Anomaly Detection in Distributed Refrigeration and Air-Conditioning Systems Using Federated Learning

Toshiaki Hirata
Tokyo Information Design
Professional University
hirata@tid.ac.jp

Kaoru Kaneko
Marble Corp.
kaneko.kaoru@marble-corp.co.jp

***Abstract*—The Internet of Things (IoT) has been widely applied to predict and diagnose anomalous conditions in the operation of various machines and systems. In particular, accurate and efficient diagnosis of failures is especially important for refrigeration and air-conditioning systems with equipment distributed across multiple locations. In previous research, we explored strategies for deploying vibration sensors to optimize diagnostic models for a given site. In this study, we apply federated learning (FL) to detect anomalies in distributed air-conditioning and refrigeration systems and evaluated the anomaly detection accuracy compared with local learning at each location to obtain guidelines for creating appropriate models from the perspective of network configurations and hyperparameters in FL. We confirm the effectiveness of FL for a convolutional autoencoder(CAE) which is a type of unsupervised learning in distributed refrigeration and air-conditioning systems. We clarified that, to reliably detect anomalies, we need to adjust the method of dividing the network configuration of federated representation learning(FedRep) which is a type of FL into an aggregation section and a local learning-only section, and which layer to insert padding into.**

***Index Terms*—Anomaly detection, Federated learning, IoT**

I. Introduction

Machinery installed at various facilities is a critical component of infrastructure, and its abnormal operation can have a significant impact on both the environment and business continuity. Therefore, predicting and diagnosing abnormalities are crucial for preventing such problems. We have previously focused on employing the Internet of Things (IoT) technology to implement a system for collecting vibration sensor data and diagnosing existing air-conditioning and refrigeration systems, with a particular emphasis on anomaly detection [1] [2] [3]. In refrigeration and air-conditioning systems, accurate and efficient prediction and diagnosis of failures in equipment distributed across multiple locations is important. Federated learning (FL) has garnered attention as a method for achieving large-scale learning by individually utilizing data locally retained by numerous client computers distributed across a network [4]. Because confidentiality is required for sensor data, it also addresses security concerns.

In this study, we applied FL to refrigeration and air-conditioning systems distributed across multiple locations and evaluated the anomaly detection accuracy compared with local learning at each location to obtain guidelines for creating appropriate models from the perspective of network configurations and hyperparameters in FL. We evaluated federated averaging (FedAvg) [4] and federated representation learning (FedRep) [5] using a convolutional autoencoder (CAE) [6]. FedAvg was the first major FL algorithm, and FedRep is a variation of FedAvg with a split-learning approach that is effective if the data distributions differ for each client. Our contributions are: (i)We confirmed the effectiveness of FL in distributed refrigeration and air-conditioning systems; (ii)We clarified that to reliably detect anomalies, we need to adjust the method of dividing the network configuration of FedRep into aggregation and local learning-only sections, and which layer to insert padding into.

The remainder of this paper is organized as follows: Section 2 reviews the related work in this area; Section 3 describes the target refrigeration and air-conditioning systems with IoT sensors in detail. Section 4 explains the proposed approach; Section 5 reports the experimental results, and Section 6 concludes the paper with a summary of the findings.

II. Related work

Extensive research has been conducted on anomaly detection systems using FL, including IoT-related applications and algorithm evaluations. Tran et al. [7] proposed status monitoring of manufacturing systems using long short-term memory (LSTM). Costa et al. [8] performed a comparative evaluation of CO2 time-series data collected by IoT sensors with local learning using LSTM, Heinrich et al. [9] developed software to perform data augmentation to generate synthetic time series datasets and simulate FL, Gkillas et al. [10] proposed a method based on neural network (NN) pruning for IoT applications. In addition to anomaly detection, research is progressing in numerous fields, including intrusion detection [11] and medicine [12]. Although these studies have primarily focused on time-series data, they do not provide details regarding appropriate network models or the configuration of the hyperparameters. In addition, no research has been conducted on refrigeration or air-conditioning systems.

In this study, for anomaly detection, we applied FL using a CAE to an anomaly detection system for refrigeration and air-conditioning systems using vibration sensor data, and obtained guidelines for constructing appropriate models from the perspective of network configurations and hyperparameters in FL.

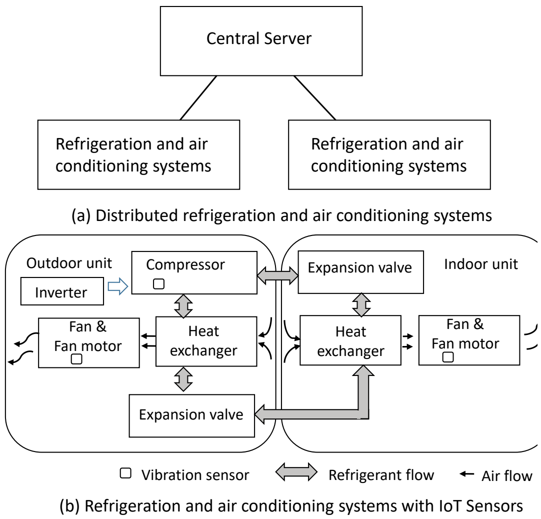

Fig. 1. Target system configuration

#	Sensor type	Measured amount(unit)	Sampling rate
1	Vibration sensor	Acceleration(m/s^2)(3axis)	50Hz

III. System configuration

Fig. 1 illustrates the configuration of the distributed refrigeration and air-conditioning systems. The target system was configured such that the refrigeration and air-conditioning units were installed at distributed locations, with a central server for data aggregation in FL. They comprised both indoor and outdoor units. The compressor, which functions as the main power unit, increases the refrigerant pressure and temperature through compression. The expansion valve decreases the refrigerant pressure and temperature by channeling it through a narrow passage while simultaneously controlling the flow rate and degree of superheating automatically. Air movement in both the indoor and outdoor areas is handled by a fan motor, and the heat exchanger enables the transfer of heat between the refrigerant and surrounding air in each environment. Along with the temperature controller, the inverter adjusts the compressor performance to maintain the target temperature.

Based on a prior experiment [1], three-axis vibration sensors were installed in outdoor and indoor fan motors and compressors to detect equipment failure in the early stages. In this study, data from a sensor installed on an outdoor fan motor were evaluated.

IV. Anomaly detection method

A. Data collection condition

Table I lists the data collection conditions of the vibration sensors. Based on sensor specifications, the sampling rate of each vibration sensor was set to 50 Hz. Vibration data were collected using the method employed in a previous study with an IoT gateway [1].

Layers	Convolution/Deconvolution	Number of filters	Filter size
1-2	Convolution	32	3
3-4	Convolution	64	3
5-7	Convolution	128	3
5-7	Deconvolution	128	3
3-4	Deconvolution	64	3
1-2	Deconvolution	32	3

B. Data analysis procedure

Our study employed a CAE [6] to analyze the vibration data. Autoencoders [13], which are NNs designed to model normal operational conditions, were utilized with the CAE specifically incorporating convolution and deconvolution layers for data processing (see Table II). The input to the CAE comprised 324-point amplitude spectrum(reshaped to 18×18), derived from fast Fourier transform (FFT) calculations for each three-axis vibration measurement. The size and number of filters were determined through preliminary experiments [1]. As the input data reflect normal operating conditions, the gap between the input and output remains relatively minor. Anomalies were identified by checking whether these gaps exceeded a defined threshold. The anomaly score E is calculated as the mean squared error derived from the comparison of the predicted data x and the observed data x'.

$$E = \frac{\sum_{n=1}^{N}(\parallel x_n - x'_n \parallel)^2}{N} \tag{1}$$

where the dimension N of x and x' is 972(i.e.,324 $\times$ 3). We applied FL using the CAE by implementing FedAvg and FedRep. FedAvg is a learning method in which a server provides a weighted average of the weights of each layer in the client models based on the amount of data available from each to generate a training model that reflects the volume of data available to each client. By contrast, FedRep is a learning method that divides the model into a client-only trained section, trained on data available only to a given client, and a server-aggregated section, whose updates are aggregated across clients at the server. This approach enables the model to learn shared knowledge while generating personalized models for each client. We outline the analytical procedure for FedAvg and FedRep as follows.

- FedAvg:

Step1: Each client model trained locally and sends its parameters to the server.

Step2: The server calculates the weights of each layer of the client models using a weighted average according to the number of data samples sent by the clients to update the global model.

Step3: The server distributes the updated model to each client.

Step4: Each client performs training locally based on the updated model and thus returns to Step1.

- FedRep:

#	Data group	Device type/ Sensor location	Data collection period	Operating condition
1	Data group1	Air conditioner Outdoor fan motor	From 12/9/2020 to 1/31/2021	Operation mode: Heatin Temperature settings: 20
2	Data group2	Air conditioner Outdoor fan motor	From 7/7/2021 to 10/31/2021	Operation mode: Coolin Temperature setting: 26 °
			From 11/1/2021 to 12/23/2021	Operation mode: Heatin Temperature setting:20 °
3	Data group3	Refrigerator Outdoor fan motor	From 9/16/2022 to 12/26/2022	Temperature settings: -23

Step1: Each client model learns locally and sends its pa eters to the server.

Step2: The server calculates the average of the weights of layer in the global model(the model aggregation sec according to the number of data samples provide each client.

Step3: The server updates the aggregated global model an distributes it to each client.

Step4: Each client performs training locally based on th updated model and thus returns to Step1.

The FedRep network model division which represent server-aggregated section and a client-only trained section i evaluated in the following three patterns .

Pattern1:

Convolution layers 1–7 were used for the serve aggregated section, whereas the other layers were use for the client-only trained section.

Pattern2:

Convolution layers 5-7 and deconvolution layers 7-5 wer used for the server-aggregated section, whereas the other layers were used for the client-only trained section.

Pattern3:

Convolution layer 7 and deconvolution layer 7 were used for the server-aggregated section, whereas the other layers were used for the client-only trained section.

V. EXPERIMENTAL RESULTS

A. Test items and condition

We performed an evaluation using FL with the three data groups listed in Table III. Data group1 and data group2 are different in the collection period of the air conditioner data. Data group3 included refrigerator data.

B. Learning models and evaluation data

From each data group listed in Table III, three pseudo-clients and three types of learning models were constructed for evaluation, as summarized in Table IV. The training data were part of the normal operating data from the training period, excluding days on which failure testing was conducted.

The evaluation objectives for each learning model are as follows:

Model1:

Evaluating data from different periods(seasons) as if they were from different clients.

TABLE IV

Learning model	Client	Data group	Learning data period and type	Number of learning data(Million)
Model1	Client1	Data group1	From 12/9/2020 to 1/25/2021(AC:H)	30
	Client2	Data group2	From 9/11/2021 to 10/24/2021(AC:C)	36
	Client3	Data group2	From 7/7/2021 to 7/13/2021(AC:C)	15
Model2	Client1	Data group1	From 12/9/2020 to 1/25/2021(AC:H)	30
	Client2	Data group2	From 9/11/2021 to 10/24/2021(AC:C)	36
	Client3	Data group3	From 9/22/2022 to 11/28/2022(R)	16
Model3	Client1	Data group1	From 12/9/2020 to 1/25/2021(AC:H)	30
	Client2	Data group2	From 9/11/2021 to 10/24/2021(AC:C)	4
	Client3	Data group2	From 7/7/2021 to 7/13/2021(AC:C)	4

AC:H Air Conditioner Heating AC:C Air Conditioner Cooling R: Refrigerator

TABLE V

Test case	Learning model	Target client	Evaluation data
Case1	Model1	Client1	12/15/2020 failure test(AC:H)
Case2	Model1	Client1	1/19/2021 failure test(AC:H)
Case3	Model1	Client2	12/23/2021 failure test(AC:H)
Case4	Model2	Client3	10/27/2022 Failure test(R)
Case5	Model3	Client2	12/23/2021 failure test(AC:H)

AC:H Air Conditioner Heating R: Refrigerator

Model2:

In Model1, the training data of Client3 was changed to a different type of equipment (refrigerator).

Model3:

In Model1, the amount of training data for Clients2 and 3 was reduced to evaluate the imbalance in data volume.

The evaluation data and test cases obtained using the aforementioned models are summarized in Table V. In each case, the data from the failure test conducted during the data collection period for each data group were used as evaluation data. In these failure tests, the volume of air was reduced by 40%, 50%, and 80% by adjusting the closing rate of the intake port in the outdoor fan motor. Of note, in the evaluation, we simulated multiple clients as well as a central server that performed the aggregation on the same computer.

C. Test results

Experiments were conducted with and without padding for all the layers in the network model. Table VI summarizes the anomaly detection accuracy for each test case and the network model without padding. Here, the anomaly detection accuracy is determined by calculating the maximum value of the abnormal section and the average value of the normal section during the failure test period from the moving average of the anomaly score (window width = 50), and then determining the ratio of these values(see Fig. 2). This indicates the separation performance between normal and abnormal sections and also applies to subsequent evaluations. Here, FedRep1, 2 and 3 represent division patterns1, 2, and 3, respectively. Table VI summarizes the model with the highest accuracy in each case, as shown with the hatched cells. In each case, the accuracies of FedAvg and FedRep(especially FedRep) were

Analysis method	Case1	Case2	Case3	Case4	Case5
Local	2.21	1.76	3.83	2.71	3.3
FedRep1	2.71	1.61	3.5	3.02	3.1
FedRep2	3.07	1.80	4.32	2.84	4.75
FedRep3	2.81	1.83	3.47	3.01	4.44

Analysis method	Case1	Case2	Case3	Case4	Case5
Local	8.15	4.14	3.27	2.75	4.74
FedRep1	39.4	8.80	34.2	1.60	6.36
FedRep2	21.4	10.0	5.58	1.85	4.78
FedRep3	12.2	9.10	9.0	2.96	5.90
FedAvg	54.1	37.27	42.82	2.97	21.95

generally higher than those of Local. This demonstrates the effectiveness of FL in general and FedRep in particular. In the case with imbalanced training data (Case5), the accuracy of Local decreased compared with the case without imbalance (Case3), but that of FedRep2, 3 and FedAvg did not decrease, which demonstrates the effectiveness of FL. FedRep2 and 3 were generally more accurate than FedRep1. Fig. 2 shows the transition in the CAE anomaly score for Case2 with the Local, FedAvg and FedRep2 without padding. The threshold was set to normal and abnormal status can be distinguished from the result to the maximum extent. In such cases, we can distinguish between normal and abnormal conditions. Hereinafter, the jumps on the right end of the graphs showing the transitions of each CAE anomaly score are due to the influence of the work based on the experiment and are not abnormal.

Similarly, Table VII lists the anomaly detection accuracy with padding for all layers and summarizes the models with the highest accuracy for each case, as shown by the hatched cells. As in the case without padding, the accuracies of FedAvg and FedRep were generally more accurate than that of Local. However, FedAvg exhibited the highest accuracy. FedRep1 and 2 were generally more accurate than FedRep3. Fig. 3 shows the transition of the CAE anomaly score in Case2 with padding for all layers. The model with padding exhibited improved accuracy compared with the model without padding. This is because the medium-layer data size of the network is maintained by convolution and deconvolution such that data close to the training data can be well restored. However, we observed overfitting where the accuracy was extremely high, particularly for FedAvg. In this case, detection may not occur in the areas that should be detected (see the hatched parts in Fig. 3, as in most of the other cases). This implies that even areas with a low anomaly score in the abnormal section(40% and 50% reduction in air volume) could be restored, which would result in failing to detect a valid anomaly.

Fig. 4 shows the input and output data of the learning model

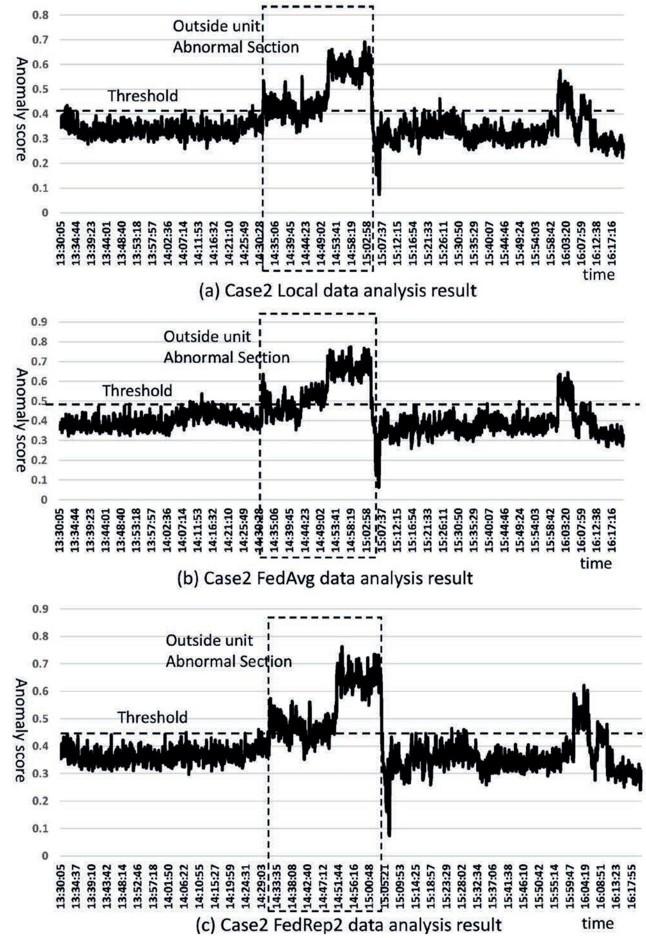

Fig. 2. Transition in CAE anomaly score in Case2(without padding:Local,FedAvg and FedRep2)

in Case2 with FedRep1 (with an amplitude spectrum derived from the FFT). The left side shows the normal section with a specific point, whereas the right side shows the abnormal section with the highest anomaly score. It can be observed that with padding for all layers, the case is restored better than that without padding because the input and output match well. This supports the results of the accuracy evaluation.

D. Padding evaluation

Following the results of Subsection C, we evaluate the effects of the padding value of each layer on the anomaly detection accuracy. We evaluated Local, FedAvg and FedRep1, 2, 3 for cases where padding was added to the inner n(n:1,3,5) layers and the outer n(n:1,3,5) layers in Case2. For instance, Inner1 indicates that padding has been inserted into convolution layer 7 and deconvolution layer 7, and Outer1 indicates that padding has been inserted into convolution layer 1 and deconvolution layer 1. Table VIII lists the anomaly detection accuracy for each network model according to padding. Areas of overdetection (normal areas determined to be abnormal) were observed in parts of Local, FedAvg, and FedRep1 (light-hatched cells). In contrast, the darkly hatched cells are those

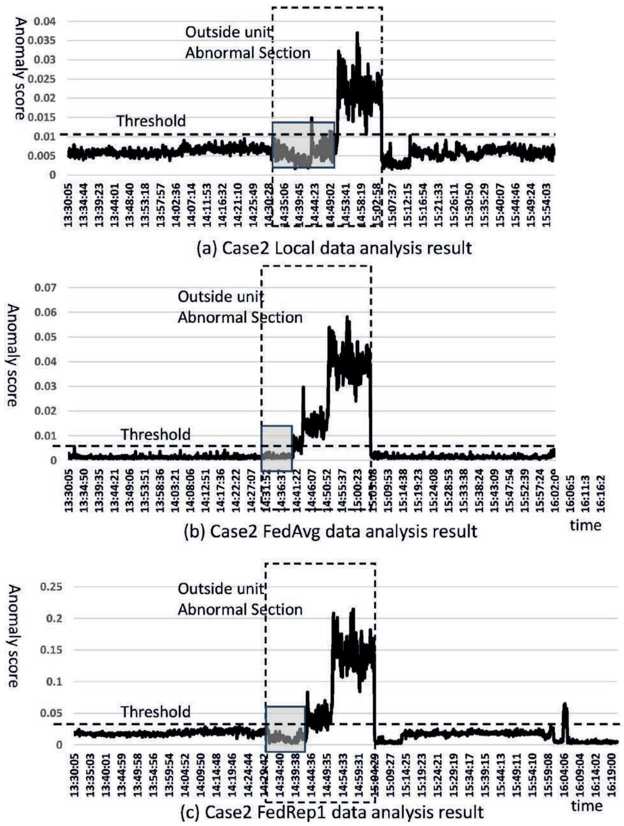

Fig. 3. Transition in CAE anomaly score in Case2(with padding for all layers:Local,FedAvg and FedRep1)

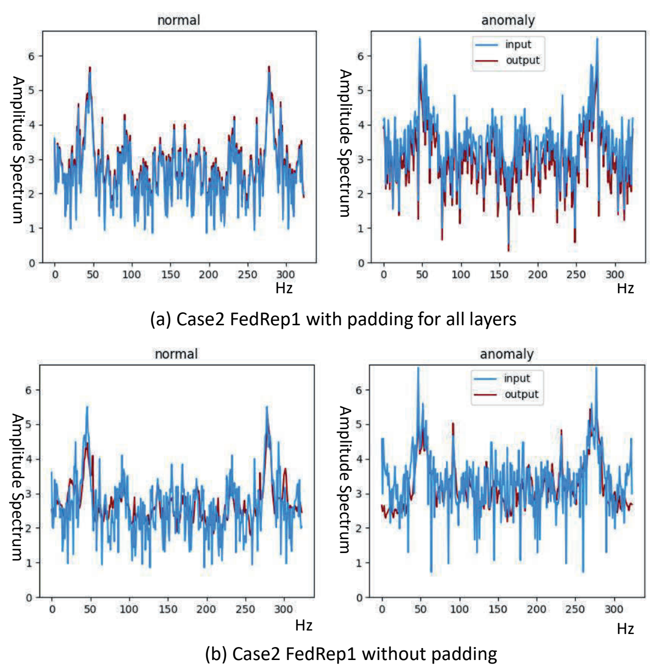

Fig. 4. Differences between input and output data of Case2 FedRep1 with padding for all layers and without padding

where nondetection areas occur, as explained in the case of padding for all layers. The accuracy of inner n and outer n

#	Padding	Intermediate-layer data size	Local	FedRep1	FedRep2	FedRep3	FedAvg
1	No padding	16	1.76	1.61	1.80	1.83	1.75
2	Inner1	36	2.00	1.85	1.81	1.83	2.0
3	Inner3	100	2.06	1.50	1.78	1.64	2.03
4	Inner5	196	2.01	2.38	2.00	2.29	1.97
5	Outer1	36	2.04	1.94	1.90	2.15	1.99
6	Outer3	100	1.99	1.99	2.36	2.22	2.04
7	Outer5	196	2.03	1.99	7.84	2.00	2.06
8	Padding for all layers	324	4.14	8.80	10.0	9.10	37.27

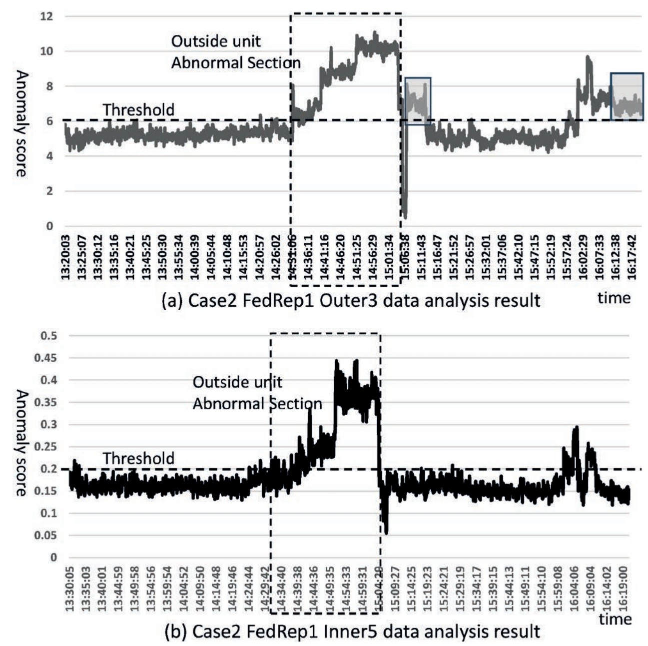

Fig. 5. Transition in CAE anomaly score in Case2(Padding evaluation)

was approximately halfway between that without padding and that with padding for all layers. Black-framed cases achieved higher accuracy without overdetection or nondetection. Table VIII lists the intermediate-layer data sizes for each padding case. The accuracy did not increase with the intermediate-layer data size, but Inner3/Outer3 and Inner5/Outer5 were generally more accurate than Inner1/Outer1. Fig. 5 shows the transition of the CAE anomaly score in Case2 of the padding evaluation. Fig. 5(a) shows overdetection case(see hatched parts) and Fig. 5(b) shows an appropriate case. Thus, FedRep is considered an appropriate model with no overdetection or nondetection and appropriate accuracy. To fully obtain the effect of FedRep, it is necessary to create an appropriate anomaly detection model by appropriately selecting the padding along with the division pattern.

Fig. 6 shows the input and output data for Case2 with FedRep1 Inner5. Compared with Fig. 4, the degree of agreement was equal to or less than that with padding for all layers and greater than that without padding.

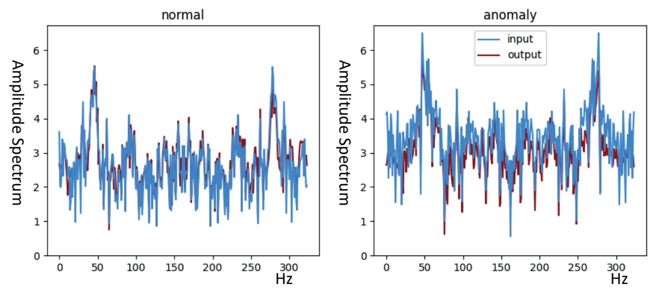

Fig. 6. Differences between input and output data of Case2 FedRep1 Inner5

E. Summary

The above results are summarized as follows:

- The accuracies of FedAvg and FedRep was higher than that of Local, demonstrating the effectiveness of FL.
- When the amount of training data is unbalanced, without padding, the accuracy of Local drops compared with when no imbalance occurs, but those of FedRep and FedAvg improve, demonstrating the effectiveness of FL.
- In FL using a CAE, padding at each layer affects accuracy. With padding for all layers, although the accuracy improves, nondetection or overdetection occur. FedRep is considered an appropriate model with no overdetection or nondetection and appropriate accuracy. In FedRep, we should create an appropriate anomaly detection model by appropriately selecting padding along with the division pattern.
- The method of dividing the network in FedRep affects the accuracy; however, examining the factors behind this will be a future challenge.

VI. Conclusion

In this study, we applied FL to refrigeration and air-conditioning systems distributed across multiple locations and evaluated the anomaly detection accuracy of FL compared with local learning at each location. Consequently, the following points were clarified:

- FL, particularly FedRep, was shown to be effective for anomaly detection using sensor data in distributed refrigeration and air-conditioning systems.
- In FedRep, it was revealed that in addition to the method of dividing the aggregated layers, we need to adjust padding to obtain appropriate accuracy without failing to detect or overdetect abnormalities.

The findings of this study were supported by experimental results. Future study will need to clarify why the way the network is divided in FedRep affects accuracy.

References

[1] T.Hirata,K.Yoshida,K.Koido,S.Takahashi,Anomaly detection in air conditioners using IoT technologies,Proceedings of COMPSAC2021, pp.1553-1559, Jul.2021.

[2] T.Hirata,Y.Kuriyama,T.Hinohara,S.Takahashi,K.Yoshida,T.Ogasawara,M.Go, Model Creation Method for Anomaly Detection in Refrigeration and Air Conditioning Systems,Proceedings of ICOIN2024, pp.29-34, Jan.2024.

[3] T.Hirata,Y.Kuriyama,T.Hinohara,S.Takahashi,K.Yoshida,T.Ogasawara,M.Go, Improved Anomaly Detection in Air Conditioners Using IoT Technologies, Proceedings of ICOIN2025, pp.505-509, Jan.2025.

[4] H.B.McMahan, E.Moore, D.Ramage, S.Hampson, B.A.y.Arcas, Communication-Efficient Learning of Deep Networks from Decentralized Data, Proceedings of International Conference on Artificial Intelligence and Statistics, 2017, p.1273-1282.

[5] L.Collins, H.Hassani, A.Mokhtari, S.Shakkottai, Exploiting Shared Representations for Personalized Federated Learning, Proceedings of the 38th International Conference on Machine Learning, PMLR 139, 2021.

[6] A. Alqahtani, X. Xie, J. Deng, M.W.Jones: A Deep Convolutional Auto-encoder with embedded clustering,2018 25th IEEE International Conference on Image Processing.

[7] D.H.Tran, V.L.Nguyen, I.B.K.Y.Utama, Y. M.Jang, An Improved Sensor Anomaly Detection Method in IoT System using Federated Learning, 2022 Thirteenth International Conference on Ubiquitous and Future Networks.

[8] L.F.D.Costa, L.S.Furtado, P.H.G.Rocha, P.A.L.Rego, F.A.M.Trinta, Time series prediction in IoT: a comparative study of federated versus centralized learning, 2023 IEEE 20th Consumer Communications & Networking Conference.

[9] F.Heinrich,T.Egger,H.Ghaeni,B.Kormann,F.Wenninger, Enabling Federated Learning: Generating Synthetic Clients through Time Series Data Augmentation, 2025 Smart Systems Integration Conference and Exhibition.

[10] A.Gkillas,A.Lalos,Resource Efficient Federated Learning for Deep Anomaly Detection in Industrial IoT applications, 2023 24th International Conference on Digital Signal Processing.

[11] S.Ghosh,A.S.M.M.Jameel, A.E.Gamal, Improving Transferability of Network Intrusion Detection in a Federated Learning Setup, 2024 IEEE International Conference on Machine Learning for Communication and Networking.

[12] A.Korkmaz,A.Alhonainy,P.Rao, An Evaluation of Federated Learning Techniques for Secure and Privacy-Preserving Machine Learning on Medical Datasets, 2022 IEEE Applied Imagery Pattern Recognition Workshop.

[13] G.E. Hinton and R. Salakhutdinov. Reducing the dimensionality of data with neural networks. Science, Vol.313, No.5786, pp.504-507, 2006.