Cluster-driven Personalized Federated Learning for Natural Gas Load Forecasting

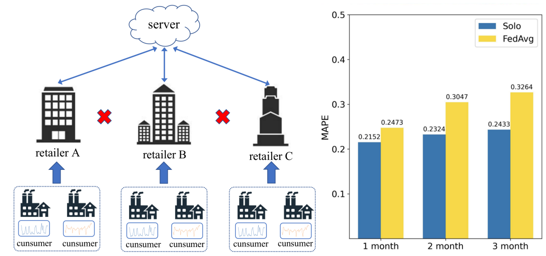
Shubao Zhao , **Jia Liu** , **Guoliang Ma** , **Jie Yang** , **Di Liu** , **Zengxiang Li**
Digital Research Institute, ENN Group, China
zhaoshubao, liujiaam, maguolianga, yangjiev, liudif, lizengxiang@enn.cn

Abstract

Natural gas load forecasting is essential to retailers in terms of profit-making and service quality. In practice, a retailer has limited consumer load data to build an accurate prediction model. Federated learning enables retailers to train a global model collaboratively, without compromising data privacy. However, it could not behave well on all consumers due to their diversity, e.g., different load patterns. To address this data heterogeneity issue, we propose a cluster-driven personalized federated learning (CPFL) framework. Firstly, a knowledge-based federated clustering is proposed to categorize similar consumers from different retailers into clusters in a privacy-preserving manner. Then, vanilla federated learning is adopted to pre-train a global model, leveraging all available data from retailers. Finally, the pre-trained model is fine-tuned and personalized to each cluster respectively, using an attention-based model aggregation strategy according to the contribution difference of individual consumers in the cluster. Comprehensive experiments are conducted using a real-world data set with 2000 consumers from eight retailers, and the results show our proposed CPFL framework outperforms the state-of-the-art personalized federated learning approaches for time-series forecasting.

1 Introduction

Natural gas load forecasting [Liu *et al.*, 2021; Wei *et al.*, 2019] is essential to retailers in terms of profit-making (e.g., demand management, resource coordination, and pipeline network planning) and service quality (e.g., safety management, personalized contract and proactive scheduling). With the rapid development of AI technologies, complex deep learning models trained by large amounts of data extract time-series features and thus enable accurate forecasting. However, the retailer usually serves a certain number of consumers within a region or even a small city. Its data is likely insufficient to support a high-performance deep learning model. In addition, it is probable to deduce consumers' operation and business confidential information through their energy load. Retailers are committed to preserve consumers data privacy without sharing to any others. Consequently, reputable retailers are reluctant to upload data for centralized model training in practice.



(a) Vanilla federated learning. (b) MAPE of Solo and FedAvg.

Figure 1: Vanilla federated learning leverages data from different retailers but could not address data heterogeneity issues among them. Experiments show that the federated model is even worse than the solo models trained using the data of individual consumers respectively.

A new AI paradigm federated learning (FL) is proposed in [McMahan *et al.*, 2017], which enables participants to jointly train a model without sharing their data. It has obtained great achievements in various industry domains, including healthcare, finance service and smart cities [Yang *et al.*, 2019]. Recently, federated learning has also been used in load forecasting [Taïk and Cherkaoui, 2020; Husnoo *et al.*, 2022; Fekri *et al.*, 2022]. However, most FL application scenarios assume data from all participants are independent and identically distributed (IID). In real-world cases, this assumption is hard to be satisfied due to the diversity among participants. We conduct experiments with vanilla federated learning using federated average (FedAvg) model aggregation strategy on 2000 consumers. As shown in Figure 1(b), the federated model trained by all available data performs even worse than individual models trained for each consumer.

In order to alleviate the impact of data heterogeneity and thus improving model performance, several personalized federated learning mechanisms have been proposed for various application domains including CV and NLP [Tan *et al.*, 2022]. Specifically, in the time series forecasting field, [Wang *et al.*, 2022] propose to fine-tune the global federated model to personalized models for individual electricity consumers. Unfortunately, the very limited data of each consumer make the personalized model prone to bias and over-fitting. Moreover, a large number of models can also increase the management cost of the retailers. In addition, [Guo *et al.*, 2022] propose to build personalized federated models using similar data within the cluster only. However, this method could not fully exploit the advantages of federated learning as retailers' data out of the cluster are not used. Furthermore, the proposed similarity-based methods are data-driven without exploring the domain knowledge (e.g., time-series data processing and pattern recognition). Therefore, designing an effective personalized federated learning framework that can adapt to individual consumers properly remains crucial.

According to our domain knowledge, consumers from different retailers may have similar load patterns, for instance, if they are from the same region and/or the same industry sector. Therefore, a cluster of similar consumers who shares the same model could achieve higher prediction accuracy, while avoiding over-fitting and reducing management cost. From the other perspective, the data of consumers from different clusters may also be valuable especially for extracting low-level common features. In order to take advantage of both perspectives, we propose a two-phase cluster-driven personalized federated learning framework. Firstly, a global model is pre-trained using vanilla federated learning by leveraging all available data from retailers. Then, the global model is fine-tuned for each cluster in a federated learning manner using its consumers' data from different retailers. In other words, consumers within the same cluster cooperated with each other to build a robust and accurate personalized model, while keeping their data within its retailer's sovereignty. Furthermore, consumers in the same cluster are not identical, and thus have different contributions to the personalized model. Hence, an attention-based model aggregation strategy is adopted in the fine-tuning federated learning phase, considering different weights of consumers' local model updates. Experimental results on 2000 consumers from eight retailers show that our proposed framework outperforms the aforementioned personalized federated learning approaches [Wang *et al.*, 2022; Guo *et al.*, 2022]. Moreover, the performance of our framework is less sensitive to the number of clusters, compared with other cluster-driven personalized approaches.

The main contributions of this paper are as follows:

- A cluster-driven personalized federated learning for natural gas load forecasting is proposed to leverage all available data while considering their heterogeneity.
- A knowledge-based federated clustering is proposed to categorize similar customers into clusters in a privacy-preserving manner, based on our industry domain knowledge and time-series data analytic experiences.
- A two-phased federated learning mechanism is proposed, enabling retailers to pre-training a global model and then fine-tuning it to personalized models with respect to consumer clusters collaboratively in a federated learning manner.
- An attention-based model aggregation strategy is introduced in fine-tuning phase, to capture the contributions of individual consumers to the personalized federated model of their cluster.

The rest of this paper is organized as follows. Section 2 reviews the related work on personalized federated learning especially for time-series forecasting,. Section 3 introduces CPFL framework and technical details of each module. Section 4 reports experiments on a real-world dataset. Section 5 concludes this paper.

2 Related Work

2.1 Time-series Forecasting

Over the years, time series forecasting is a hot topic in both industry and academy communities. The continuous efforts could be classified into the following three groups with the increasing model complexity: 1) statistical approaches such as autoregressive integrated moving average(ARIMA) [Pradhan *et al.*, 2016], 2) machine learning models such as XGBoost [Li and Zhang, 2018], 3) deep learning models such as LSTM [Peng *et al.*, 2021] and transformer [Xu *et al.*, 2021]. Complex deep learning models trained by large amounts of data could extract time series features automatically, and relieve the burden of time-consuming data pre-processing and feature engineering. However, in practice, a retailer serving parts of consumers within a region usually has insufficient data to train a high-performance deep learning model. Therefore, it is desirable to enable retailers to collaborate with each other to train AI models using their data in a secure way.

2.2 Federated Learning

With the emergence of Federated Learning [McMahan *et al.*, 2017], participants could jointly train a model by exchanging model updates instead of the raw data, reducing the risk of confidential information leakage. As society is more and more concerned with privacy protection, federated learning becomes popular in various applications. The most relevant studies to this paper are [Fekri *et al.*, 2022; Wang *et al.*, 2019] apply federated learning to load forecasting with smart meter IoT data. Despite the popularity of FL, researchers have pointed out that the performance of vanilla FL models may degrade significantly, in the case that data from different participants could not satisfy IID assumption [Sattler *et al.*, 2019; Sattler *et al.*, 2020; Briggs *et al.*, 2020]. To address the issue, different model aggregation strategies (e.g., FedProx [Li *et al.*, 2020]) and Scaffold [Karimireddy *et al.*, 2020]) are proposed to replace classical federate averaging. [Ji *et al.*, 2019] proposed an attention-based model aggregation strategy considering different contributions of participant model updates to the global model. However, the aggregated global model could not obtain optimum performance for all participants due to high diversity. Consequently, it is still challenging to deploy FL models in real-world application scenarios.

2.3 Personalized Federated Learning

As data heterogeneity issue is very common in applications and critical to FL model performance, a number of researches are conducted in these few years [Tan *et al.*, 2022] to address the issue. There are two research directions most relevant to our work. The first is to obtain a personalized FL model for individuals [Wang *et al.*, 2019; Chen *et al.*, 2020]. In the healthcare sector, [Chen *et al.*,

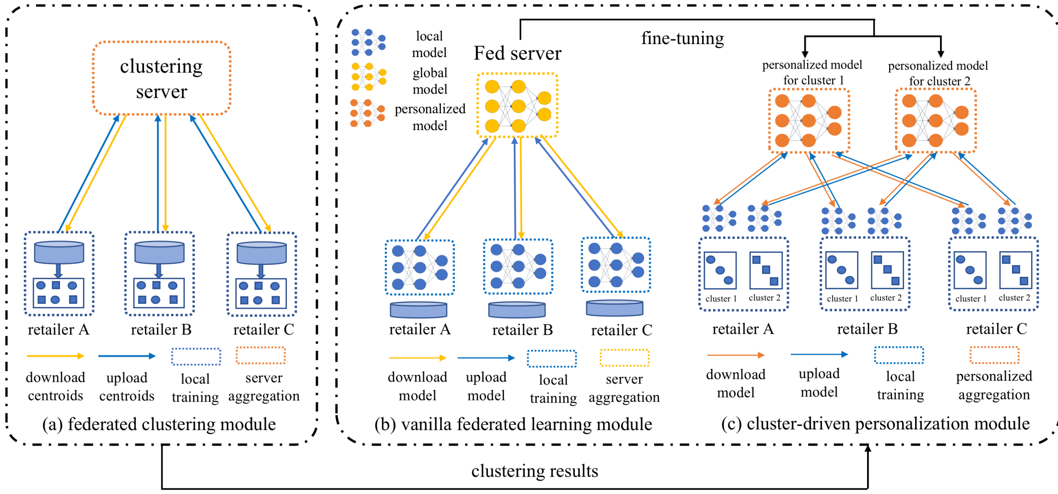


Figure 2: The framework of cluster-driven personalized federated learning.

2020] propose a federated transfer learning framework that enables collaborations among large amounts of wearable devices, which aggregates device local models through federated learning and then obtains a personalized model through transfer learning. In the energy sector, [Wang *et al.*, 2022] propose a personalized federated learning for individual consumer electricity load forecasting. The algorithm first trains a global federated model and then performs fine-tuning with each consumer's data to obtain its personalized model. The second is to obtain a personalized FL model for each cluster of similar individuals. In the healthcare sector, [Huang *et al.*, 2019] propose to cluster the distributed data into clinically meaningful communities and learnt one personalized model for each community. In the manufacturing sector, [Guo *et al.*, 2022] propose a multi-task machinery fault diagnosis algorithm, according to the similarity of equipment IoT data features. In addition, [Sattler *et al.*, 2020] propose a clustered federated learning based on multi-task optimization.

The aforementioned personalized federated learning approaches are either prone to bias and over-fitting due to very limited data of individuals or could not fully utilize the valuable data from all participants. Instead, our proposed framework enables natural gas retailers to participate in two-phase federated learning, which could fully exploit federated learning advantages by leveraging all available consumer data without compromising their privacy and improve performance by fine-tuning the pre-trained model for each cluster according to different contributions of its consumers.

3 Cluster-driven personalized federated learning Framework

Figure 2 illustrates cluster-driven personalized federated learning (CPFL) framework, which consists of the following three key functional modules:

- Knowledge-based federated clustering: categorizing customers into clusters in a privacy-preserving manner.
- Vanilla federated learning: pre-train a global federated model across retailers collaboratively.
- Cluster-driven personalization: fine-tune the pre-trained global federated model for each cluster in a federated learning manner.

Our proposed framework has three important features to distinguish from other personalized federated learning approaches for time-series forecasting in the literature. Firstly, a knowledge-based federated clustering categorizes consumers with similar load patterns into the same cluster, according to our industry domain knowledge and time-series data processing experiences. In addition, the federated clustering is proposed to enable retailers to obtain global information for clustering consumers without compromising their privacy. Secondly, retailers participate in two-phase federated learning, pre-training global model by leveraging as much data as possible, and fine-tuning the model for each cluster to achieve a robust and accurate personalized model. Thirdly, we introduced an attention-based aggregation strategy in fine-tuning phase to consider the contribution difference of consumers to the cluster, for the purpose of improving performance further.

3.1 Knowledge-based Federated Clustering

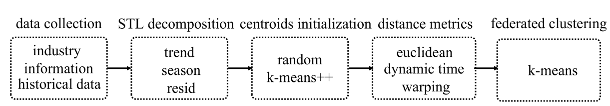


Figure 3: The flowchart of knowledge-based federated clustering.

The goal of knowledge-based federated clustering is to categorize similar consumers into clusters without revealing the private data of retailers. The algorithm, as shown in Figure 3, was developed based on our specific natural gas distribution domain knowledge (e.g., the seasonal pattern of domestic heating and manufacturing factories) and data scientists' experiences in time series data processing.

The algorithm takes historical natural gas consumption data and general information of different consumers into ac-

Algorithm 1 Federated clustering algorithm

Input: The dataset $U = \{\pi_1, \pi_2, ..., \pi_p\}$ from p retailers, number of clusters k, initial cluster centroids C_i, local epochs L, and global epochs G.
Output: Cluster centroids $C = \{C_1, C_2, ..., C_k\}$ and clustering labels.

```
Client initializes cluster centroids C and send to server.
for g = 1, 2, ..., G do
    The server sends the cluster centroids C to the client.
    for l = 1, 2, ..., L do
        Assign cluster label for each samples.
        Client update the cluster centroids using the formula (2).
    end for
    Client upload the cluster centroids to the server.
    Server update the cluster centroids using the formula (3).
end for
```

count. It decomposes the time series data through Seasonal and Trend decomposition using Loess (STL) and then clustering consumers based on the importance of these components. The federated clustering(k-means) module is shown in Figure 2(a), supporting frequently used centroids initialization and distance metrics.

As shown in Algorithm 1, the cluster centroids of participants are initialized using random or k-means++ algorithms. Each participant updates local cluster centroids based on its consumers, using a k-means clustering algorithm. The server receives the cluster centroids from all the clients and aggregates them to obtain global cluster centroids. This process iterates many times until the model converges, i.e., the updates of cluster centroids are marginal.

Suppose that the training datasets from different retailers $U = \{\pi_1, \pi_2, ..., \pi_p\}$, where $\pi_l = \{x_1, x_2, ..., x_n\}$ denotes the data from retailer l, are partitioned into K clusters $C = \{C_1, C_2, ..., C_k\}$. The objective function of federated k-means can be formulated as:

$$J = \sum_{l=1}^{p} \sum_{j=1}^{k} \sum_{i=1}^{n} ||x_i^{(\pi_l)_j} - C_j||^2 \tag{1}$$

Here, $x_i^{(\pi_l)_j}$ denotes the training data in cluster centroid C_j and belonging to retailer π_l. $||x_i^{(\pi_l)_j} - C_j||^2$ denotes the distance from $x_i^{(\pi_l)_j}$ to C_j. Client and server update cluster centroids C_i of retailer π_l using formulates 2 and 3 respectively.

$$C_i^{\pi_l} = \frac{1}{|C_i^{\pi_l}|} \sum_{x \in C_i^{\pi_l}} x \tag{2}$$

$$C_i = \frac{1}{p} \sum_{C_i^{\pi_l} \in U} C_i^{\pi_l} \tag{3}$$

3.2 Two-phase Personalized Federated Learning

The proposed CPFL is as shown in Figure 2, adopting two-phases personalized federated learning as shown in Algorithm 2. A global federated model is first pre-trained across different retailers, using following objective function:

$$\arg\min_{w_g} L(w_g) = \sum_{\pi_l \in U} \sum_{x_i \in \pi_l} loss(f(x_i), y_i) \tag{4}$$

Algorithm 2 Personalized federated learning

Input: Data from different retailers $U = \{\pi_1, \pi_2, ..., \pi_p\}$, results of federated k-means $C = \{C_1, C_2, ..., C_k\}$, global federated epochs G, personalized federated epochs L.
Output: Personalized federated learning model for each cluster.

```
//Federated learning
for g = 1, 2, ..., G do
    for l = 1, 2, ..., p do
        Client π_l update model weights w_{π_l}.
    end for
    Server aggregates the model weights of each client to obtain
    global model weights w_g.
end for
//Cluster-specific personalization
for i = 1, 2, ..., k do
    Cluster C_i initializes personalized model weights w_{c_i} using
    w_g.
    for j = 1, 2, ..., L do
        for l = 1, 2, ..., p do
            Client π_l update model weights w_{π_l}.
        end for
        Server aggregates the model weights of each client to obtain
        personalized model weights w_{c_i}.
    end for
end for
```

,where $loss(f(x_i), y_i)$ denotes the loss function of the neural network. In this paper, mean square error(MSE) is selected as the loss function. $f(x_i)$ denotes the forecast result and y_i denotes the ground truth. w_g denotes the parameters to be learned. p denotes the number of participators.

The pre-trained global federated model is then fine-tuned to a personalized federated model for each cluster. Because consumers in the same cluster are considered similar, it realizes personalization in the form of federated learning instead of focusing only on the data of a specific consumer. The personalized federated model can effectively integrate useful knowledge from different consumers and meet the needs of privacy-preserving. The objective function is formulated as:

$$\arg\min_{w_{c_i}} L(w_{c_i}) = \sum_{\pi_l \in U} \sum_{c_i \in \pi_l} \sum_{x_i \in c_i} loss(f(x_i), y_i) \tag{5}$$

,where w_{c_i} denotes the parameter of c_i to be learned.

3.3 Attention-based Model Aggregation Strategy

Although we believe that consumers in the same cluster have similar load patterns, it is necessary to analyze the contributions of consumers to the personalized federated model in a more fine-grained manner. Because in a fine-grained view, they are still not similar. The forecasting performance of the personalized federated model may be sensitive to the number of clusters. If we can consider the contribution of different consumers to the personalized federated model in model aggregation, the sensitivity may be reduced. [Ji *et al.*, 2019] introduce an attention mechanism into the model aggregation and propose a Federated Attention(FedAtt) algorithm. The algorithm takes into account the distance between the client model and server model, so that the learned features of each client can be effectively selected to generate a better server model. Inspired by this, this section introduces the FedAtt algorithm for personalized model aggregation. We use euclidean distance to measure the distance between the local model and the personalized model. Suppose that cluster $C_i = \{x_1, x_2, ..., x_n\}$ has n consumers. At the t epoch, attention score can be expressed as $\alpha^t = \{\alpha_1^t, \alpha_2^t, ..., \alpha_n^t\}$. for consumer i, we calculate the euclidean distance between the local model and personalized model as follows:

$$d_i^t = ||w_i^{t+1} - w^t||^2 \tag{6}$$

The distance set between local model and personalized federated model can be expressed as $d_t = \{d_1^t, d_2^t, ..., d_n^t\}$. The attention score of i consumer on epoch t is as follows:

$$\alpha_i = d_i^t / \sum_{i=1}^{n} d_i^t \tag{7}$$

The personalized model weights update is as follows:

$$w^{t+1} = w^t - \sum_{i=1}^{n} \alpha_i (w^t - w_i^{t+1}) \tag{8}$$

4 Case Studies

4.1 Experimental Settings

A real-world dataset with natural gas monthly consumption of 2000 consumers from eight retailers is utilized in our experiments. These consumers come from different sector (e.g., Residential, Industrial and Commercial). Since those consumers may start natural gas service at different times, their load data cover the different lengths of time periods. To enable meaningful training, all consumers in the dataset have at least three years of load data. We choose the data from last year as the test set and the data before that as the train set for both knowledge-based federated clustering and personalized federated learning. MSE, MAE, and MAPE metrics are used to evaluate the performance of forecasting models with 1, 2 and 3 months forecasting horizons. In order to verify the performance advantages of our proposed CPFL framework, several model training methods are implemented and evaluated based on the aforementioned dataset, using the same LSTM network with a hidden layer and fully connected layer.

- Solo: train forecasting models for consumers individually.
- Centralized: collected data from all retailers and train a centralized model.
- FedAvg [McMahan *et al.*, 2017]: a vanilla federated learning among retailers using federated averaging model aggregation strategy.
- FedAtt [Ji *et al.*, 2019]: federated learning among retailers using attention-based model aggregation strategy.
- PFL-Solo [Wang *et al.*, 2022]: fine-tuning is used to personalize the global federated model for individual consumers.
- CFL [Guo *et al.*, 2022]: train a personalized federated model for each cluster, without using data from other clusters.
- CPFL-FedAvg: our proposed CPFL framework, without attention-based aggregation strategy.
- CPFL-FedAtt: our proposed CPFL framework, with attention-based aggregation strategy.

4.2 Forecasting Performance

Table 1 compares the performance of the aforementioned algorithms with different forecasting horizons. FedAvg and FedAtt perform even worse than Solo, although FL enables them to leverage all available data. The data heterogeneity makes them difficult to train optimum global federated models which are suitable for all consumers. Centralized model with access to all data outperforms FedAvg and FedAtt. CFL achieves better performance than Solo, Centralized, FedAvg and FedAtt, as it alleviates data heterogeneity to a certain extent through clustering. PFL-Solo achieves better performance than Solo, Centralized, FedAvg, FedAtt and CFL, as the pre-trained FL model fully leverages available data and data heterogeneity issue is alleviated by fine-tuning the pre-trained model for individual consumers. CPFL-FedAvg and CPFL-FedAtt are our proposed algorithms without and with an attention-based model aggregation strategy respectively. Both of them outperform the aforementioned personalized FL algorithms, as our CPFL adopts two-phase FL leveraging all available data and fine-tuning personalized FL model to clusters lowering the risk of over-fitting. CPFL-FedAtt outperforms CPFL-FedAvg, as it takes into account the contributions of different consumers to the personalized model in the fine-tuning and personalization phase.

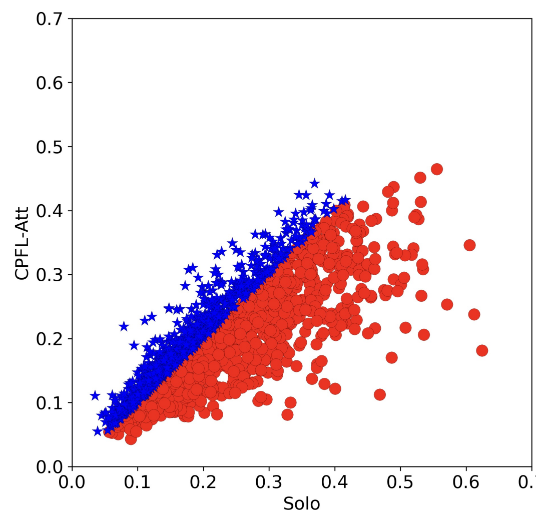


Figure 4: MAPE of CPFL-Att against Solo.

4.3 Forecasting Performance Distribution

Figure 4 illustrates the advantages of CPFL-FedAtt over Solo in terms of MAPE for 2000 consumers. The red circle represents the performance enhancement while the blue star represents the performance degradation after participating in our proposed federated learning framework using CPFL-FedAtt algorithms. It is obvious that more consumers could achieve

Methods	1 month			2 month			3 month		
	MSE	MAE	MAPE	MSE	MAE	MAPE	MSE	MAE	MAPE
Solo	0.0148	0.0919	0.2152	0.0170	0.0990	0.2324	0.0188	0.1037	0.2433
Centralized	0.0138	0.0883	0.1991	0.0181	0.1030	0.2308	0.0209	0.1118	0.2484
FedAvg	0.0155	0.0990	0.2473	0.0212	0.1180	0.3047	0.0238	0.1252	0.3264
FedAtt	0.0154	0.0987	0.2453	0.0208	0.1168	0.2995	0.0237	0.1249	0.3237
CFL	0.0130	0.0863	0.2018	0.0155	0.0955	0.2274	0.0168	0.0999	0.2396
PFL-Solo	0.0129	0.0851	0.2000	0.0158	0.0954	0.2262	0.0174	0.1000	0.2370
CPFL-FedAvg	0.0129	0.0848	0.1982	0.0151	0.0930	0.2212	0.0162	0.0968	0.2321
CPFL-FedAtt	**0.0126**	**0.0834**	**0.1930**	**0.0148**	**0.0911**	**0.2137**	**0.0161**	**0.0956**	**0.2261**

Table 1: Performance Comparison among model training algorithms with different forecasting horizons.

higher performance using CPFL-FedAtt than Solo, and the extent of performance improvement is also more significant. Figure 5 reports the MAPE distribution of the Solo and CPFL-Att. The number of consumers with MAPE less than 0.2 increased significantly after participating CPFL-FedAtt, while consumers with MAPE greater than 0.3 decreased significantly. In other words, CPFL-Att model enables more consumers achieve desirable forecasting accuracy, which is beneficial to both consumers and retailers for proper demand management and resource scheduling.

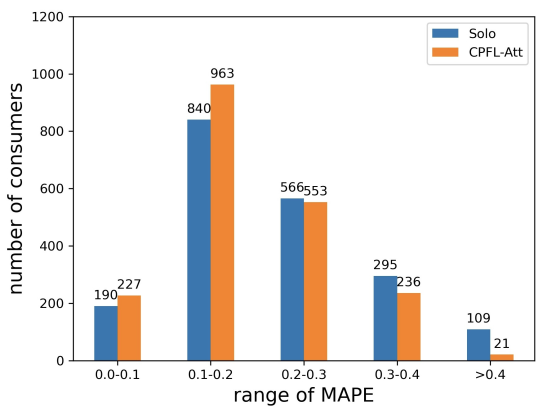


Figure 5: MAPE distribution of our model and Solo.

4.4 Performance on Different Number of Clusters

The number of clusters is one of the most critical parameters in cluster-driven approaches. With the increasing number of consumers in the same cluster, a better model generalization could be expected, but the data heterogeneity among many consumers may affect forecasting performance. Figure 6 reports forecasting performance (in terms of MAPE) of CFL, CPFL-FedAvg and CPFL-FedAtt with different numbers of clusters. CFL uses data within the cluster only, and thus its performance is sensitive to the number of clusters. To the extreme, if the number of clusters is very small, the CFL performs similarly to vanilla federated learning. In contrast, if the number of clusters is very large, CFL performs similarly to Solo models. CPFL-Avg reports better performance than CFL. If the number of clusters is very large, the CPFL-Avg performs similarly to PFL-Solo. That is, fine-tuning the global model for each consumer. This could lead to bias and over-fitting. CPFL-Avg leverage all data from retailers and fine-tune a personalized model for each cluster but ignores the different contributions of consumers to the personalized model. CPFL-Att not only uses the data of all consumers but also takes into account the contribution of different consumers to the personalized model. This makes the model more robust and less sensitive to the number of clusters.

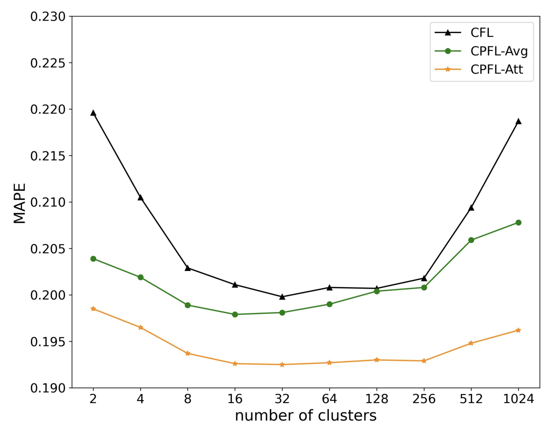


Figure 6: MSE with different number of clusters.

5 Conclusion and Future Work

In this paper, we proposed a cluster-driven personalized federated learning (CPFL) framework for natural gas load forecasting. It enables retailers to work together to categorize their consumers into a number of clusters based on domain knowledge, and then train a personalized model for each cluster by two-phased federated learning, for the purpose of leveraging all available data and addressing data heterogeneity issues properly. According to our experiments based on a real-world dataset, the CPFL framework can achieve better performance than other personalized FL approaches recently proposed in the literature.

In the future, we will further improve the performance of the CPFL framework from multiple perspectives. For example, we will explore the feasibility of much larger deep learning networks (e.g., transformer) for datasets with more consumers and daily (or hourly) energy consumption. We may also consider the relationship among clusters to get the most valuable knowledge selectively (e.g., using graph neural network) for model performance enhancement.

References

[Briggs *et al.*, 2020] Christopher Briggs, Zhong Fan, and Peter Andras. Federated learning with hierarchical clustering of local updates to improve training on non-iid data. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–9. IEEE, 2020.

[Chen *et al.*, 2020] Yiqiang Chen, Xin Qin, Jindong Wang, Chaohui Yu, and Wen Gao. Fedhealth: A federated transfer learning framework for wearable healthcare. *IEEE Intelligent Systems*, 35(4):83–93, 2020.

[Fekri *et al.*, 2022] Mohammad Navid Fekri, Katarina Grolinger, and Syed Mir. Distributed load forecasting using smart meter data: Federated learning with recurrent neural networks. *International Journal of Electrical Power & Energy Systems*, 137:107669, 2022.

[Guo *et al.*, 2022] Sheng Guo, Zengxiang Li, Hui Liu, Shubao Zhao, and Cheng Hao Jin. Personalized federated learning for multi-task fault diagnosis of rotating machinery. 2022.

[Huang *et al.*, 2019] Li Huang, Andrew L Shea, Huining Qian, Aditya Masurkar, Hao Deng, and Dianbo Liu. Patient clustering improves efficiency of federated machine learning to predict mortality and hospital stay time using distributed electronic medical records. *Journal of biomedical informatics*, 99:103291, 2019.

[Husnoo *et al.*, 2022] Muhammad Akbar Husnoo, Adnan Anwar, Nasser Hosseinzadeh, Shama Naz Islam, Abdun Naser Mahmood, and Robin Doss. Fedrep: Towards horizontal federated load forecasting for retail energy providers. *arXiv preprint arXiv:2203.00219*, 2022.

[Ji *et al.*, 2019] Shaoxiong Ji, Shirui Pan, Guodong Long, Xue Li, Jing Jiang, and Zi Huang. Learning private neural language modeling with attentive aggregation. In *2019 International joint conference on neural networks (IJCNN)*, pages 1–8. IEEE, 2019.

[Karimireddy *et al.*, 2020] Sai Praneeth Karimireddy, Satyen Kale, Mehryar Mohri, Sashank Reddi, Sebastian Stich, and Ananda Theertha Suresh. Scaffold: Stochastic controlled averaging for federated learning. In *International Conference on Machine Learning*, pages 5132–5143. PMLR, 2020.

[Li and Zhang, 2018] Pin Li and Jin-Suo Zhang. A new hybrid method for china's energy supply security forecasting based on arima and xgboost. *Energies*, 11(7):1687, 2018.

[Li *et al.*, 2020] Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks. *Proceedings of Machine Learning and Systems*, 2:429–450, 2020.

[Liu *et al.*, 2021] Jinyuan Liu, Shouxi Wang, Nan Wei, Xi Chen, Hanyu Xie, and Jie Wang. Natural gas consumption forecasting: A discussion on forecasting history and future challenges. *Journal of Natural Gas Science and Engineering*, 90:103930, 2021.

[McMahan *et al.*, 2017] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pages 1273–1282. PMLR, 2017.

[Peng *et al.*, 2021] Shanbi Peng, Ruolei Chen, Bin Yu, Min Xiang, Xiugao Lin, and Enbin Liu. Daily natural gas load forecasting based on the combination of long short term memory, local mean decomposition, and wavelet threshold denoising algorithm. *Journal of Natural Gas Science and Engineering*, 95:104175, 2021.

[Pradhan *et al.*, 2016] Prabodh Pradhan, Bhagirathi Nayak, and Sunil Kumar Dhal. Time series data prediction of natural gas consumption using arima model. *International Journal of Information Technology & Management Information System*, 7(3):1–7, 2016.

[Sattler *et al.*, 2019] Felix Sattler, Simon Wiedemann, Klaus-Robert Müller, and Wojciech Samek. Robust and communication-efficient federated learning from non-iid data. *IEEE transactions on neural networks and learning systems*, 31(9):3400–3413, 2019.

[Sattler *et al.*, 2020] Felix Sattler, Klaus-Robert Müller, and Wojciech Samek. Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints. *IEEE transactions on neural networks and learning systems*, 32(8):3710–3722, 2020.

[Taïk and Cherkaoui, 2020] Afaf Taïk and Soumaya Cherkaoui. Electrical load forecasting using edge computing and federated learning. In *ICC 2020-2020 IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE, 2020.

[Tan *et al.*, 2022] Alysa Ziying Tan, Han Yu, Lizhen Cui, and Qiang Yang. Toward personalized federated learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[Wang *et al.*, 2019] Kangkang Wang, Rajiv Mathews, Chloé Kiddon, Hubert Eichner, Françoise Beaufays, and Daniel Ramage. Federated evaluation of on-device personalization. *arXiv preprint arXiv:1910.10252*, 2019.

[Wang *et al.*, 2022] Yi Wang, Ning Gao, and Gabriela Hug. Personalized federated learning for individual consumer load forecasting. *CSEE Journal of Power and Energy Systems*, 2022.

[Wei *et al.*, 2019] Nan Wei, Changjun Li, Xiaolong Peng, Yang Li, and Fanhua Zeng. Daily natural gas consumption forecasting via the application of a novel hybrid model. *Applied Energy*, 250:358–368, 2019.

[Xu *et al.*, 2021] Jiehui Xu, Jianmin Wang, Mingsheng Long, et al. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in Neural Information Processing Systems*, 34, 2021.

[Yang *et al.*, 2019] Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2):1–19, 2019.