Federated Learning With Potential Partnership Identification for Accurate Prediction in Flexible Manufacturing System

Chao Ren, Bin Jiang, *Fellow, IEEE*, and Ningyun Lu, *Member, IEEE*

***Abstract*—In recent years, intelligent manufacturing has integrated industrial data and artificial intelligence technology, which has been a widely concerned development direction in the manufacturing industry. Industrial data integrity is the key factor for the successful implementation of intelligent manufacturing. However, in flexible manufacturing systems with multivariety and small-batch, it is hard to collect production data from all working conditions. Actually, for the purpose of status monitoring, data acquisition and annotation on complex mechanical components is also time-consuming and labor-intensive, which requires the assistance of professional domain knowledge. Faced with the challenge of incomplete data quantity and quality, federated learning is a promising paradigm of collaborative modeling, which ensures data privacy and fully utilizes distributed data information from different industrial users. However, due to the heterogeneity of data among industrial users, cooperation benefits cannot satisfy all industrial users. In this article, a novel federated learning cooperation framework is proposed to guide participants to choose the appropriate coalition and improve the benefits of participants. In this framework, the self-organizing incremental neural network is employed to generate prototypes that can effectively capture the distributional characteristics of raw data, obviating the necessity for industrial users to provide their raw data and labels. It offers recommendations for industrial users to foster collaboration by assessing the similarity among these prototypes. The collaborative tool wear prediction experiments demonstrate the effectiveness of the framework on industrial data.**

***Index Terms*—Federated learning (FL), flexible manufacturing systems, intelligent manufacturing, potential partnership.**

I. Introduction

DUE to the rapid development of artificial intelligence, big data, cloud and edge technologies, etc., more and more enterprises are investing in intelligent manufacturing to optimize and manage production and product transactions proactively. In the process of transitioning to intelligent manufacturing, manufacturers employ two primary technical approaches. The first is control and optimization technology, while the second is predictive monitoring and maintenance technology. Control and optimization methods are implemented for tasks like inventory optimization, resource optimization, and process control. Concurrently, the implementation of predictive monitoring and maintenance methods facilitates advancements in product quality monitoring, equipment maintenance, and yield prediction.

Equipment maintenance is the foundation of intelligent manufacturing, which can be achieved through data-driven predictive maintenance and automation optimization, ensuring stable and efficient operation of the equipment. With the increasing emphasis on device operation and maintenance, as well as the continuous advancement of artificial intelligence technology, prognostics and health management (PHM) has emerged as a prominent research field in both academia and industry, such as fault detection on rotating component [1], remaining useful life prediction of systems exhibiting degenerative characteristics [2], [3], [4].

Large and comprehensive data are the key to effective PHM. However, in flexible manufacturing systems, it is difficult to obtain complete data. In terms of data collection, many issues may limit the number of samples for modeling. First, assembling sensors on machine equipment, especially high-end devices like computer numerical control (CNC) machine tools, poses a challenge as it is difficult to do so without compromising the reliability and security of the equipment. In this situation, companies typically modify a pilot machine. However, the conditions for data collection experiments are often limited, which results in the collected data not being able to reflect all the working conditions of the machine. Second, data acquisition on complex mechanical components is time-consuming and labor-intensive, and high-quality data annotation requires the assistance of professional domain knowledge. Taking the measurement of tool wear as an example, a sample can only be obtained after a complete processing process on the workpiece [5]. It is unrealistic to collect and annotate samples of all operating conditions. In conclusion, it is difficult for a single industrial user to collect a comprehensive dataset.

In order to obtain sufficient data for modeling, a new idea is to make industrial users who have similar production lines collaborate. Industrial users share their incomplete data to compose a relatively comprehensive training dataset aiming for improved model performance. Nevertheless, due to conflicts of interest, even if industrial users have machines with similar functions

Manuscript received 28 September 2023; revised 17 January 2024 and 16 April 2024; accepted 3 June 2024. This work was supported in part by the National Key R&D Program of China under Grant 2021YFB3301300, in part by the National Natural Science Foundation of China under Grant 62273176, in part by the Natural Science Foundation of Jiangsu Province of China under Grant BK20222012, and in part by the Postgraduate Research & Practice Innovation Program of Jiangsu Province under Grant KYCX24_0588. Associate Editor: Steven Li. *(Corresponding author: Bin Jiang.)*

The authors are with the College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China (e-mail: aloysren@nuaa.edu.cn; binjiang@nuaa.edu.cn; luningyun@nuaa.edu.cn).

Digital Object Identifier 10.1109/TR.2024.3427813

and structures, they are reluctant to share data collected on their production line, which results in a "data island" problem in real industry.

Federated learning (FL) is proposed as a collaborative machine learning framework [6], which allows more participants to jointly train models without sharing data, and make full use of the decentralized data. There have been some research works employing FL paradigm in various real-world applications for condition monitoring. For example, Li et al. [7] incorporated adversarial learning into the FL framework to solve the domain shift problem in bearing fault diagnosis. Qin et al. [8] conducted distributed modeling by adopting train bogie data from multiple railways, identifying the faults of air springs, lateral dampers, and antiyaw dampers. Lu et al. [9] proposed a collaborative training paradigm for offshore wind turbine fault diagnosis, which includes an event-triggered mechanism to reduce communication costs and computational workload at the stage of local model parameter transmission. Chung et al. [10] proposed a federated analysis framework aimed at enhancing model generalization, specifically addressing the challenge of the forgetting of old knowledge when the learned model is adapted to new clients. Due to the advantages of FL, it has the potential to solve the problem of insufficient industrial small samples by aggregating data from industrial users.

However, FL cannot guarantee to bring ideal benefits to cooperative industrial users, because FL cannot guarantee the training effect in the presence of data heterogeneity. Collaborating without considering the quality of data in the hands of industrial users is likely to result in counterproductive consequences due to negative transfer [11]. More importantly, in reality, once cooperation begins, it cannot be easily terminated before the end of the cooperation agreement, and the costs invested by industrial users in data acquisition, computation resources, and communication resources will be wasted if the cooperation partners are not suitable. In reality, before reaching a cooperation agreement, users usually consider the conditions of their potential partners, and data quality is one of the most important considerations in FL paradigm. For user R, potential partners of user R are defined as users willing to collaborate and capable of contributing to user Rs benefits before the formal cooperation is reached. Identifying potential partnership in advance during the decision-making stage can bolster confidence and benefits in cooperation. As far as we know, most FL research focus on improving performance, such as client selection technologies [12], model training and optimization methods [13], [14], [15], and model updating strategies [16]. The potential assumption of the abovementioned research is that participants have formally initiated collaboration.

Note that there have been studies on clustering FL to determine cooperative relationships during the training process. The clustering of clients can be deployed on the server or client side. For instance, in [17], a postprocessing clustering method is proposed based on the gradients of clients updated during training stage. On the other hand, the client can also adaptively select clustering categories. In [18], a machine learning model is trained in each cluster, and each client runs the updated models and is relocated to the best-fitting cluster. The number of models corresponds to the number of clusters. However, no matter which type of clustering is chosen, the abovementioned methods consume a large amount of computing and communication resources when forming clusters in large-scale setting [19], [20]. In addition, some advanced graph FL methods have been proposed to evaluate the relationship of clients. Yao et al. [21] relaxed the assumption that the number of clients and client relationships remain unchanged, and solved the personalized model update problem brought by newly participating clients by constructing a collaboration graph. The process of finding trustworthy neighbors also occurs in the stage where cooperation has already begun. If potential cooperative relationships between clients can be determined before training model, it will eliminate the tedious clustering steps in training process and save resources. To achieve this goal, the key lies in how to obtain the distribution information of client data without privacy leakage.

Self-organizing incremental neural network (SOINN) [22] is an effective method on evaluating distribution information of user data. SOINN is a two-layer neural network based on competitive learning. The incremental nature of SOINN enables it to discover new patterns in data streams and learn them without affecting the results of previous learning. SOINN generates prototype samples in an unsupervised manner that can represent the distribution of the original samples. The introduction of SOINN transforms the evaluation of user data quality into a problem of evaluating prototype samples. The K-means clustering method offers the benefits of straightforward computation, rapid processing speed, and provides a convenient means to assess the similarity of prototype samples.

This article focuses on the issue of participant uncertainty cooperation in the user's decision-making phase. An federated learning with potential partnership identification (FLPPI) framework is proposed. The core of the framework is to identify potential cooperative relationships before participants collaborate so that they can choose suitable partners as much as possible and filter out unsuitable participants. The main contributions can be summarized as follows.

1) The willingness of industrial users to collaborate during the decision-making stage before FL is considered, a topic that has received relatively little attention in the current literature. Based on the basic fact that users need to invest costs as soon as they start cooperating, this method starts from the user's perspective and focuses on their potential partner selection before cooperation.
2) The proposed framework evaluates the similarity of user data while protecting user data privacy, and conveniently and quickly recommends potential cooperative alliances to users. This framework enables participants with approximately homogeneous data to form coalitions.
3) An SOINN generates a small number of prototype samples for each user in a shared instance space, and the similarity of the original data is evaluated by the similarity of the prototype samples in the instance space.
4) The core concept of this research has the potential for broad application in facilitating collaboration among industrial users in diverse industrial settings, such as critical

component life monitoring and mechanical component fault diagnosis.

The rest of this article is organized as follows. Section II introduces related work of FL from the perspective of user cooperation benefits. The proposed FLPPI framework is introduced in Section III. In Section IV, key component state prediction tasks in manufacturing system are conducted to verify the effectiveness of proposed framework. Finally, Section V concludes this article.

II. Related Work

This section provides an overview of pertinent research regarding user cooperation interest protection within the context of FL. In terms of the node of selecting cooperation, incentivized FL represents a strategy for safeguarding user interests postcooperation, whereas cooperative evaluation-based FL is an approach aimed at protecting user interests prior to cooperation.

A. FL With Incentive Mechanism

Users are unwilling to participate in FL without any return, especially considering that they need to pay the cost of computation and communication resources. The incentive mechanism is proposed to encourage them to tolerate these costs and make contributions. In the design of incentive mechanisms, rewards can be determined by the participant data contribution [23], participant reputation [24], and participant resources [25]. Song et al. [23] proposed a measure called contribution index used to quantify the contribution of data providers. The contribution index can be calculated from the intermediate results of FL avoiding extra rounds of model training. Kang et al. [24] entrusted user reputation management to consortium blockchain. An effective incentive mechanism is designed based on contract theory, in which higher reputation workers can obtain more rewards from the task publisher. Khan et al. [25] considered computation resources and communication resources in the formulation of the reward rate. The local training strategies of users are optimized under dynamic reward rates, and ultimately the Stackelberg equilibrium is achieved between users and the base station.

In nature, FL with an incentive mechanism is a conditional cooperation framework among participants. They differ from our approach in that these methods are based on the premise that cooperation has already been executed, while our method is adopted before participants decide whether to cooperate and with whom.

B. FL Based on Collaboration Evaluation

Researchers have noticed that collaborating with all clients is not necessarily the optimal solution for each client. Participants should be given the right to voluntarily choose their partners. Choosing partners who can improve the performance of the local model is a better strategy compared to uniform federation. Donahue et al. [26] view the collaboration problem in FL as a coalitional game. In the three forms of FL, core-stable allocation schemes are derived theoretically. It is noted that theoretical derivation is currently limited in mean estimation tasks and linear regression tasks. For nonlinear modeling scenarios, further cooperative game research is needed. For the generalized modeling tasks, Cui et al. [11] summarized the selfishness principle and rationality principle for individuals in cooperation, and the optimal collaborator sets are determined under the two principles. The cooperation payoff is determined by the performance of the model applied to the local dataset, which means each participant needs to provide a portion of raw data to train and verify model performance.

In our work, the evaluation of cooperation does not require raw data, especially key label information, which protects data privacy. In the meanwhile, the proposed framework provide participants with a low computational complexity method to select potential partners.

III. Proposed Method

A. Problem Definition

Suppose there are N industrial users with their own data $D_a = \{X_a, Y_a\}, a \in \{1, 2, \ldots, N\}$. Each user could learn a model $M^a(\theta^a)$ based on local data, θ^a is the parameter of model $M^a(\theta^a)$. To enhance model generalization, industrial users are actively seeking collaboration to make full use of data from others. Under the classic FL setup [6], the user first downloads the model $M(\theta_{t-1})$ from the central server in the tth round. Then, the model is trained in local data

$$\theta_{t-1}^a \leftarrow \theta_{t-1}^a - \delta \frac{\partial L}{\partial \theta_{t-1}^a} \tag{1}$$

where L is the loss between the inference results of model and the value of training samples. δ is the learning rate for model updating. After completing a specified number of model training iterations, each user uploads their model parameters θ_{t-1}^a to the central server. Then, the central server weights the model parameters uploaded by each user

$$\theta_t = \sum\nolimits_{a=1}^{N} \frac{n_a}{n} \theta_{t-1}^a \tag{2}$$

where n_a represents the data size of the ath user and n represents the total data volume. The model $M(\theta_t)$ is waiting to be downloaded by the users in the next round. After T rounds, each user gets the final model $M(\theta_T)$. Fig. 1 illustrates the overarching structure of the FL framework.

In flexible manufacturing, various industrial users have heterogeneous distribution data due to differences in production and processing parameters. On the one hand, when all participants engage in FL, uneven distribution of user data can lead to more local training rounds and communication rounds, resulting in higher costs for each user. On the other hand, for each user, once cooperation begins, it cannot be easily terminated. This means that inappropriate cooperation can result in wasted upfront costs. Therefore, it is foreseeable that when industrial users actively choose suitable partners rather than passively accepting cooperation, they can filter out unsuitable partners, which is more likely to achieve satisfactory benefits while avoiding unnecessary costs. Fig. 2 demonstrates the flowchart of federated cooperation

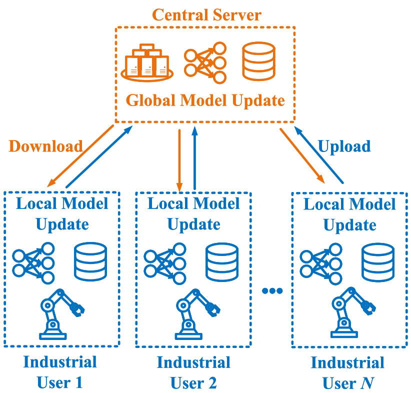


Fig. 1. General scheme of the FL framework.

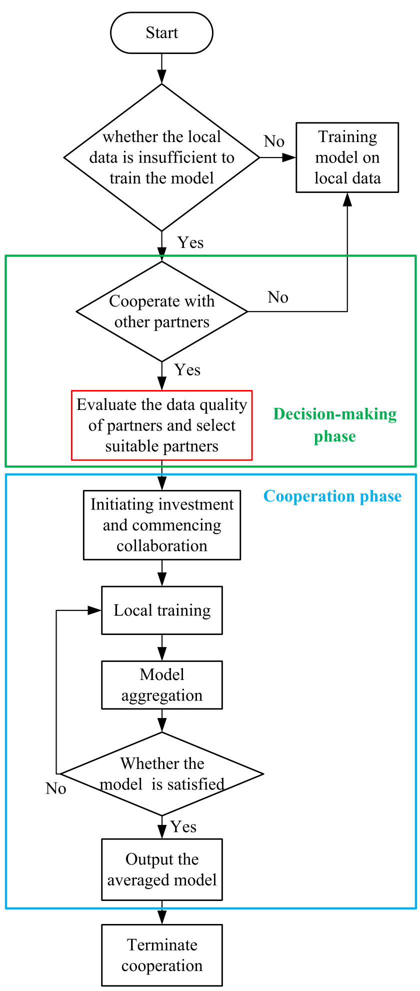


Fig. 2. Flowchart of the federated cooperation framework.

for a single user. User involvement in FL typically encompasses two key stages: the decision-making stage and the cooperation stage. Existing literature predominantly concentrates on research related to the collaborative stage, often overlooking user willingness to participate during the decision-making phase. In the highlighted red box, we evaluate the data quality of partners and select suitable partners, which is geared toward assisting users in making informed decisions before cooperating, thus helping them avoid resource wastage.

How to define a potential partnership? According to [17], FL contains an implicit assumption that the trained global model may simultaneously fit all clients data generating distributions. Hence, in this article, industrial users with data distributions that are alike are regarded as fitting collaborators. Prior to implementing FL, owing to data privacy concerns, users were unable to access the data held by other users, thus making it challenging to assess data quality. This article aims to evaluate the distribution of user data while protecting user data privacy, and then recommend suitable partners for users.

B. Federated Learning With Potential Partnership Identification

In the proposed framework, there are N industrial users who are willing to cooperate, and each user holds data $D_a = \{X_a, Y_a\}, a \in \{1, 2, \ldots, N\}$. The default data owned by industrial users is one-dimensional signals collected through industrial sensors, such as vibration signals, current signals, and so on.

1) Data Preprocessing: Each user acquires $D_a = \{(x_i^a, y_i^a)\}_{i=1}^{n_a}$ using a nonoverlapping sliding window consisting of S sampling points.

2) Data Compression and Reconstruction: In order to protect data privacy, the original data is compressed and reconstructed. The compressed and reconstructed data retains the potential global structure information and local neighborhood information of the original data. The characteristics of industrial processes are time-varying [27], so when processing data, it is also necessary to consider the dynamic nature of the data. Based on the abovementioned considerations, this article introduces the SOINN.

The advantages of employing SOINN are reflected in several aspects. First, SOINN creates a limited set of prototype samples to represent the original data, safeguarding data privacy. Second, SOINN autonomously processes raw data from individual users, eliminating the need for prior user knowledge. Users generate their own prototype samples within a shared instance space, ensuring subsequent similarity measurements are rational. Finally, SOINN generates prototype samples through incremental learning. As the prototype samples are gradually expanded, the distance information is fully preserved, which allows the prototype samples to capture the nonstationary characteristics inherent in time-varying data.

For ath user, randomly select two samples from D_a to create an initial prototype set $p_a = \{c_1, c_2\}$, $c_1, c_2 \in \mathbb{R}^S$. Then, the values and quantities of the prototype would be dynamically updated. Specifically, after inputting the ith sample x_i^a, find the two most similar prototypes in the prototype set p_a

$$s_{a1} = \arg\min_{c \in p_a} ||x_i^a - c|| \tag{3}$$

$$s_{a2} = \arg\min_{c \in p_a \backslash \{s_{a1}\}} ||x_i^a - c|| \tag{4}$$

where c represents a prototype in prototype set p_a. The two prototypes are activated and a neighborhood relationship (s_{a1}, s_{a2}) is established. When two prototype nodes are not activated simultaneously after a certain period, the connection is deleted. The distance threshold T of prototype node is usually taken as the maximum intraclass distance or the minimum interclass distance. For instance, calculate the distance threshold $T_{s_{a1}}$ of the prototype node s_{a1}. If s_{a1} has a neighbor set $B_{s_{a1}}$

$$T_{s_{a1}} = \max_{c \in B_{s_{a1}}} ||s_{a1} - c|| \tag{5}$$

else

$$T_{s_{a1}} = \min_{c \in p_a \setminus s_{a1}} ||s_{a1} - c||. \tag{6}$$

Whether to update the prototype set p_a depends on the distance between the new sample and the activated prototype node. If $||x_i^a - s_{a1}|| > T_{s_{a1}}$ or $||x_i^a - s_{a2}|| > T_{s_{a2}}$

$$p_a = p_a \cup \{x_i^a\}. \tag{7}$$

Else, the activated prototype nodes $\{s_{a1}, s_{a2}\}$ will be updated

$$s_{a1} = s_{a1} + \varepsilon(x_i^a - s_{a1}) \tag{8}$$

$$s_{a2} = s_{a2} + \varepsilon(x_i^a - s_{a2}) \tag{9}$$

where ε represents the step size for the update process. By dynamically adjusting prototypes, the prototype set will build a spatially ordered mapping to the input data. By repeating the aforementioned process, N industrial users can generate their respective prototype sets $P = \{p_1, p_2, \ldots, p_N\}$ where p_a represents the prototype sample set of the ath user. The number of prototype set P is

$$n'_P = \sum_{a=1}^{N} n'_a \tag{10}$$

where n'_a is the number of prototype set p_a.

In this step, we replace the user's original sample set with a prototype set, which effectively safeguards industrial user data privacy. On the one hand, the reconstructed prototype samples have a smaller quantity compared to the original samples. On the other hand, for supervised learning tasks, labels are indispensable. The prototype sample does not have a corresponding label. Unlike textual or image data, industrial sensor signals are difficult to annotate without prior knowledge.

3) Cluster Analysis: When industrial users select appropriate partners to establish an alliance, their data should exhibit similar distribution. K-means clustering provides a convenient means to assess data similarity. Randomly select K prototypes $\{\mu_1, \mu_2, \ldots, \mu_K\}$ from the prototype set P and calculate the category to which each prototype c_i belongs

$$C(i) = \arg\min_k ||c_i - \mu_k||^2 \tag{11}$$

where $C(i) \in \{1, 2, \ldots, K\}, \mu_k \in \{\mu_1, \mu_2, \ldots, \mu_K\}$. Next, proceed to update the prototype cluster center

$$\mu_k = \frac{\sum_{i=1}^{n'_P} 1\{C(i) = k\} c_i}{\sum_{i=1}^{n'_P} 1\{C(i) = k\}}. \tag{12}$$

Iterate through (11) and (12) until convergence is reached. All prototypes form K clusters $C_1, C_2, \ldots, C_K$. Each cluster is a collection of different prototype samples.

It should be noted that under the proposed framework, the value of K is determined using the Silhouette method [28]. Conduct experiments with various K values and calculate the Silhouette score F for each prototype cluster c_i

$$F(i) = \frac{h(i) - g(i)}{\max\{g(i), h(i)\}} \tag{13}$$

where $h(i)$ represents the average distance between the prototype point c_i and other points in the cluster to which it belongs, and $g(i)$ represents the minimum average distance between the prototype point c_i and points in other clusters

$$h(i) = \frac{1}{|C_I - 1|} \sum_{c_j \in C_I, j \neq i} d(c_i, c_j) \tag{14}$$

$$g(i) = \min_{J \neq I} \frac{1}{|C_J|} \sum_{c_j \in C_J} d(c_i, c_j) \tag{15}$$

where C_I is the prototype set containing c_i and $d(c_i, c_j)$ represents the distance measurement between two prototypes, usually taken as Euclidean distance.

It is worth noting that K-means is not the only option in cluster analysis. When applying the proposed framework, the selection of clustering methods needs to consider the characteristics of the data, the level of noise, the need for computational efficiency, and the requirement for interpretability of the results.

4) Coalition Formation: For ath industrial user, cast a vote for the category to which the prototype sample set $p_a = \{c_1, c_2, \ldots, c_a\}$ belongs, and the category that receives the majority of the votes from the prototypes is considered the optimal coalition for the user

$$U_a = \arg\max_k \sum\nolimits_{i=1}^{n'_a} \{C(i) = k\}. \tag{16}$$

Users belonging to the same coalition are considered potentially suitable partners. The quantity of coalitions established is either less than or equal to the number of clusters within the cluster, as there can be scenarios where a part of clusters may be not selected by any users.

5) Model Training: After the coalitions are established, the members of the coalition collaborate in accordance with the FL paradigm.

Fig. 3 illustrates the architecture of the FLPPI framework. When multiple industrial users express their willingness to participate in cooperation, local prototypes will be created and then shared with a third party. The third party will assess the similarity of all prototypes and recommend appropriate coalitions to the users. Subsequently, users proceed to the cooperation stage. The specific algorithm flow is described in Algorithm 1.

IV. Case Study

Flexible manufacturing systems typically have the characteristics of variable batch sizes and multiple varieties. After receiving different production orders, the workshop and production line will be flexibly arranged with production tasks

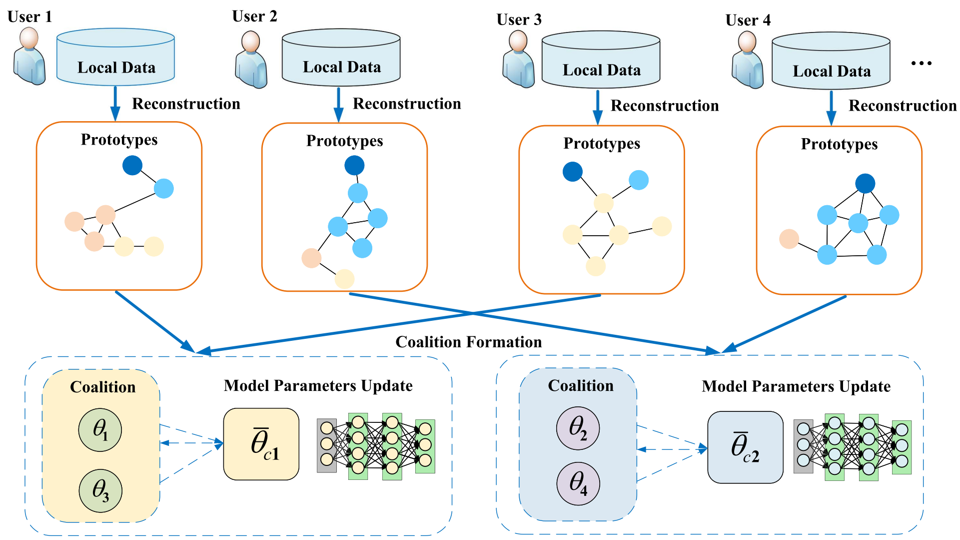


Fig. 3. Architecture of the FLPPI framework.

Algorithm 1: FLPPI.

Input: M: model structure, $D = \{D_1, D_2, \ldots, D_N\}$: local data of N users
Output: $M(\theta^1), M(\theta^2), \ldots, M(\theta^N)$: N local model

for all D **do**
  Reconstruct prototypes p according to (3)–(10)
end for
Send p to third-party
Using K-means to evaluate the similarity among the prototypes according to (11) and (12)
Form coalitions $U = \{U_1, U_2, \ldots, U_{N'}\}$ according to (16)
for all U **do**
  Initialize parameters of global model in each U
  for each global epoch **do**
    for each user in U **do**
      Download model from server
      for each local epoch **do**
        User train the model according to (1)
      end for
      Upload the trained model to server
    end for
    Perform model aggregation on the server side according to (2)
  end for
  The global model from U is distributed to every user within U
end for

according to the order requirements. To achieve this goal, a flexible manufacturing system owned by an industrial user consists of at least two or more CNC machine tools, a set of material transportation systems, and a computer control system. Different CNC machine tools are assigned different machining tasks to achieve collaboration. For example, some CNC machine tools are responsible for completing rough machining of products, such as drilling, milling, and other tasks; Some CNC machine tools are responsible for completing the precision machining of products, such as grinding, gear grinding, and other tasks. Due to the variability of order processing tasks, the working conditions of each machine tool in a flexible manufacturing system also change accordingly. It is unrealistic to collect sufficient high-quality samples from each working condition. FL frameworks can fully utilize a small number of samples from each industrial user for modeling while protecting data privacy.

Predicting wear in CNC machine tools is a common task in manufacturing processes. Accurate tool wear prediction models assist in determining the optimal time for tool replacement during machining operations, thereby ensuring consistent machining quality. The experimental section selected tool wear data from different working conditions to simulate the monitoring data of machine tools for different industrial users. This section assesses the performance of the model, which was trained by industrial users within the proposed cooperation framework.

A. Data Description

1) Milling Data: To evaluate the effectiveness of the proposed framework, the milling wear dataset provided by NASA was used in the case study [29]. This dataset is a cutting experiment conducted on a milling machine under different experimental conditions. Three types of sensors were used to collect current signals, vibration signals, and acoustic emission signals. We selected 5 operating conditions in the experiment, as shown in Table I. Assuming 5 different industrial users hold monitoring data for 5 different operating conditions.

2) Tool Wear Data: The experimental data for this study were obtained from the DMU 80P douBLOCK machining center of the NUAA Ideahouse team [30]. Nine solid carbide endmills of

TABLE I
FIVE SELECTED EXPERIMENTAL PARAMETERS

Number	Depth of cut (mm)	Feed (mm/rev)	Material
Z1	1.5	0.5	cast iron
Z2	0.75	0.25	cast iron
Z3	1.5	0.25	cast iron
Z4	1.5	0.25	steel
Z5	0.75	0.5	steel

Fig. 4. Machining process for TC4 on the experimental platform.

TABLE II
CUTTING PARAMETERS AND TOOL PARAMETERS

Number	Feed per tooth (mm/r)	Spindle speed (r/min)	Axial cutting depth (mm)
W1	0.045	1750	2.5
W2	0.045	1800	3
W3	0.045	1850	3.5
W4	0.05	1750	3
W5	0.05	1800	3.5
W6	0.05	1850	2.5
W7	0.055	1750	3.5
W8	0.055	1800	2.5
W9	0.055	1850	3

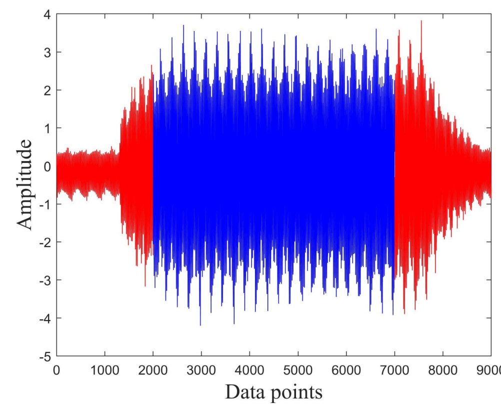


Fig. 5. Alternating current signals collected during a cutting process.

12×75×R1 were used to machine the pocket of titanium alloy (TC4) workpiece. The machining process involved utilizing a circular toolpath, moving from the inside to the outside. The machining process is shown in Fig. 4.

In the experiment, vibration signals were collected using $\mathrm{PCB}^{\mathrm{TM}}$ acceleration sensors. The spindle power signal was captured in real time from the internal programmable logic controller register of the machine tool using the object linking and embedding (OLE) for process control unified architecture communication protocol. Specifically, in the experiment, we selected vibration signals from two channels, along with spindle current and spindle power, totaling four sets of signals used for modeling. The tool wear label values were established by measuring the maximum width of the side wear area at specific intervals, employing a Sinico XK-T600 V microscope with a precision of 0.01 mm. Wear values ranged from 0 to 0.3 mm. The workpiece material utilized was TC4. The cutting process encompassed nine distinct working conditions, each configured based on varying parameters, including feed per tooth, spindle speed, and axial cutting depth. These conditions are comprehensively detailed in Table II. Suppose there are 9 industrial users, all of whom produce similar product types on their respective production lines. The industrial users possess data corresponding to the abovementioned working conditions, respectively.

B. Experimental Settings

1) Data Processing on Milling Data: The data collected during each cutting process is treated as a sample, as illustrated in Fig. 5. Taking the spindle alternating current signal as an example, due to ineffective data during tool approach and tool retraction, we extract the middle 5000 data points as new sample. A total of 12 features were extracted from the samples, as shown in Table III.

TABLE III
DETAILS OF FEATURE EXTRACTION

Time domain features		Frequency domain features
Mean	Root mean square	Mean of power spectrum
Median	Skewness	Mean square error of power spectrum
Variance	Crest factor	Skewness of power spectrum
Peak	Peak-to-peak	Peak of power spectrum

TABLE IV
KEY PARAMETER SETTINGS IN FEDERATED TRAINING

Hyperparameter	Value
Global rounds	100
Local epochs	20
Local batch size	10
Local learning rate	0.01
Optimizer	SGD
Number of total layers	4
Number of input layer nodes	49
Number of first hidden layer nodes	20
Number of second hidden layer nodes	20
Number of output layer nodes	1
Activation function	sigmoid

2) Data Processing on Tool Wear Data: Since the microscope recorded the wear labels in a discrete manner during the experiment, and the stored signal data was in the form of temporal data, it was necessary to perform interpolation on the wear labels. In this experiment, a linear interpolation method was used, cutting every 900 data points into a sample and labeling

TABLE V
TOOL WEAR PREDICTION RESULT OF DIFFERENT FRAMEWORKS

Framework	Metric	Average prediction error of tool wear under different frameworks								
		W1	W2	W3	W4	W5	W6	W7	W8	W9
FLPPI	MSE	0.0004	0.0037	0.0005	0.0009	0.0025	0.0009	0.0037	0.0014	0.0013
	MAE	0.0168	0.0526	0.0384	0.0266	0.0426	0.0246	0.0514	0.0341	0.0221
	R	0.9785	0.8393	0.9791	0.9787	0.932	0.878	0.9376	0.9759	0.8904
FedAvg	MSE	0.0013	0.0089	0.0049	0.0025	0.0058	0.0004	0.0053	0.0006	0.0018
	MAE	0.0325	0.0832	0.0636	0.0458	0.0722	0.0152	0.0584	0.0206	0.0267
	R	0.9432	0.7881	0.9635	0.9713	0.9296	0.968	0.9236	0.9729	0.8821
FedProx	MSE	0.0012	0.0097	0.0067	0.0022	0.0054	0.0014	0.0042	0.0016	0.0015
	MAE	0.0311	0.0867	0.0699	0.0439	0.0693	0.0259	0.0525	0.0427	0.0237
	R	0.9386	0.7498	0.9694	0.9658	0.9006	0.8761	0.9338	0.9676	0.8132
FedDyn	MSE	0.0036	0.0068	0.0058	0.0035	0.0031	0.0036	0.0087	0.0115	0.0052
	MAE	0.0511	0.0743	0.0641	0.0509	0.0465	0.0502	0.0869	0.0976	0.0612
	R	0.8116	0.6234	0.7254	0.7102	0.8762	0.7778	0.7378	0.6869	0.8051

the samples. The time-domain and frequency-domain features were extracted to improve the efficiency of model training and avoid dimensional disaster issues, as shown in Table III.

3) Performance Evaluation Metric: Each user randomly selects 60% of the data as the training set, and the remaining 40% of the data as the test set. In each coalition, users trained the local model in the training set and the third-party server average model weight as the global model. The performance of the global model trained in the coalition is tested on the test sets of each user. In this experiment, mean squared error (MSE), mean absolute error (MAE), and correlation coefficient (R) are employed as a metric to assess model performance

$$\mathrm{MSE} = \frac{\sum_{i=1}^{n}\left(y_i - y_i^p\right)^2}{n} \tag{17}$$

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n}\left|y_i - y_i^p\right|}{n} \tag{18}$$

$$R = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(y_i^p - \bar{y}^p)}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{n}(y_i^p - \bar{y}^p)^2}} \tag{19}$$

where y_i represents the value of label and y_i^p stands for the inference value of model. In addition, in order to eliminate the aleatoric uncertainty caused by the model trained under the FL framework, the model was trained 10 times, and the average error of the model on the test set was taken as the result.

C. Analysis of Prototype Distribution

To validate the effectiveness of SOINN used in the framework, it is essential to analyze the distribution of prototypes among users. Maximum mean discrepancy (MMD) [31] is introduced to reveal the distribution of raw data and prototype data. Nine users have the following numbers of original samples in their possession: 1489, 907, 701, 1585, 356, 659, 444, 1721, and 323. After data reconstruction, the number of prototype samples for each user is as follows: 41, 55, 83, 97, 116, 67, 63, 46, and 81. Figs. 6 and 7 describe the differences in distribution between the nine sets of user data and user prototype data. W1–W9 represents the raw data of 9 industrial users. S1–S9 represents prototype data for 9 industrial users. The distribution of the two heat maps is almost identical, indicating that the trend of changes in distribution differences remains fundamentally consistent and SOINN effectively preserves the distribution characteristics of user data. Simultaneously, in the experiment, we introduced the Pearson coefficient to calculate the correlation between the MMD values of the user original data and the MMD values of the user prototype data. The coefficient value of 0.969 reaffirms the effectiveness of SOINN in preserving data distribution.

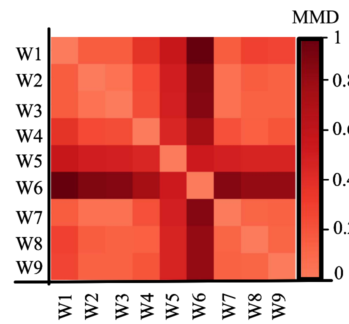


Fig. 6. Disparities in the distribution among nine sets of user raw data.

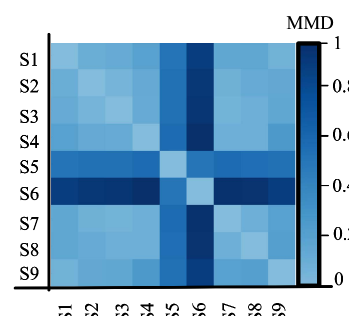


Fig. 7. Disparities in the distribution among nine sets of user prototype data.

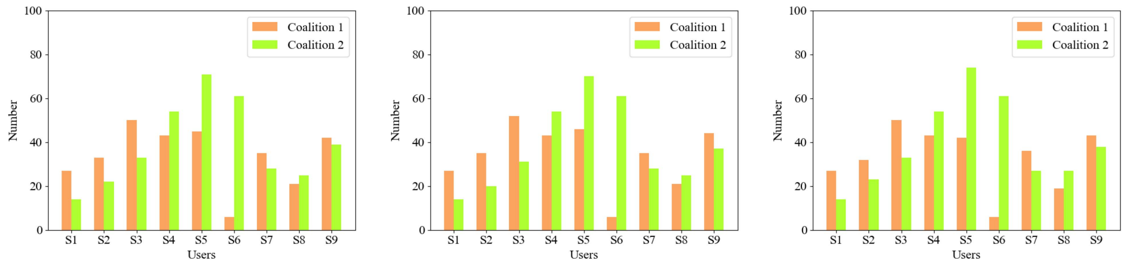


Fig. 8. Voting results of recommendation coalitions after clustering prototype samples using K-means, spectral clustering, and Gaussian mixture models.

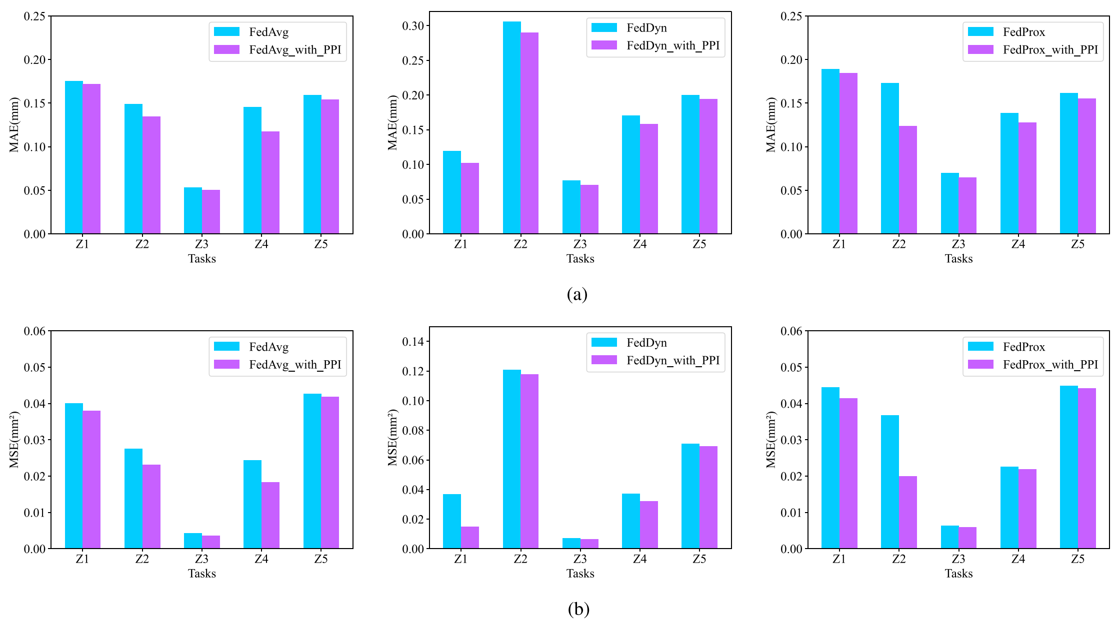


Fig. 9. Comparison of model performance with and without PPI mechanism. (a) Prediction performance of models calculated by MAE under different framework on the milling dataset. (b) Prediction performance of models calculated by MSE under different framework on the milling dataset.

D. Analysis of Voting Mechanism

We choose three classic clustering methods: K-means, spectral clustering, and Gaussian mixture clustering, and further explain the impact of constructing coalitions based on their clustering results. After conducting experiments on tool wear data, we determined the number of clusters to be 2 based on the Silhouette method. Fig. 8 illustrates the voting results of recommendation coalitions after clustering prototype samples using K-means, spectral clustering, and Gaussian mixture models. According to the voting results, industrial users 1, 2, 3, 7, and 9 form coalition 1, while industrial users 4, 5, 6, and 8 form coalition 2. It can be observed that the voting mechanism can smooth out the slight differences in clustering performance among different clustering methods.

E. Compared Schemes

In the experiment, regarding model selection under the FL framework, the tool wear mechanism [32] is introduced into the data-driven model

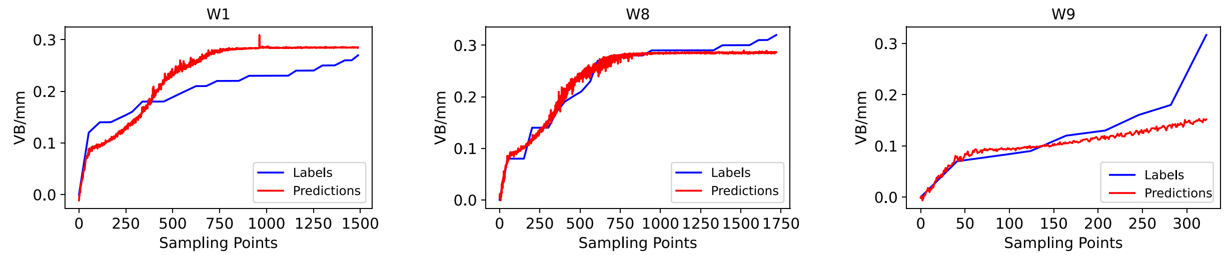
$$\frac{d(VB)}{dt} = c_1\left(1 + bt + 2c_2t^2\right)\exp\left(a + bt + c_2t^2\right) \tag{20}$$

where VB represents the maximum wear bandwidth of the tool back face and t represents the wear time. The constants a, b, c_1, and c_2 require optimization. Specifically, physics-informed neural network (PINN) [33] is introduced to fuse the data-driven model and mechanism model. All comparison schemes utilized the PINN model as the base model. The loss function in PINN consists of two parts: physic-guided loss and data-driven loss. The data-driven loss takes the average value of MSE loss between the predicted value of the model and the true tool wear value. The physic-guided loss takes the tool wear mechanism in (20) as the regularization term

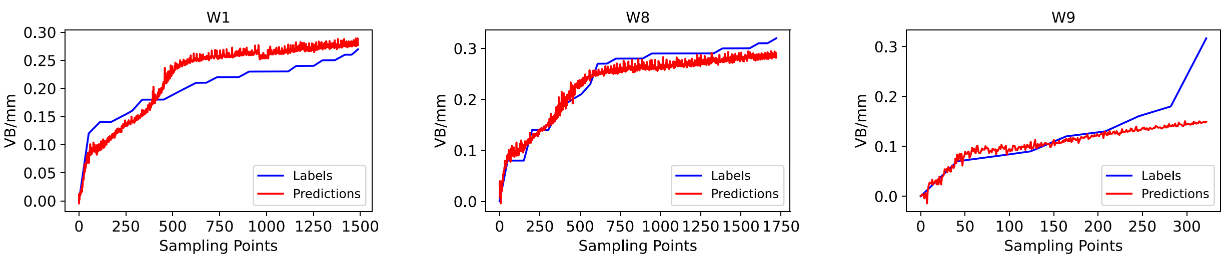
$$L_P = \left\| \frac{\partial \hat{y}}{\partial t} - \frac{d(VB)}{dt} \right\| \tag{21}$$

where $\frac{\partial \hat{y}}{\partial t}$ represents the partial differentiation of the tool wear value predicted by the model with respect to the wear time. The hyperparameters during model training are listed in Table IV.

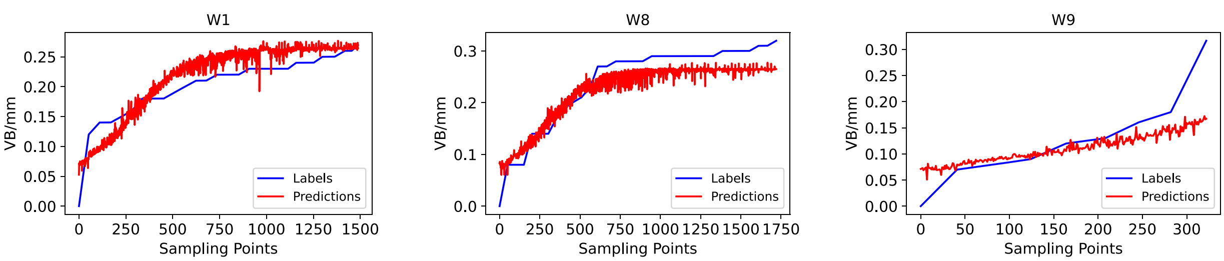
Three classic FL frameworks FederatedAveraging (FedAvg) [6], FedProx [34], FedDyn [35] were chosen for comparison. Specifically, FedProx is an FL framework that incorporates a proximal term to enhance model convergence and address



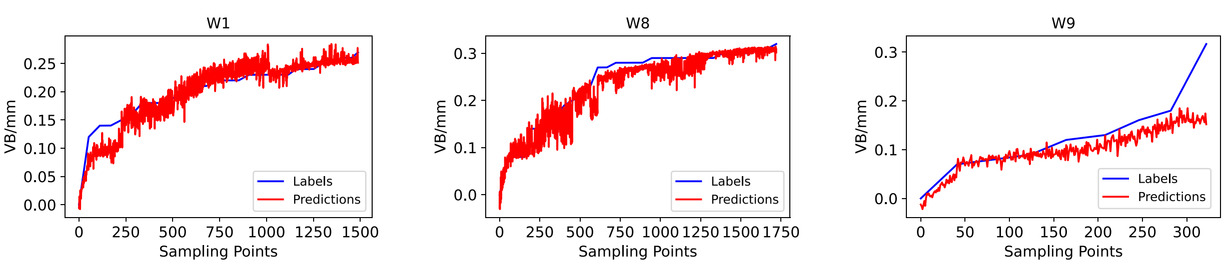
(a)



(b)



(c)



(d)

Fig. 10. Performance of models trained by different frameworks in tool wear prediction tasks. (a) Prediction performance of models trained under FedAvg on W1, W8, W9. (b) Prediction performance of models trained under FedProx on W1, W8, W9. (c) Prediction performance of models trained under FedDyn on W1, W8, W9. (d) Prediction performance of models trained under FLPPI on W1, W8, W9.

data heterogeneity among participating clients. FedDyn is an FL framework that dynamically adjusts communication rounds and local model updates to improve the convergence rate and robustness of federated models.

F. Experimental Results

Fig. 9 illustrates the predictive performance of the training model on the milling dataset using FedAvg, FedDyn, and FedProx, integrating the proposed potential partnership identification (PPI) mechanism. It is evident that, with the incorporation of the PPI mechanism, the models trained by the three classical FL frameworks exhibit reduced prediction errors. This improvement is attributed to the PPI mechanism filtering out inappropriate samples, thereby enhancing the overall accuracy of the model.

Table V presents the model performance achieved under different frameworks on the tool wear dataset. A comparison of the three classic frameworks in the table reveals that the model trained with FedAvg performs exceptionally well for W3, W6, and W8, while the model trained with FedProx excels in W1, W4, W7, and W9. In addition, the model trained with FedDyn demonstrates superior performance in W2 and W5. These results indicate that the predictive capabilities of models trained under the three frameworks are quite similar when it comes to tool wear regression tasks. Notably, our proposed framework outperforms the others in tool wear tasks for seven users, with its performance only falling behind FedAvg in tool wear prediction tasks for W6 and W8. This is because in our experiment, the heterogeneity of the data held by users is not significant, which leads to the performance of the model being affected by the number of modeling samples. After building coalitions under the proposed framework, the number of samples used for modeling decreases. In reality, the data heterogeneity caused by multiple industrial users processing products of different specifications will be greater than the data heterogeneity caused by slight changes in processing parameters, because the setting of processing parameters will be significantly different when processing products of different specifications. In our proposed scheme, the greater the heterogeneity of data, the more significant the effect of FL through building alliances will be. The proposed solution will have a significant improvement effect in real-world federal cooperation scenarios. In summary, the model trained under our proposed framework exhibits superior performance in predicting tool wear compared to the alternatives. Fig. 10 illustrates the specific prediction performance of different models on partial user data. It is evident that under the proposed framework, the fitting degree of the trained model is higher on the tool wear prediction task.

V. Conclusion

In this study, a federated collaboration framework based on potential partnership recognition is proposed to tackle the uncertainty of cooperation among industrial users participating in FL. The framework evaluates the similarity of data in the hands of users while ensuring data privacy, and recommends potential partners for users, so that users under each coalition can obtain better profits. Typical tool wear prediction experiments demonstrate that the model trained under the proposed framework delivers superior prediction performance.

There are still some challenging points to be studied. The proposed framework clusters all prototype samples to ascertain the number of coalitions. The construction of coalitions is highly correlated with the distribution of prototype samples. In reality, in order to improve cooperation efficiency and effectiveness, it is also necessary to consider the computing and communication resources owned by users when seeking for potential partner. We will further take into account the aforementioned factors to form more suitable coalitions by game theory [36]. In addition, when evaluating the quality of user data, the number of prototype samples is affected by the similarity threshold of SOINNs. How to adaptively set a unified similarity threshold is also a challenging task.

References

[1] J. Miao, J. Wang, and Q. Miao, "An enhanced multifeature fusion method for rotating component fault diagnosis in different working conditions," *IEEE Trans. Rel.*, vol. 70, no. 4, pp. 1611–1620, Dec. 2021.

[2] H. Zhang, M. Chen, X. Xi, and D. Zhou, "Remaining useful life prediction for degradation processes with long-range dependence," *IEEE Trans. Rel.*, vol. 66, no. 4, pp. 1368–1379, Dec. 2017.

[3] C. Wang, N. Lu, Y. Cheng, and B. Jiang, "A data-driven aero-engine degradation prognostic strategy," *IEEE Trans. Cybern.*, vol. 51, no. 3, pp. 1531–1541, Mar. 2021.

[4] Y. Li, M. Fei, L. Jia, N. Lu, O. Kaynak, and E. Zio, "Novel outlier-robust accelerated degradation testing model and lifetime analysis method considering time-stress-dependent factors," *IEEE Trans. Ind. Inform.*, vol. 20, no. 8, pp. 9907–9917, Apr. 2024.

[5] T. Deng, Y. Li, X. Liu, and L. Wang, "Federated learning-based collaborative manufacturing for complex parts," *J. Intell. Manuf.*, vol. 34, no. 7, pp. 3025–3038, 2023.

[6] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. Conf. Artif. Intell. Statist.*, PMLR, 2017, pp. 1273–1282.

[7] W. Zhang and X. Li, "Federated transfer learning for intelligent fault diagnostics using deep adversarial networks with data privacy," *IEEE/ASME Trans. Mechatron.*, vol. 27, no. 1, pp. 430–439, Feb. 2022.

[8] N. Qin, J. Du, Y. Zhang, D. Huang, and B. Wu, "Fault diagnosis of multirailway high-speed train bogies by improved federated learning," *IEEE Trans. Veh. Technol.*, vol. 72, no. 6, pp. 7184–7194, Jun. 2023.

[9] S. Lu, Z. Gao, P. Zhang, Q. Xu, T. Xie, and A. Zhang, "Event-triggered federated learning for fault diagnosis of offshore wind turbines with decentralized data," *IEEE Trans. Autom. Sci. Eng.*, vol. 21, no. 2, pp. 1271–1283, Apr. 2024.

[10] S. Chung and R. Al Kontar, "Federated condition monitoring signal prediction with improved generalization," *IEEE Trans. Rel.*, vol. 73, no. 1, pp. 438–450, Mar. 2024.

[11] S. Cui, J. Liang, W. Pan, K. Chen, C. Zhang, and F. Wang, "Collaboration equilibrium in federated learning," in *Proc. 28th ACM SIGKDD Conf. Knowl. Discov. Data Mining*, 2022, pp. 241–251.

[12] L. Fu, H. Zhang, G. Gao, M. Zhang, and X. Liu, "Client selection in federated learning: Principles, challenges, and opportunities," *IEEE Internet Things J.*, vol. 10, no. 24, pp. 21811–21819, Dec. 2023.

[13] N. Skatchkovsky and O. Simeone, "Optimizing pipelined computation and communication for latency-constrained edge learning," *IEEE Commun. Lett.*, vol. 23, no. 9, pp. 1542–1546, Sep. 2019.

[14] B. Lin, Y. Huang, J. Zhang, J. Hu, X. Chen, and J. Li, "Cost-driven off-loading for DNN-based applications over cloud, edge, and end devices," *IEEE Trans. Ind. Inform.*, vol. 16, no. 8, pp. 5456–5466, Aug. 2020.

[15] D. Rothchild et al., "Fetchsgd: Communication-efficient federated learning with sketching," in *Proc. Int. Conf. Mach. Learn.*, PMLR, 2020, pp. 8253–8265.

[16] Z. Liu, J. Guo, W. Yang, J. Fan, K.-Y. Lam, and J. Zhao, "Privacy-preserving aggregation in federated learning: A survey," *IEEE Trans. Big Data*, to be published, doi: 10.1109/TBDATA.2022.3190835.

[17] F. Sattler, K.-R. Müller, and W. Samek, "Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 8, pp. 3710–3722, Aug. 2021.

[18] Y. Kim, E.Al Hakim, J. Haraldson, H. Eriksson, J. M. B. da Silva, and C. Fischione, "Dynamic clustering in federated learning," in *Proc. IEEE Int. Conf. Commun.*, 2021, pp. 1–6.

[19] A. Z. Tan, H. Yu, L. Cui, and Q. Yang, "Towards personalized federated learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 12, pp. 9587–9603, Dec. 2023.

[20] Y. Yan, X. Tong, and S. Wang, "Clustered federated learning in heterogeneous environment," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: 10.1109/TNNLS.2023.3264740.

[21] Z. Yao, P. Song, and C. Zhao, "Finding trustworthy neighbors: Graph aided federated learning for few-shot industrial fault diagnosis with data heterogeneity," *J. Process Control*, vol. 129, 2023, Art. no. 103038.

[22] S. Furao and O. Hasegawa, "An incremental network for on-line unsupervised classification and topology learning," *Neural Netw.*, vol. 19, no. 1, pp. 90–106, 2006.

[23] T. Song, Y. Tong, and S. Wei, "Profit allocation for federated learning," in *Proc. IEEE Int. Conf. Big Data*, 2019, pp. 2577–2586.

[24] J. Kang, Z. Xiong, D. Niyato, S. Xie, and J. Zhang, "Incentive mechanism for reliable federated learning: A joint optimization approach to combining reputation and contract theory," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 10700–10714, Dec. 2019.

[25] L. U. Khan et al., "Federated learning for edge networks: Resource optimization and incentive mechanism," *IEEE Commun. Mag.*, vol. 58, no. 10, pp. 88–93, Oct. 2020.

[26] K. Donahue and J. Kleinberg, "Model-sharing games: Analyzing federated learning under voluntary participation," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 5303–5311.

[27] Y. Wu, D. Liu, X. Yuan, and Y. Wang, "A just-in-time fine-tuning framework for deep learning of SAE in adaptive data-driven modeling of time-varying industrial processes," *IEEE Sensors J.*, vol. 21, no. 3, pp. 3497–3505, Feb. 2021.

[28] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, 1987.

[29] A. Agogino and K. Goebel, "BEST Lab, UC Berkeley Milling Data Set, NASA Prognostics Data Repository," NASA Ames Research Center, Moffett Field, CA, 2007. [Online]. Available: https://www.nasa.gov/intelligent-systems-division/discovery-and-systems-health/pcoe/pcoe-data-set-repository/

[30] C. Liu, Y. Li, J. Li, and J. Hua, "A meta-invariant feature space method for accurate tool wear prediction under cross conditions," *IEEE Trans. Ind. Inform.*, vol. 18, no. 2, pp. 922–931, Feb. 2022.

[31] A. Gretton, K. Borgwardt, M. Rasch, B. Schölkopf, and A. Smola, "A kernel method for the two-sample-problem," in *Proc. Adv. Neural Inf. Process. Syst.*, 2006, pp. 513–520.

[32] H. Hanachi, W. Yu, I. Y. Kim, J. Liu, and C. K. Mechefske, "Hybrid data-driven physics-based model fusion framework for tool wear prediction," *Int. J. Adv. Manuf. Technol.*, vol. 101, pp. 2861–2872, 2019.

[33] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *J. Comput. Phys.*, vol. 378, pp. 686–707, 2019.

[34] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," *Proc. Mach. Learn. Syst.*, vol. 2, pp. 429–450, 2020.

[35] D. A. E. Acar, Y. Zhao, R. M. Navarro, M. Mattina, P. N. Whatmough, and V. Saligrama, "Federated learning based on dynamic regularization," 2021, *arXiv:2111.04263*.

[36] Y. Xu, H. Yang, B. Jiang, and M. M. Polycarpou, "Distributed optimal fault estimation and fault-tolerant control for interconnected systems: A stackelberg differential graphical game approach," *IEEE Trans. Autom. Control*, vol. 67, no. 2, pp. 926–933, Feb. 2022.

Chao Ren received the M.S. degree in control engineering from the College of Electrical and Power Engineering, Taiyuan University of Technology, Taiyuan, China, in 2020. He is currently working toward the Ph.D. degree in control theory and control engineering with the Nanjing University of Aeronautics and Astronautics, Nanjing, China.

His current research interests include industrial artificial intelligence, industrial fault diagnosis, and industrial process monitoring.

Bin Jiang (Fellow, IEEE) received the Ph.D. degree in automatic control from Northeastern University, Shenyang, China, in 1995.

He was a Postdoctoral Fellow, a Research Fellow, an Invited Professor, and a Visiting Professor in Singapore, France, USA, and Canada, respectively. He is currently a Chair Professor of the Cheung Kong Scholar Program with the Ministry of Education and the President of the Nanjing University of Aeronautics and Astronautics, Nanjing, China. He has authored eight books and more than 100 referred international journal articles. His current research interests include intelligent fault diagnosis and fault tolerant control and their applications to helicopters, satellites, and high-speed trains.

Dr. Jiang is a fellow of the Chinese Association of Automation (CAA). He was a recipient of the Second-Class Prize of National Natural Science Award of China. He currently an Editor for *International Journal of Control, Automation and Systems*, an Associate Editor or an Editorial Board Member for a number of journals, such as IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE TRANSACTIONS ON CYBERNETICS, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, and *Journal of the Franklin Institute*. He is also a Chair of Control Systems Chapter in IEEE Nanjing Section, and a member of the IFAC Technical Committee on Fault Detection, Supervision, and Safety of Technical Processes.

Ningyun Lu (Member, IEEE) received the Ph.D. degree in detection technology and automatic equipment from Northeastern University, Shenyang, China, in 2004.

From 2002 to 2005, she was a Research Associate and a Postdoctoral Fellow with the Hong Kong University of Science and Technology, Hong Kong, China. She is currently a Full Professor with the College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China. Her research interests include data-driven fault prognosis and diagnosis and their applications to various industrial processes.