model						Federated Model						Centralized Model			
	LSTM		TS2Vec		FOCL		Fed LSTM		IFCA		FedCon-LCF		LSTM		FOCL	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
1	0.0458	0.1634	0.0325	0.1329	0.0326	0.1332	0.0441	0.172	0.0402	0.1462	**0.032**	**0.1304**	0.0435	0.1563	0.0322	0.132
2	0.0633	0.2041	0.0458	0.1718	0.0435	0.1676	0.063	0.2045	0.0498	0.1721	**0.0434**	**0.1661**	0.0622	0.1928	0.0437	0.1689
3	0.0439	0.1615	0.0355	0.145	0.0342	0.1413	0.0439	0.1567	0.0399	0.1498	**0.0338**	**0.1397**	0.0434	0.1597	0.0335	0.1392
4	0.0425	0.154	**0.037**	0.1497	0.0382	0.1488	0.0449	0.1698	0.0421	0.1522	0.0374	**0.1466**	0.0449	0.1605	0.0359	0.1458
5	0.0544	0.1858	0.0351	0.1406	0.0325	0.1332	0.0496	0.1815	0.0329	0.1342	**0.0309**	**0.1309**	0.0517	0.1732	0.0304	0.1312
6	0.066	0.2055	0.0399	0.1496	0.0401	**0.1495**	0.0616	0.1971	0.0492	0.1577	**0.0399**	0.1499	0.0611	0.1926	0.0392	0.1481
7	0.0541	0.1816	0.0429	0.156	0.043	0.1555	0.0537	0.1732	0.047	0.1599	**0.0426**	**0.1538**	0.0533	0.1785	0.0414	0.1536
8	0.036	0.1454	0.0317	0.1289	0.0316	**0.1269**	0.038	0.1381	0.036	0.151	**0.0315**	0.1289	0.0373	0.1483	0.0306	0.1272
9	0.0819	0.2337	0.0464	0.1631	0.0461	0.1616	0.0603	0.1936	0.0496	0.1766	**0.0453**	**0.1597**	0.0614	0.1906	0.0446	0.1592
Average	*0.0514*	*0.176*	*0.037*	*0.1455*	*0.0363*	*0.1432*	*0.0485*	*0.1718*	*0.0430*	*0.1555*	*0.0357*	*0.1415*	*0.0483*	*0.1674*	*0.0354*	*0.1417*

Table 6: Model Improvement for Different Forecasting Ranges

Forecasting Range	FOCL v.s. LSTM		FOCL v.s. TS2Vec		FedCon-LCF v.s. LSTM		FedCon-LCF v.s. FedLSTM		FedCon-LCF v.s. IFCA	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
3	11.95%	6.94%	1.63%	1.27%	14.29%	8.43%	11.71%	5.62%	13.75%	6.28%
6	30.94%	20.80%	1.25%	0.83%	32.46%	22.01%	27.74%	18.46%	24.47%	24.55%
9	22.55%	14.92%	2.02%	1.29%	23.92%	16.04%	23.39%	15.25%	15.28%	9.77%
12	29.38%	18.64%	1.89%	1.58%	30.54%	19.60%	26.39%	17.64%	16.91%	9.02%
Average	*23.70%*	*15.32%*	*1.70%*	*1.24%*	*25.30%*	*16.52%*	*22.31%*	*14.24%*	*17.60%*	*12.40%*

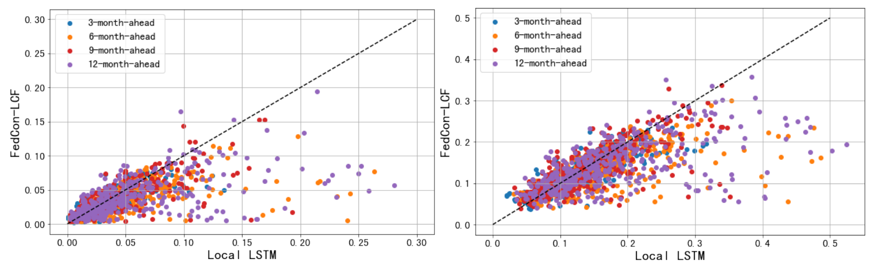
Every gas company updates the federated model with local data while exchanging model parameters with a trustful third-party server. While the centralized setting disregards privacy and presumes that data from all gas companies can be freely collected. Every gas company can use the aggregated data to train the centralized model.

We include local models of LSTM, TS2Vec, and FOCL, federated models of FedLSTM, IFCA, and FedCon-LCF, and centralized LSTM and FOCL models. The models forecasting performance on each gas company in terms of MSE and MAE for 3-month-ahead, 6-month-ahead, 9-month-ahead, and 12-month-ahead forecasting is shown in Table 2-5 respectively. Because the centralized model sets no restriction on data utilization, we treat it as the reference, and we mark the best performance among local and federated models in bold. The weighted average performance (weighted by the number of clients served by each gas company) is shown in the last row of the table in italics. And the average performance improvement when comparing our proposed model with other local and federated models for different forecasting ranges is shown in Table 6.

When we compare the local models, we can find that the LSTM model performs the worst. It is because even though one gas company collects data from tens of clients, the number of monthly data is still too limited to train a good LSTM model that directly captures the relationships between covariates and the forecasting target from raw time series. In contrast, the TS2Vec and FOCL approach that first learns meaningful representations of time series and then fits downstream linear regression model shows more promising results. For the average performance of four forecasting ranges, the proposed FOCL can achieve improvements over benchmark LSTM by 23.70% and 15.32% in terms of MSE and MAE respectively. And the performance of the FOCL method surpasses that of the original TS2Vec, demonstrating the effectiveness of the improved hierarchical contrastive loss and multi-scale regression loss designs. The average performance of FOCL can achieve improvements over TS2Vec of 1.70% and 1.24% in terms of MSE and MAE respectively.

When we extend the models to the federated setting, we can find that the FOCL and LSTM performance can be improved by the federated training thanks to the collaboration of multiple gas companies. However, the diverse data from different industries may affect the performance of the FedLSTM model, and such a problem also exists when we construct a centralized LSTM model. In contrast, the proposed FedCon-LCF approach federatedly pretrain the encoder that focuses on learning how to extract representations that may be less influenced by the heterogeneous conditional probability distributions of covariates and targets problem. And the challenge of heterogeneous consumption patterns is tackled by the downstream local clustered finetuning approach. For the clustering-based federated model IFCA, it achieves better performance than the FedLSTM model because of its advantages in dealing with heterogeneous data distribution. However, the IFCA model adopts the end-to-end modeling approach without capturing the universal representations for different ranges of forecasting, it performs worse than our proposed FedCon-LCF model. As a result, the average performance over four forecasting ranges of FedCon-LCF can achieve large improvements over local LSTM of 25.30% and 16.52% in terms of MSE and MAE. The FedCon-LCF outperforms FedLSTM with average improvements of 22.31% and 14.24% in terms of MSE and MAE. Additionally, FedCon-LCF surpasses IFCA with average improvements of 17.6% and 12.4% in terms of MSE and MAE. We can also observe that the forecasting task becomes more difficult and leads to a higher loss as the forecasting range becomes longer. However, the proposed FedCon-LCF method can make better use of data resources of multiple gas companies and exhibit more dominant advantages for longer-term forecasting.

We investigate how FedCon-LCF and local LSTM model perform on certain clients for different forecasting ranges, and the comparison in terms of MSE and MAE is presented in Fig. 4. Different colors denote distinct forecasting ranges, and the x-axis and y-axis respectively represent the performance of the local LSTM and the FedCon-LCF model. The dots below the black diagonal dashed line represent clients for which the FedCon-LCF model outperforms the local LSTM. We can observe that the majority of clients can improve their forecasting performance by adopting the proposed FedCon-LCF model, and some clients can see significant improvements in 6-month-ahead, 9-month-ahead, and 12-month-ahead forecasting. For the purpose of demonstration, we randomly choose one client from each of the three categories (domestic, industrial, and commercial) to show the FedCon-LCF's forecasting results of the last year in the test dataset, as shown in Fig. 5. We can observe that the FedCon-LCF method is capable of producing accurate forecasting on diverse consumption patterns in different forecasting ranges, which reflects the efficacy of the proposed strategy of local



(a) Model comparison in terms of MSE (b) Model comparison in terms of MAE

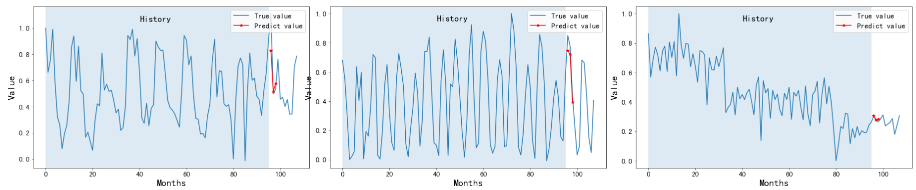
Figure 4: The performance of the FedCon-LCF in comparison to the local benchmark on 413 clients. Different colors represent 3-month-ahead, 6-month-ahead, 9-month-ahead, and 12-month-ahead forecasting respectively. The x-axis shows the performance of the local LSTM, and the y-axis represents the performance of the FedCon-LCF model. The dots indicate that FedCon-LCF performs better than local LSTM in the area below the diagonal dashed line. (a) The model performance comparison in terms of MSE. (b) The model performance comparison in terms of MAE.

clustered regression and multi-scale regression loss.

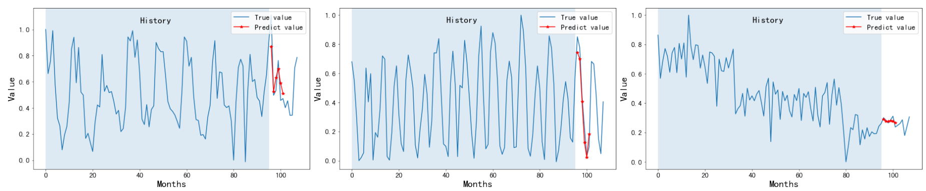
When we investigate how the FedCon-LCF model performs for companies from different cities, we can find that the model achieves the best performance for Company 1, where clients in that city are mainly of the domestic and industrial type. We further examine the FedCon-LCF model's performance on clients from domestic, industrial, and commercial categories, as shown in Fig. 6. Because domestic clients make up the majority of the training samples, referred to the Table 1, the FedCon-LCF performs better for the domestic category than it does for the other two categories. For the industrial and commercial categories, the FedCon-LCF model performs better on industrial clients than the commercial in the settings of 3-month-ahead, and 6-month-ahead forecasting, while the situation is reversed in the settings of 9-month-ahead, and 12-month-ahead forecasting. We reckon that more industrial clients' samples than commercial clients are used in model training making the FedCon-LCF model to better predict industrial gas demand in a relatively short period of time. However, industries may have varied gas consumption patterns in different seasons, making it difficult to estimate over a longer period of time.

As we compare the proposed FedCon-LCF model to the referencing centralized model, we observe that FedCon-LCF can outperform the centralized LSTM model and achieves comparable performance as the centralized FOCL model. It should be noted that the centralized FOCL model is constructed without the clustering strategy and each gas company builds a single regression model for all clients with different gas usage types. But the FedCon-LCF model adopts the clustered regression strategy to account for the heterogeneous consumption patterns and thus achieves even better performance than the centralized FOCL model on some specific gas companies.

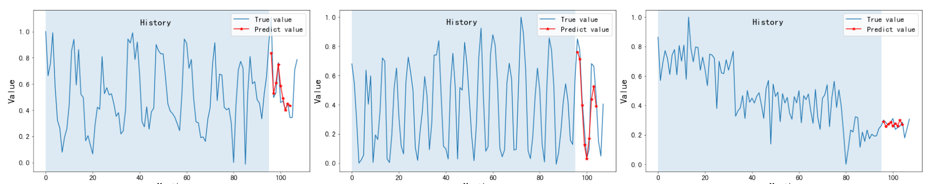
In order to validate the effectiveness of the proposed strategy, we conduct the ablation study and show the results in Fig. 7. We evaluate the model



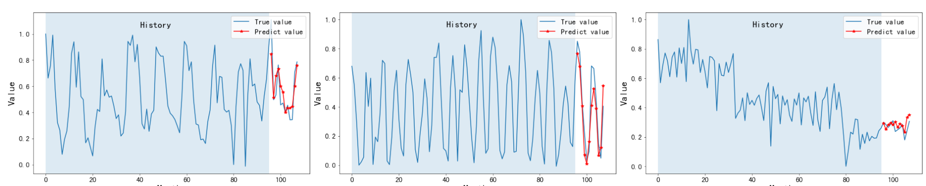
(a) Domestic gas demand 3-month-ahead forecasting (b) Industrial gas demand 3-month-ahead forecasting (c) Commercial gas demand 3-month-ahead forecasting



(d) Domestic gas demand 6-month-ahead forecasting (e) Industrial gas demand 6-month-ahead forecasting (f) Commercial gas demand 6-month-ahead forecasting



(g) Domestic gas demand 9-month-ahead forecasting (h) Industrial gas demand 9-month-ahead forecasting (i) Commercial gas demand 9-month-ahead forecasting



(j) Domestic gas demand 12-month-ahead forecasting (k) Industrial gas demand 12-month-ahead forecasting (l) Commercial gas demand 12-month-ahead forecasting

Figure 5: Mid-term gas demand forecasting of different usage categories. The subfigures in the left column, middle column, and right column are gas demand forecasts of domestic usage, industrial usage, and commercial usage respectively. 3-month-ahead, 6-month-ahead, 9-month-ahead, and 12-month-ahead gas demand forecasting are shown by the four rows of subfigures in order from top to bottom. The true value is indicated by the blue line, while the predicted value is shown by the red dots.

performance in three settings, the first is to conduct federated contrastive pre-training but without local clustered regression, where only one regression model is constructed for each company. The second setting is to train the FedCon-LCF model without the hierarchical contrastive loss $\mathcal{L}_{con}$ (setting $\lambda = 1$), where the model is trained with only multi-scale regression loss. The third setting is to

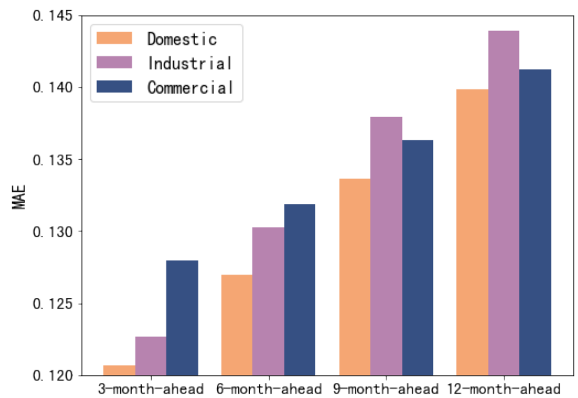


Figure 6: The performance of the FedCon-LCF model on clients of three categories: domestic, industrial, and commercial.

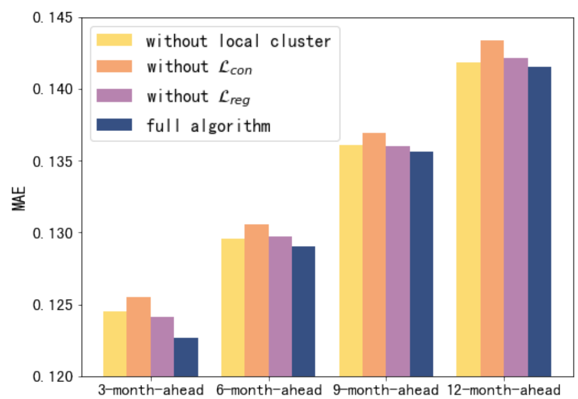


Figure 7: The performance of the FedCon-LCF model with three ablation settings: (1) without local clustered regression, (2) without hierarchical contrastive loss $\mathcal{L}_{con}$, (3) without multi-scale regression loss $\mathcal{L}_{reg}$.

train the FedCon-LCF model without the multi-scale regression loss $\mathcal{L}_{reg}$ (setting $\lambda = 0$). We can observe that all the strategies are beneficial for improving forecasting, and hierarchical contrastive loss serves the most significant function since it is the primary component that directs the model to produce fine-grained representations from the time series.

We also evaluate the time consumption for training the federated model, which is shown in Fig. 8. As FedLSTM and IFCA model is trained with the end-to-end paradigm, models have to be trained for 3-month-ahead, 6-month-ahead, 9-month-ahead, and 12-month-ahead forecasting separately. Because IFCA involves clustering in the federated training procedure, it takes the longest time in model training. In contrast, the proposed FedCon-LCF model only needs to be trained once for all possible forecasting ranges and attempts to capture uni-

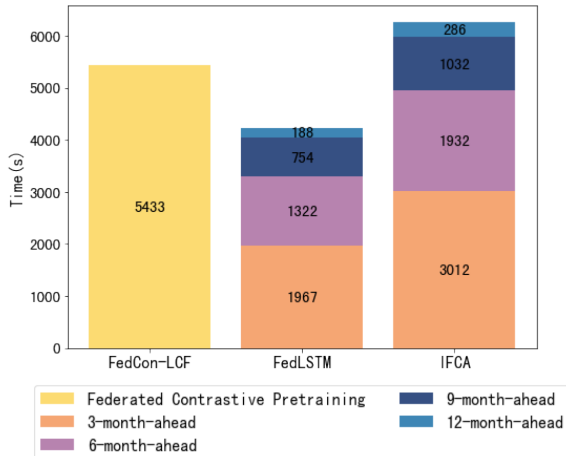


Figure 8: The training time of the FedCon-LCF, FedLSTM, and IFCA model.

versal representations. The training time of FedCon-LCF mainly focuses on the federated contrastive pretraining procedure, where the local clustered regression can be accomplished in a short time that can be neglected. Although the FedCon-LCF model requires more time to train than the FedLSTM model does, it eliminates repeating model training and can be completed in a manageable amount of time, which confirms its practicability.

5. Conclusions

This paper developed a novel FedCon-LCF paradigm to account for the severe data scarcity and diverse consumption pattern challenges in order to accomplish effective mid-term gas demand forecasting. The FOCL approach is proposed to extract information and generate fine-grained representations from the raw gas demand time series by integrating an improved hierarchical contrastive loss with multi-scale regression loss. We conduct comprehensive experiments using data from real-world gas companies, and the evaluation of models' performance for 3-month ahead, 6-month ahead, 9-month ahead, and 12-month ahead forecasting demonstrates the superiority of the proposed method. The FOCL is more capable than the conventional end-to-end forecasting LSTM model for mid-term gas demand forecasting tasks, and it also outperforms the original TS2Vec model in terms of the capacity to produce informative representations. Additionally, the FedCon-LCF approach enables the collaboration of multiple gas companies to enhance forecasts without violating data privacy and to construct several regression models for various client clusters. The effectiveness of the proposed FedCon-LCF approach is further demonstrated by the comparable performance to centralized methods and robustness for mid-term gas demand forecasting with different ranges under the situation of heterogeneous consumption patterns.

6. Acknowledgement

The work was supported in part by the National Key R&D Program of China (2022YFE0141200), in part by the National Natural Science Foundation of China (72242105).

References

[1] bp Statistical Review of World Energy 2022 edition.
URL https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2022-full-report.pdf

[2] bp Energy Outlook 2023 edition.
URL https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/energy-outlook/bp-energy-outlook-2023.pdf

[3] B. Soldo, Forecasting natural gas consumption, Applied energy 92 (2012) 26–37.

[4] J. G. Tamba, S. N. Essiane, E. F. Sapnken, F. D. Koffi, J. L. Nsouandélé, B. Soldo, D. Njomo, Forecasting natural gas: A literature survey, International Journal of Energy Economics and Policy 8 (3) (2018) 216.

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

[6] P. Potočnik, J. Šilc, G. Papa, et al., A comparison of models for forecasting the residential natural gas demand of an urban area, Energy 167 (2019) 511–522.

[7] H. Su, E. Zio, J. Zhang, M. Xu, X. Li, Z. Zhang, A hybrid hourly natural gas demand forecasting method based on the integration of wavelet transform and enhanced deep-rnn model, Energy 178 (2019) 585–597.

[8] N. Wei, C. Li, X. Peng, Y. Li, F. Zeng, Daily natural gas consumption forecasting via the application of a novel hybrid model, Applied Energy 250 (2019) 358–368.

[9] F. Li, H. Zheng, X. Li, F. Yang, Day-ahead city natural gas load forecasting based on decomposition-fusion technique and diversified ensemble learning model, Applied Energy 303 (2021) 117623.

[10] Y. Chen, X. Xu, T. Koch, Day-ahead high-resolution forecasting of natural gas demand and supply in germany with a hybrid model, Applied Energy 262 (2020) 114486.

[11] N. Izadyar, H. C. Ong, S. Shamshirband, H. Ghadamian, C. W. Tong, Intelligent forecasting of residential heating demand for the district heating system based on the monthly overall natural gas consumption, Energy and Buildings 104 (2015) 208–214.

[12] N. Izadyar, H. Ghadamian, H. C. Ong, C. W. Tong, S. Shamshirband, et al., Appraisal of the support vector machine to forecast residential heating demand for the district heating system based on the monthly overall natural gas consumption, Energy 93 (2015) 1558–1567.

[13] O. F. Beyca, B. C. Ervural, E. Tatoglu, P. G. Ozuyar, S. Zaim, Using machine learning tools for forecasting natural gas consumption in the province of istanbul, Energy Economics 80 (2019) 937–949.

[14] W. Qiao, W. Liu, E. Liu, A combination model based on wavelet transform for predicting the difference between monthly natural gas production and consumption of us, Energy 235 (2021) 121216.

[15] Z. Yue, Y. Wang, J. Duan, T. Yang, C. Huang, Y. Tong, B. Xu, Ts2vec: Towards universal representation of time series, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 36, 2022, pp. 8980–8987.

[16] B. McMahan, E. Moore, D. Ramage, S. Hampson, B. A. y Arcas, Communication-efficient learning of deep networks from decentralized data, in: Artificial intelligence and statistics, PMLR, 2017, pp. 1273–1282.

[17] J. D. Fernández, S. P. Menci, C. M. Lee, A. Rieger, G. Fridgen, Privacy-preserving federated learning for residential short-term load forecasting, Applied Energy 326 (2022) 119915.

[18] L. Tang, H. Xie, X. Wang, Z. Bie, Privacy-preserving knowledge sharing for few-shot building energy prediction: A federated learning approach, Applied Energy 337 (2023) 120860.

[19] Y. Li, R. Wang, Y. Li, M. Zhang, C. Long, Wind power forecasting considering data privacy protection: A federated deep reinforcement learning approach, Applied Energy 329 (2023) 120291.

[20] X. Liu, Y. Liang, Y. Zheng, B. Hooi, R. Zimmermann, Spatio-temporal graph contrastive learning, arXiv preprint arXiv:2108.11873 (2021).

[21] T. W. Liao, Clustering of time series data—a survey, Pattern recognition 38 (11) (2005) 1857–1874.

[22] Y. Liu, Y. Zhou, K. Yang, X. Wang, Unsupervised deep learning for iot time series, IEEE Internet of Things Journal (2023).

[23] Q. Ma, J. Zheng, S. Li, G. W. Cottrell, Learning representations for time series clustering, Advances in neural information processing systems 32 (2019).

[24] Q. Lei, J. Yi, R. Vaculin, L. Wu, I. S. Dhillon, Similarity preserving representation learning for time series clustering, arXiv preprint arXiv:1702.03584 (2017).

[25] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural computation 9 (8) (1997) 1735–1780.

[26] A. Ghosh, J. Chung, D. Yin, K. Ramchandran, An efficient framework for clustered federated learning, Advances in Neural Information Processing Systems 33 (2020) 19586–19597.

[27] D. Arthur, S. Vassilvitskii, K-means++ the advantages of careful seeding, in: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, 2007, pp. 1027–1035.