Future Generation Computer Systems 135 (2022) 244–258

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

A state-of-the-art survey on solving non-IID data in Federated Learning

Xiaodong Ma [a], Jia Zhu [a,*], Zhihao Lin [b], Shanxuan Chen [b], Yangjie Qin [b]

[a] The Key Laboratory of Intelligent Education Technology and Application of Zhejiang Province, Zhejiang Normal University, Jinhua, Zhejiang 321004, China
[b] School of Computer Science, South China Normal University (Shipai Campus), Guangzhou, Guangdong 510631, China

ARTICLE INFO

Article history:
Received 2 December 2020
Received in revised form 29 April 2022
Accepted 1 May 2022
Available online 7 May 2022

Keywords:
Federated Learning
Non-IID data
Machine learning
Statistical heterogeneity

ABSTRACT

Federated Learning (FL) proposed in recent years has received significant attention from researchers in that it can enable multiple clients to cooperatively train global models without revealing private data. This training mode protects users' privacy without violating the supervision, and aggregates scattered data to exert great potential. However, the data samples on each participating device of FL are usually not independent and identically distributed (IID), which leads to serious statistical heterogeneity challenges for FL. In this article, we analyze and establish the definition of non-IID data problems, and put forward a series of challenges that this problem may bring to FL. We classify existing methods to solve this problem from the researcher's entry point and subsequent sub-methods, aiming to provide a comprehensive study for solving the problem of non-IID data in FL. Our research shows that non-IID data will not only reduce the performance of the FL model, but also damage the active participation of users in the FL process. Compared with methods based on data-side sharing, enhancement, and selection, it is more common for researchers to improve federated learning algorithms from models, algorithms, and frameworks to solve non-IID problems. To the best of our knowledge, although many efforts have been made to address the problem of non-IID data, there are currently few authoritative systematic reviews in this field and are not up-to-date. In this article, we will fill the gaps in FL and provide researchers with the state-of-the-art research results to solve non-IID problems in FL and promote the further implementation of FL.

1. Introduction

The successful application of machine learning (ML) in scientific research and business decision-making benefits from the driving force of data. Data with high representation ability can help us build more complex and accurate ML models. We need data. However, there are contradictions in the current data. On the one hand, the data is explosive. People generate a large amount of data every day through ubiquitous intelligent devices, such as photos, language, behavior, and so on. On the other hand, data is hungry. Traditional ML relies on the main server to store data and train models. Due to the current data is distributed to various mobile terminals, it is difficult to aggregate and store the data. With the emergence of the General Data Protection Regulation (GDPR) [1], the challenge of data aggregation becomes greater [2,3]. In this scenario, scholars began to turn their attention from data aggregation to model aggregation. Federated learning (FL) is emerging as a new distributed machine learning framework.

Federated learning (FL), first proposed by McMahan [4], is a framework that enables multiple users known as clients to collaboratively train a shared global model without exposing the data from their local devices [5]. A typical architecture and training process of an FL system is shown in Fig. 1. A central server coordinates the FL process composed of multiple rounds. At the beginning of each round, the central server sends the initialized global model to the participating clients. Then, each client trains the model on its local data and communicates only the model updates back to the central server. Last, the central server collects these updates from all clients and update the global model to end this round of the FL training [6–8]. As a result, the user data will never be directly shared with the third party in the whole FL training process. FL can not only ensure that training data is only kept on personal devices, but also promote collaborative machine learning of complex models among distributed devices.

In recent years, FL has attracted much attention and has been successfully applied to many applications, such as the prediction of the next words [9,10] and the detection of visual objects [11]. Compared with the traditional ML, FL can make more effective

* Corresponding author.
E-mail addresses: felicity@m.scnu.edu.cn (X. Ma), jiazhu@zjnu.edu.cn (J. Zhu), zhihaolin@m.scnu.edu.cn (Z. Lin), shanxuanchen@m.scnu.edu.cn (S. Chen), eurusyoung@m.scnu.edu.cn (Y. Qin).

https://doi.org/10.1016/j.future.2022.05.003
0167-739X/

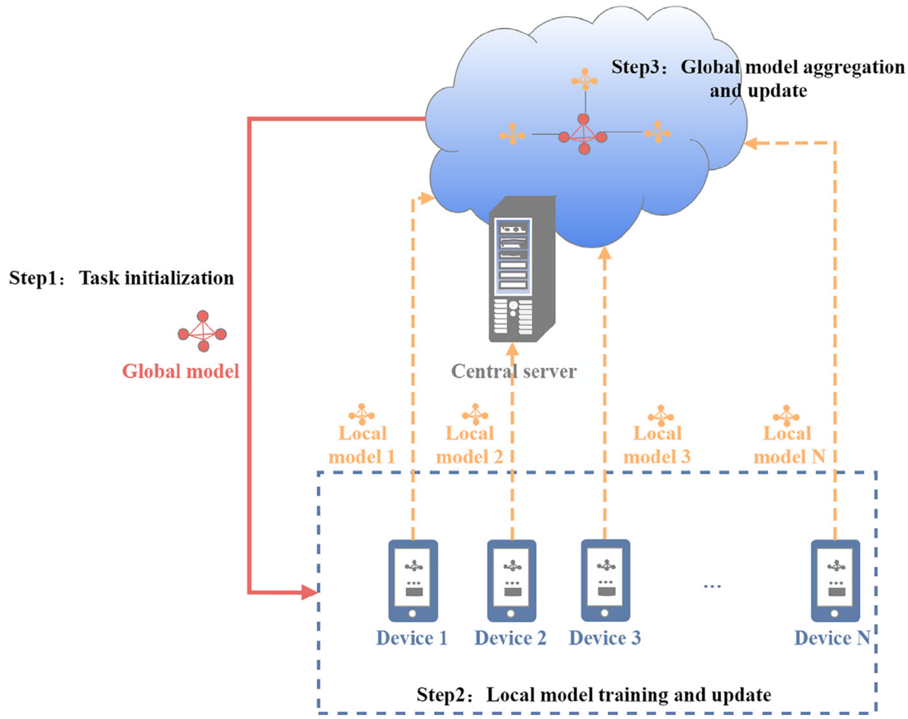


Fig. 1. The Federated learning process: (A) Task initialization by the central server. (B) The devices train the model further on their local datasets and update. (C) The central server performs global model aggregation and updates.

use of network bandwidth, protect data privacy, and train a high-quality global model [12–14]. However, FL has many or even millions of client nodes, and the data distribution of each node is seriously different. At the same time, there is high communication latency and instability between the client and the central server. These factors make FL face many challenges when facing practical application requirements.

In this paper, we will focus on the heterogeneity of data. Usually, the heterogeneity of data is caused by the non-IID distribution of data. For example, a hospital specializes in certain medical diseases, so there will be a lot of medical data related to this disease, on the contrary, there will be very few medical data on other diseases. Most FL research work assumes that data is IID on distributed edge devices, which is what we want rather than the real-world data distribution. In order to solve the problem of non-IID data in FL, many scholars have been made many efforts. For example, McMahan [9] mentioned non-IID data as the main challenge of FL in their pioneering papers, and mentioned that the FedAvg algorithm can present ideal results on non-IID data. Subsequent scholars have improved FedAvg, which makes FL show more efficient communication and better performance on non-IID data [15–17].

Table 1 lists the relevant review papers of FL. This paper selects the 15 most cited review articles. Among them, most of the reviewed literatures involve the basic introduction of federated learning in concepts, methods, and applications. The literature review covers data privacy and security, communication strategy and efficiency, personalization, Internet of Things, etc. To our knowledge, there are currently few systematic literature reviews on non-IID data in FL, e.g. there is only one highly cited article reviewing non-IID problems in federated learning in Table 1. However, the discussion of [18] on solving the non-IID problem in federated learning is not comprehensive and time-sensitive. Therefore, the goal of this paper is to give a comprehensive and up-to-date literature review on the research of non-IID data in FL. The contributions of this paper are as follow:

- We cover the research published from 2016 to present (April 2022) on solving the problem of non-IID data in FL.
- We have made an in-depth summary of the existing research on non-IID data of FL, which reflects the latest research level.
- According to the summary results, we provide the future trend to solve the problem of non-IID distribution of data in FL to support further research in related fields.

This paper is organized as follows. Section 2 introduces the problem of non-IID data in FL and summarizes the many challenges it brings. Section 3 systematically discusses the research status of FL in solving non-IID data distribution problems from the four dimensions of data, model, algorithm and framework. Furthermore, datasets and data heterogeneity settings commonly used in existing work are summarized. Then Section 4 gives the future research trends in this field. Finally, Section 5 ends with the conclusions.

2. The non-IID data in federated learning

The problem of non-IID data distribution can also become a problem of data heterogeneity, which exists in many ML applications and distributed learning methods. All along, ML algorithms are usually trained based on the assumption that data are independent and identically distributed (IID) [32]. Due to the difference of non-IID data in data quantity and category distribution, many distributed training of deep neural networks suffer from serious precision loss. Hsieh [33] further indicated that the difficulty of solving the problem of non-IID data depends on its deviation degree. As a distributed deep learning algorithm, FL can train global models without sharing user data, which has attracted much attention in recent years. However, it is unrealistic to ensure that the data in the edge device is IID in the practical application environment of FL. Therefore, the distribution of non-IID data has become one of the challenges

Table 1
Overview of FL related research fields (top 15 cited)

Ref	Major contributions	Year
[19]	An introduction to the definition, architecture, and applications of federated learning frameworks, and provides a comprehensive survey of existing work	2019
[20]	An introduction to statistical challenges, system challenges, and privacy issues in federated learning, and a comprehensive survey of existing work in healthcare informatics	2021
[6]	An introduction to the concepts, techniques, protocols, and applications of federated learning, and discuss their challenges and benefits Provide detailed service use-cases	2020
[21]	An introduction to the need for personalized federated learning and provides a comprehensive review of existing personalized federated learning	2020
[22]	An introduction to the significance, technical challenges and future directions of applying federated learning in in-vehicle IoT, and conduct a comprehensive review of existing work	2020
[23]	An introduction to the federated learning system model, design, application areas, privacy security and resource management, and discuss important challenges and future directions of federated learning	2020
[24]	An introduction to the data partition, privacy mechanism, machine learning model, communication architecture and system heterogeneity of federated learning, and sort out the current challenges and future directions	2021
[25]	An introduction to the concepts, techniques, and learning methods of federated learning frameworks, and provide a comprehensive review of existing work in communication and networking	2021
[26]	An introduction to the current challenges in smart city sensing and provides a comprehensive review of existing federated learning work in smart city sensing	2020
[27]	An introduction to the protection mechanism and attack strategy of federated learning privacy and security, and conduct a comprehensive review of the existing work	2021
[28]	An introduction to the concepts of federated learning and provide a comprehensive review of work in federated learning and neural architecture search	2021
[29]	An introduction to the problems and challenges of federated learning applied to the IoT environment, and provide a comprehensive review of the work to address them	2021
[30]	An introduction to the issues and challenges of privacy-preserving federated learning, and provide a comprehensive review of the work to address them	2021
[18]	An introduction to non-IID data distribution in federated learning, discuss its challenges and implications, and provide a comprehensive review of existing work to address them	2021
[31]	An introduction to the vulnerabilities, threats, and challenges in federated learning protocols, and provide a comprehensive review of existing attack and defense methods	2021

affecting the development of FL, which has attracted more and more attention.

Assume that the respective local data distributions of client i and client j participating in federated learning are P_i and P_j. Then in federated learning, data heterogeneity can be expressed as the difference between P_i and P_j for different clients i and j. Therefore, when performing a supervised task with feature x and label y in federated learning, a participating client is randomly selected, and then the local data distribution of selected client i $P_i(x, y)$ extract feature label pairs from (x, y). Then the data heterogeneity in federated learning can be subdivided into feature distribution skew, label distribution skew, same label different features, same feature different labels, and quantity skew. In order to further subdivide and better observe the form of data heterogeneity difference, this paper writes $P_i(x, y)$ as $P_i(y \mid x)P_i(x)$ and $P_i(x \mid y)P_i(y)$. This paper breaks down the types of data heterogeneity in federated learning as follows:

- **Feature distribution skew:** Feature distribution skew is when the distribution of the feature $P_i(x)$ varies from customer to customer, but the distribution of $P_i(y \mid x)$ is the same. For example, on the MNIST dataset, customer i likes and often writes the number 5 in thin font, while customer j likes and often writes the number 5 in bold. In the data distribution for customer j, the number 5 has a higher probability of being written in bold. But in the data distribution for client i, the number 5 is more likely to be written in thin font, i.e. $P_i(x)$ is different. However, the fonts are all 5, i.e. $P_i(y \mid x)$ is the same.
- **Label distribution skew:** Label distribution skew means that the label $P_i(y)$ distribution of different clients is different, but the $P_i(x \mid y)$ situation is the same. For example, on the MNIST dataset, in the distribution of client i, 90% of the digits are 7 and 10% of the digits are other digits. In client j's distribution, 95% of the numbers are 7 and 5% of the numbers are other numbers. That is, the distribution of $P_i(y)$ is different. But if $y = 7$ is given, then the corresponding feature x has roughly the same probability of being 7. That is, the $P_i(x \mid y)$ distributions are the same.
- **Same label, different features:** Different features with the same label means that the distribution of $P_i(x \mid y)$ is different for different clients, and the distribution of $P_i(y)$ is the same. For example, the same buildings, Eastern countries and Western countries have different architectural styles. That is, the building style $P_i(x \mid y)$ of the client i belonging to the eastern country is different from the building style $P_i(x \mid y)$ of the client j belonging to the western country. But they are all buildings, that is, $P_i(y)$ is the same.
- **Same features, different label:** Different labels with the same feature means that the distribution of $P_i(y \mid x)$ is different for different clients, and the distribution of $P_i(x)$ is the same. For example, in the same living environment, people's attitudes towards life are different. Jack advocates optimism (feature), Myron is the opposite. That is, the $P_i(y \mid x)$ distribution of different clients is different. But they are all attitudes, that is, $P_i(x)$ is the same.
- **Quantity skew:** Quantity skew refers to the large difference in the quantity of different client data $P_i(x, y)$. For example,

client i has 100 data samples, and client j has 10,000,000 data samples. That is, the data amount of $P_i(x, y)$ is very different.

To understand the challenges and troubles brought by non-IID data distribution to FL, it is necessary to understand the Stochastic gradient descent (SGD) algorithm [34]. At present, the successful application of a large number of deep learning depends almost entirely on SGD variables for optimization. Compared with batch gradient descent (BGD), SGD updates the gradient of each sample (random sampling) every time. Therefore, the SGD algorithm has a faster update speed and is easier to converge to the local minimum. Because the batch gradient is required to be calculated once in each round of communication, SGD can be naturally applied to the federated optimization problem.

As early as 2016, McMahan [9] proposed to use the FedAvg algorithm of SGD to solve the global aggregation problem of FL. At that time, the FedAvg algorithm attracted a lot of attention. Researchers have analyzed and discussed the theoretical and experimental results [15,17,35]. In McMahan's pioneering paper, the problem of non-IID data is regarded as one of the major challenges of FL. The article points out that the FedAvg algorithm can still be robust and effective in non-IID data problems. However, some subsequent researchers have analyzed and discussed the performance of the FedAvg model in a non-IID data environment. The results show that the heterogeneity of data will significantly reduce the convergence speed and accuracy of the model, especially for highly skewed non-IID data sets. At the same time, experimental results show that the FedAvg algorithm is unstable or even divergent on non-IID data [36].

Fig. 2 is a statistical graph of the challenges faced by FL drawn by [37], where the numbers represent the literature statistics of related challenges. That is, [37] statistically found that there are 33 papers that identified data heterogeneity as the main challenge for federated learning. Therefore, we can see that among the many challenges faced by FL, data heterogeneity is the second biggest challenge after communication efficiency. By analyzing and summarizing the existing research results dedicated to solving the non-IID problem in FL, we can find that the client data distribution is very important to the performance of the FL model. When the data distribution of each client deviates seriously, the FL model is difficult to learn and its performance is low. Such as, non-IID data leads to many FL model performance problems, such as lower accuracy, delay of model communication, slow convergence of model, and so on [36,38,39].

In addition, under heterogeneous data conditions, a single FL model cannot provide the best performance for all clients at the same time. When there is a big gap between the global data distribution and the local data distribution, the accuracy of the model obtained by FL training is lower than that of the participants' local training model, which leads to the participants not benefiting from the FL process and the low practicability of the model [40–42].

After understanding the challenges brought by non-IID data to FL, the next chapter of this paper will focus on the existing literature used to solve the problem of non-IID data in FL.

3. Classification techniques/approaches to solve the problem of non-IID data in FL

As mentioned above, researchers gradually pay attention to solving the problem of non-IID data in FL. In this paper, we have sorted out all the work aimed at solving the non-IID problem in FL from 2016 to present (April 2022). Fig. 3 shows a taxonomy of existing methods to address this problem and common datasets and data settings. This paper summarizes the existing methods to solve the non-IID problem of FL from four dimensions: data-based, model-based, algorithm-based and framework-based. We will discuss these methods in detail in this section.

3.1. Data-based

In general, the solutions to the non-IID problem in FL from a data-based dimension mainly include data sharing, data enhancement and data selection.

3.1.1. Data sharing

Zhao [36] pointed out that the poor performance of the FedAvg algorithm on non-IID data is due to the weight difference. To solve this problem, this paper proposes to train a warm-up model by using shared data in the cloud server and then share a subset of shared datasets from the cloud server to each client, thus reducing the weight difference between the cloud server and each device and reducing the accuracy loss caused by non-IID data. As shown in Fig. 4, the global server distributes the warm-up model trained on the global shared dataset and the random α part of the global shared dataset to each client. The advantage of this scheme is that the cloud training warm-up model occurs in the initialization stage of FL, which will not produce excessive communication costs.

Unlike the former, Itahara [43] share unlabeled data to protect privacy. In [43], the device is set not only to store private data of labels but also to store unmarked data that can be shared. Through the output of local models exchanged between devices, unlabeled data are labeled, and then each local model is further trained based on these labels. In this way, the effect of data enhancement is achieved, thus improving the performance of the model. In addition, Collins [44] is devoted to learning shared data representations across clients, computing personalized low-dimensional classifiers for each client to cope with non-IID data distributions. Li [45] proposes a private shared model approach with inheritance to retain and transfer personalized knowledge of the global model with historical value. Tian [46] proposes a pre-shared data scheme to alleviate the divergence of model convergence caused by non-IID.

As a conclusion, the solutions based on data sharing are becoming more and more diverse. They act on the data difference itself and can fundamentally and quickly solve the problem itself. For example, [36] experiment shows that sharing only 5% global data on the CIFAR-10 datasets can improve the test accuracy by 30%. However, the operation of sharing data has the potential to undermine user data privacy, which is contrary to the original intention of FL. At the same time, in the algorithm based on data sharing, it is difficult to keep a grasp of which part of the data to share and the proportion of the shared data, and it is also difficult to operate under the FL of privacy protection.

3.1.2. Data enhancement

To alleviate the damage of non-IID data to FL model performance, Jeong [16] proposed a FAug algorithm to correct non-IID data, as shown in Fig. 5. Referring to the concept of [36], the FAug algorithm aims to allow devices to obtain missing data sample labels (target labels) from a small number of data samples uploaded by other devices to balance non-IID data samples. In FAug algorithm, the server samples the uploaded small data samples, and then trains them to form a GAN generator to supplement target tags for each device. Finally, the trained GAN model is distributed to each participant, and each participant completes data expansion by using their own data. It is worth mentioning that in order to ensure privacy, the target labels uploaded by users are doped with noise data. Experiments show that adding the FAug algorithm to FL can achieve higher accuracy and lower communication cost than not adding the FAug algorithm.

Unlike the former, Shin [47] proposed XorMixup hybrid data enhancement technology for one-shot FL. This technology uses XOR encoding and decoding to process the data, so as to generate

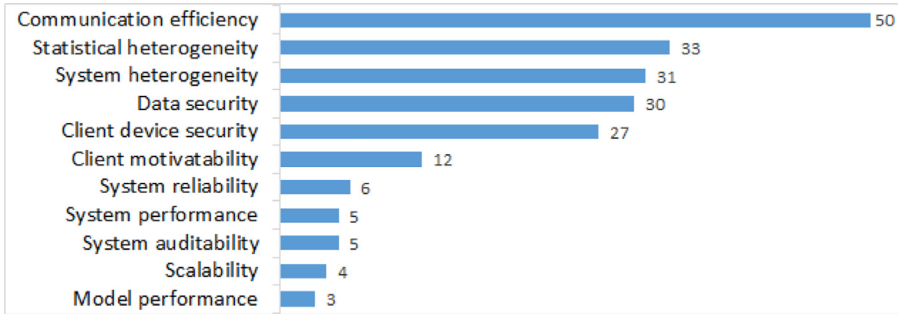


Fig. 2. Challenges in FL [37].

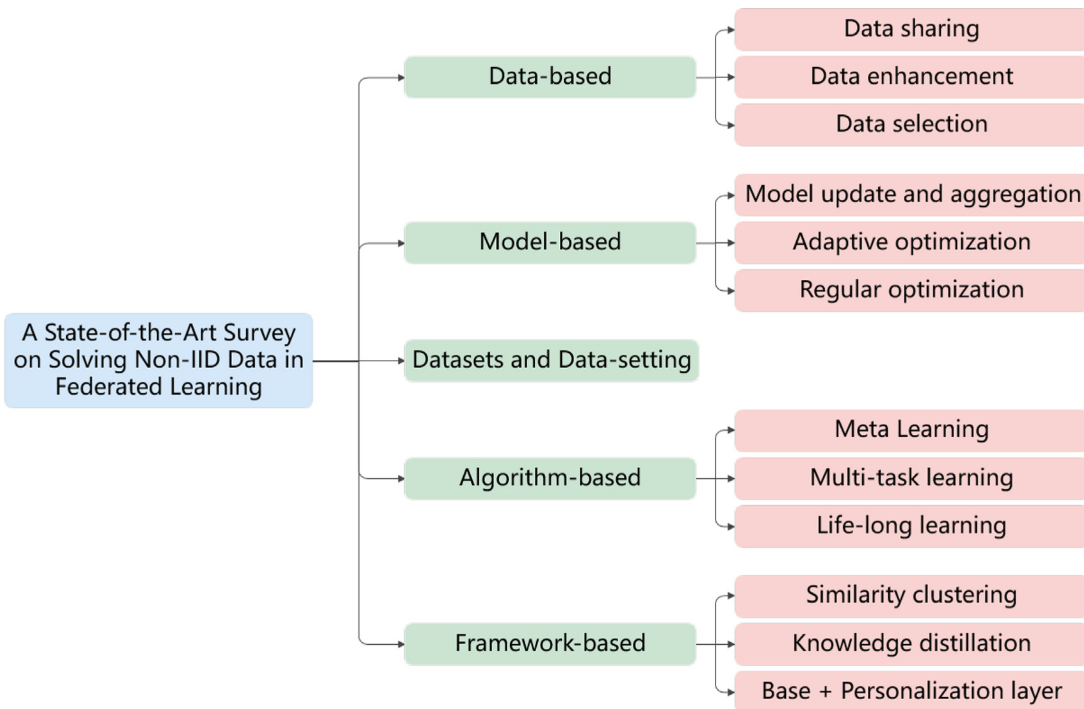


Fig. 3. Classification of main methods to solve the problem of non-IID data in FL.

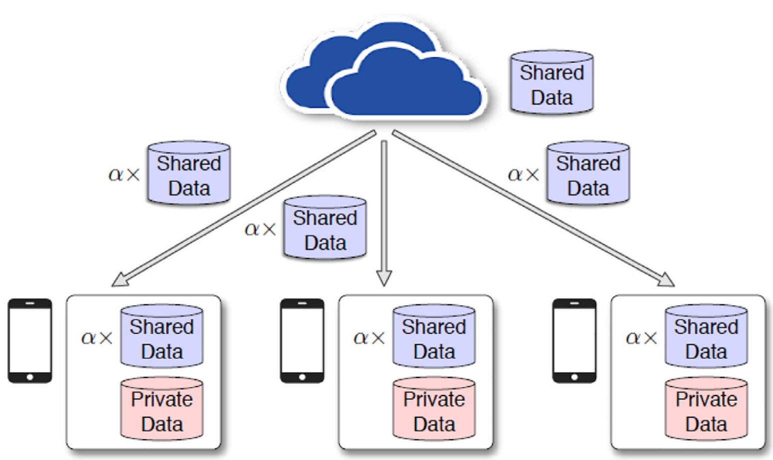


Fig. 4. Illustration of the data-sharing strategy [36].

synthetic but real sample data on the server to correct its imbalanced training dataset. In addition, in order to avoid excessive noise, Shin [47] mixes multiple samples to extract common features in the sample blending step. Given that heterogeneous data can lead to federated learning that produces a biased global model, Hao [48] proposes a fair federated learning approach that uses zero-sample data augmentation on underrepresented data to weaken the effect of statistical heterogeneity.

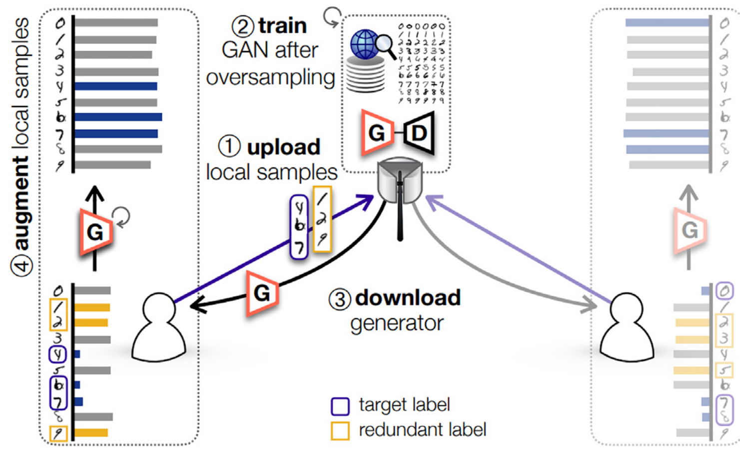


Fig. 5. FAug with 3 target and 3 redundant MNIST labels [16].

As a conclusion, data enhancement is similar to data sharing, which starts with the data source to solve the non-IID problem in FL. The difference is that the data enhancement scheme only uploads a small number of labels or samples by the client, and controls the privacy of data by doping noise or encoding–decoding. Generally speaking, the data enhancement scheme will play a greater potential in solving the problem of non-IID data in FL from the data source without destroying the data privacy.

3.1.3. Data selection

In order to deal with local data imbalance and non-IID, [49] proposes a dynamic sample selection optimization algorithm. The algorithm dynamically selects the training sample size according to the locally available data size during gradient iteration. For the clients under the non-IID data distribution, the researchers select the best clients to participate in the training by first classifying and then selecting. For example, Wang [50] assigns different workloads to mobile devices based on the same and different data distributions. Such as when the data is equally distributed, combine partitioning and linear bottleneck allocation to deal with it. When the data are distributed differently, it is transformed into an average cost minimization problem and a greedy algorithm is proposed to balance it. Wang [51] proposes an intelligent selection mechanism based on deep q-learning, which selects a subset of devices to participate in training in each communication round. Xu [52] uses model segmentation techniques to classify data distribution, and then determines the coefficient relationship between data distribution and model training quality to select important participating devices that are conducive to global model convergence.

Differences in data distribution and uncertainty in data quality bring challenges to federated learning training for wireless edge networks. Wang [53] proposes FedACS, an analysis-driven client selection framework. FedACS quantifies the class distribution heterogeneity in the federated learning environment based on Hoeffding inequality, and then selects customers with small data heterogeneity based on Thompson sampling by defining the customer selection problem to become the ideal customer pool. Similarly, [54] proposes a scheduling algorithm based on data quality, namely data-quality based scheduling algorithm based on greedy knapsack. The algorithm prioritizes reliable devices with rich and diverse datasets, where device scores are established through a priority metric. In addition, Zeng [55] proposed a deep learning method to predict the training time of a model on heterogeneous clients, that is, designing a neural network to extract the impact of different model features on the training time. Dimensionality reduction rules are then applied to extract key features with greater impact and use the key features to train and predict a global model.

As a conclusion, the solution based on data selection does not change the data and the client itself, and is more tolerant to practical applications. At the same time, the data selection scheme can select the best quality client to participate in each round of federated learning training process, so that the communication and efficient training can obtain a global model with high accuracy. However, data selection will speed up the model shift problem caused by data heterogeneity and exacerbate the unfairness of federated learning.

3.2. Model-based

In general, the solutions to the non-IID problem in FL from a model-based dimension mainly include Model update and aggregation, Adaptive optimization and Regular optimization.

3.2.1. Model update and aggregation

Non-IID data distribution undoubtedly hinders the optimization convergence and generalization ability of federated learning models. Due to the heterogeneity of the data, each client has the phenomenon of wandering, which leads to the deviation and divergence of the model. To alleviate the challenges posed by data heterogeneity, researchers improve from model update and aggregation. For example, Sannara [56] proposes to modify the model architecture of its deep neural network by identifying differences between customer-specific neurons to achieve model aggregated updates. Kwatra [57] proposed a k-anonymous federated learning framework with decision trees. The framework allows each device to train a decision tree classifier, and the aggregator merges trees for model aggregation by selecting the most common split attributes. In addition, Qin [58] designed a multi-local multi-global model aggregation update method. As shown in Fig. 6, [58] solves the federated optimization problem between multiple local and global models by introducing an optimal matching algorithm, and uses a clustering method to train non-IID data.

In addition, Abay [59] adjusts the way of local reweighting, so that each party calculates reweighting locally based on its own training dataset during preprocessing, and then uses the reweighting dataset for local training. Similarly, [60] proposes a new stochastic control averaging algorithm to correct client drift. Chai [61] proposed a novel federated learning system with an asynchronous hierarchy under unspecified training data. The system connects synchronous and asynchronous

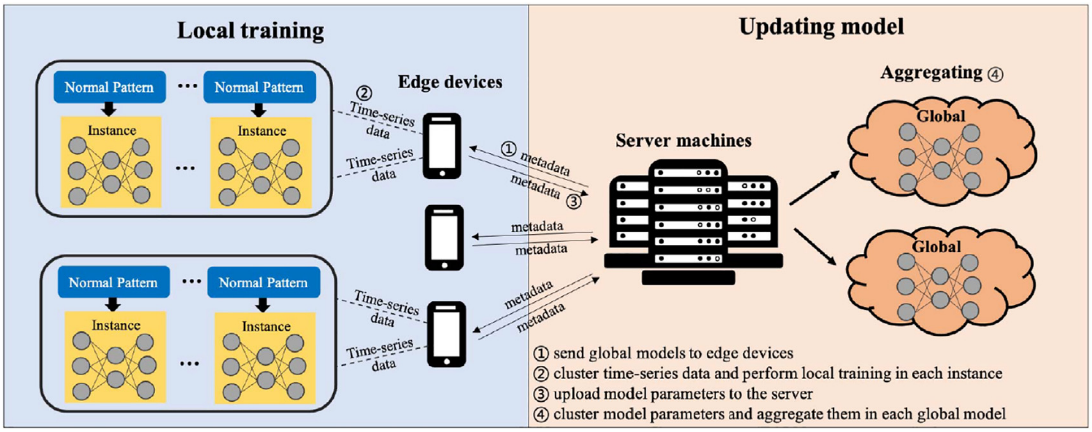


Fig. 6. Framework of a multi-local and multi-global federated learning system for anomaly detection [16].

training hierarchically, and solves the problem of falling behind through weighted aggregation heuristics to guide and balance training among clients. Similarly, Ma [62] proposes a semi-asynchronous federated learning mechanism. This mechanism allows each round of servers to aggregate a certain number of local models in the order they arrive, and adaptively adjust the learning rate based on the relative frequency with which workers participate in global updates. Chen [63] introduces an online learning process with a decay coefficient to balance the client's previous and current gradients, and introduces a dynamic learning strategy to adapt the local device to the gradient update step size to alleviate the problem of data heterogeneity. Furthermore, Tan [64] changes the communication between client and server from gradient interaction to abstract class prototype interaction, i.e. different clients collect local prototype summaries and update the global model.

As a conclusion, the strategy based on model update and aggregation acts on the update and aggregation method of the federated learning objective function. This method aims to alleviate the phenomenon of stragglers caused by the problem of data heterogeneity, and balance the data heterogeneity of each client by improving the update and aggregation methods.

3.2.2. Adaptive optimization

In the previous work, some researchers have proved that an adaptive coordination update is helpful to improve the convergence of FL [65]. At the same time, some studies have proved that the adaptive method has special advantages in the environment with heavy-tailed random gradient noise distribution, which is a common feature of non-IID data [66]. Therefore, Reddi [67] first studied the adaptive server optimizer with client learning rate attenuation. In [67], the server and client optimizers are used to propose ADAGRAD, ADAM, and YOGI optimization schemes based on FedAvg. ADAGRAD, ADAM, and YOGI pseudocodes based on FedAvg are provided in Algorithm 1. Among them, τ is used to control the self-adaptive size. The smaller τ, the higher the self-adaptation. η_l represents the client learning rate. Adjusting η_l can obtain the average difference Δ_t and accumulator v_t of the model under different optimization schemes. Experiments show that an adaptive optimizer can significantly improve the model performance of FL under heterogeneous data. Similarly, [68] uses Nesterov Accelerated Grade (NAG) to replace Adam and YOGI's classical momentum. The improved adaptive server optimizer overcomes the algorithmic challenges of fitting non-IID data, helping to train models with lower false rejection rates in fewer training epochs.

In addition, Yeganeh [69] proposes Inverse Distance Aggregation (IDA), a client-side adaptive weighting scheme based on meta-information. IDA calculates weights based on statistical meta-information, and clients can get higher weights, reducing the distance from the global average. Faced with the phenomenon of outdated clients caused by the problem of data heterogeneity, Li [70] proposed a workload prediction algorithm that automatically adjusts equipment training tasks. The algorithm uses the complete information of the historical training tasks of the device to predict the debt load of the device, thereby adaptively adjusting the training complexity of each client in each round. Aiming at the data heterogeneity problem in mobile edge computing, [71] proposed a multi-objective optimization problem to adaptively find the optimal number of iterations for privacy for each client.

As a conclusion, adaptive optimization can effectively deal with the non-IID data distribution problem in federated learning. By developing various adaptive optimization algorithms or frameworks to act on the objective functions of the global and local models, researchers aim to improve the client's ability to adapt to local data and enhance the model's resilience to the influence of heterogeneous data.

Algorithm 1 **FedAdagrad** , **FedYOGI** and **FedAdam**

Initialization: x_0, $v_{-1} \geq \tau^2$, optional decay $\beta_2 \in (0, 1)$ for FEDYOGI and FEDADAM
for $t = 0, \ldots, T-1$ **do**
  Sample subset S of clients;
  $x_{i,0}^t = x_t$
  for each client $i \in S$ in parallel **do**
    for $k = 0, \ldots, K-1$ **do**
      Compute an unbiased estimate $g_{i,k}^t$ of $\nabla F_i(x_{i,k}^t)$
      $x_{i,k+1}^t = x_{i,k}^t - \eta_l g_{i,k}^t$
    end for
    $\Delta_i^t = x_{i,K}^t - x_t$
  end for
  $\Delta_t = \frac{1}{|S|} \sum_{i \in S} \Delta_i^t$
  $v_t = v_{t-1} + \Delta_t^2$ **(FedAdagrad)**
  $v_t = v_{t-1} - (1-\beta_2)\Delta_t^2 sign(v_{t-1} - \Delta_t^2)$ **(FedYogi)**
  $v_t = \beta_2 v_{t-1} + (1-\beta_2)\Delta_t^2$ **(FedAdam)**
  $x_{t+1} = x_t + \eta \frac{\Delta_t}{\sqrt{v_t} + \tau}$
end for

3.2.3. Regular optimization

The non-IID data distribution in federated learning leads to slow convergence or even divergence of the global model. Model regularization optimization is a common strategy for preventing overfitting and improving convergence when training machine learning models. For example, in FedAvg, researchers add regular terms and nearest neighbor terms to the local objective function, which alleviates the problem of data heterogeneity and improves the stability of the model [72]. Furthermore, in order to balance the distance between the local model and the global model, [73] proposed the FedProx algorithm. The algorithm introduces an additional L2 regularization term on top of FedAvg's local objective function. This method of limiting local updates only needs to introduce a small amount of computational overhead, and the model can be adjusted to achieve good accuracy. Similarly, Li [74] proposed a teacher–student mechanism to adjust the gradients of clients in federated training from different data distributions by incorporating a regularization term of the objective function.

As a conclusion, regularized optimization is widely concerned as a beneficial method commonly used in machine learning to accelerate model convergence and improve model generalization ability. The training and convergence of federated learning models are affected by data inconsistencies of all parties, and the use of regularized optimization of global and local objective functions is beneficial to properly deal with the environment of data heterogeneity.

3.3. Algorithm-based

In general, the solutions to the non-IID problem in FL from a algorithm-based dimension mainly include meta Learning, multi-task learning and Life-long learning.

3.3.1. Meta learning

For federated learning in a typical environment of data heterogeneity, it is particularly important to learn a personalized global model for each device. Jiang [75] proposes a model-agnostic meta-learning setting to cope with the rapid adaptation of heterogeneous federated learning. The typical federated average algorithm is used as a meta-learning algorithm, and [75] is based on this for training and fine-tuning to achieve faster and better personalized training. Under the traditional computation then aggregation(CTA) protocol, Zhang [76] proposed a new meta-algorithm FedPD from the primal–dual point of view. The algorithm is robust to the nonconvex objective function and can achieve optimal optimization and good communication efficiency. In addition, Li [77] introduces a meta-learning method into a federated learning environment for spatiotemporal prediction tasks. By constructing a global spatiotemporal pattern map containing all city spatiotemporal information, [77] enables each client to use the global spatiotemporal pattern map to customize a local model and use these customized parameters to initialize the parameters of the model for spatiotemporal prediction.

As a conclusion, the learning goal of meta-learning is to quickly adapt to new similar tasks from a small amount of new data. Therefore, in federated learning, using detailed tasks of meta-learning can effectively personalize the local model of each client, so as to cope with the problem of data heterogeneity. The accuracy of the model is impaired and the phenomenon of stragglers.

3.3.2. Multi-task learning

The combination of multi-task learning and federated learning means that each participating client of federated learning learns different tasks and captures their internal relationships under the premise of privacy protection. Smith [78] proves that multi-task learning is naturally suitable for solving the statistical challenges of federation, and proposes a new system-aware optimization method MOCHA based on the distributed optimization method COCOA [79]. Optimizing federated learning to address heterogeneity-related system challenges through multi-task learning in MOCHA. As shown in Fig. 7, Corinzia [80] proposes the training process of federated multi-task learning. The parameter server learns the model relationship between multiple learning tasks according to the model parameters uploaded by each client, thereby updating each local model client's. Similarly, Chen [81] proposes a federated multi-task hierarchical attention model, which adopts multi-task learning with an attention mechanism aiming to extract feature representations from the input and learn shared representations for multiple devices.

As a conclusion, multi-task learning decomposes the federated learning task into multiple tasks for collaborative training, effectively adapting to the personalization of local heterogeneous data. During this period, each client optimizes the local model and parameter information, thereby realizing the vision of multi-party collaboration to improve the global model.

3.3.3. Life-long learning

Lifelong learning can transfer knowledge gained from new tasks or new data to all other tasks by continuously improving the shared representation of the classifier [82]. Some scholars have added the idea of life-long learning to the FL algorithm. For example, Shoham [17] proposed Elastic Weight Coalescing (EWC), which aims to prevent catastrophic forgetting when transferring from learning task A to task B. Unlike the FedAvg algorithm, which randomly selects clients to participate in the FL process, this paper uses a subset of data from all participating client nodes in FL. However, the fatal flaw of this method is that in the real application scenario of FL, it is difficult to ensure that all devices participate in each round of training.

In addition, Kopparapu [35] also proposed the FedFMC framework using the idea of life-long learning. The difference is that the latter does not treat each client in FL as a separate task to learn, but groups together similar prototype devices and treats each group as a learning task. The FedFMC framework includes fork and merging steps. In FedFMC, each participating device dynamically forks its own model into the best model after meeting the forking requirements. FedFMC then uses the FORK algorithm to make a global network of devices with a similar set of prototypes. Experiments demonstrate that FedFMC significantly improves the non-IID data problem without using a globally shared data subset and increasing communication costs.

As a conclusion, regularized optimization is widely concerned as a beneficial method commonly used in machine learning to accelerate model convergence and improve model generalization ability. The training and convergence of federated learning models are affected by data inconsistencies of all parties, and the use of regularized optimization of global and local objective functions is beneficial to properly deal with the environment of data heterogeneity.

3.4. Framework-based

In general, the solutions to the non-IID problem in FL from a framework-based dimension mainly include similarity clustering, knowledge distillation and Base + Personalization layer.

3.4.1. Similarity clustering

Although there are serious differences in the data of FL participating clients, some researchers have found that the impact of data heterogeneity on the model can be reduced by client similarity clustering. For example, Ghosh [83] proposes a client-side clustering method. As shown in Fig. 8, the method first finds an independent local optimal solution for each client, and then the client sends its model to the server. Finally, the server clusters the clients according to the local optimal solution. However, [83] does

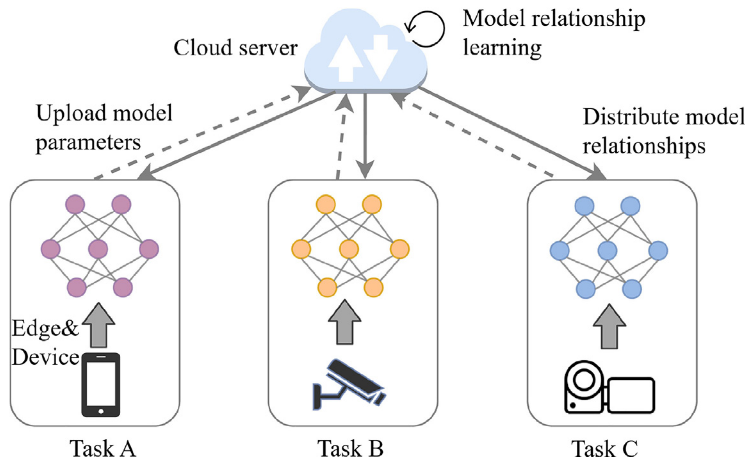


Fig. 7. Federated multi-task learning [80].

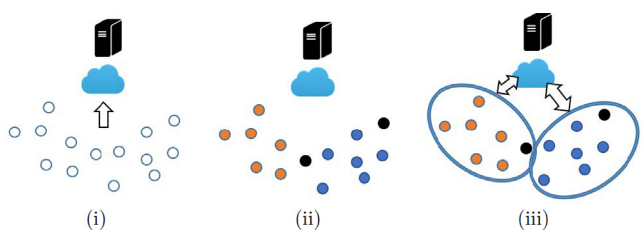


Fig. 8. A modular algorithm for Byzantine-robust optimization with heterogeneous data item [83].

not use neural networks because it assumes that the parameters of the learning algorithm are relatively low-dimensional representations. Briggs [84] introduces a clustering step after n rounds of communication in the FL process to hierarchically cluster client clusters based on the similarity between the client's local update and the global federated model. Dennis [85] proposed a one-shot clustering method for federated learning to address heterogeneity. The method first requires each device to solve a local clustering subproblem and compute the cluster average of the subproblems. After that, the central server accumulates and aggregates the results to compute the final cluster.

Clustering schemes reduce data heterogeneity challenges by identifying domains, but they fragment the data and sharing of each clustering model. Therefore, [86] proposed an iterative federated clustering algorithm (IFCA) based on cluster federated learning. This method does not require a centralized clustering algorithm, but instead uses a distributed setting of alternating minimization algorithms to alternate between estimating cluster identities and minimizing functions. Li [87] changed the paradigm by modifying the client-side clustering features into tensors of gradient vectors, computed from local data. Caldarola [88] proposed a novel cluster-driven model federated learning method. In this approach, graph convolutional networks (GCNs) are used to connect domain-specific modules to learn interactions between domains and share their knowledge. Furthermore, clustering is learned in an unsupervised setting through teacher–student classifier training. Therefore, data knowledge is shared across domains based on similarity criteria.

As a conclusion, similarity clustering reduces the impact of data heterogeneity on model performance by grouping customers with similar data distributions into similar clusters. Clustering can effectively balance the data difference, reduce the weight difference between the client cluster and the central server, and reduce the difficulty of FL training. However, adding some clustering steps to FL training may increase communication costs, especially when hundreds of customers are involved. The possible development direction in the future is to reduce the communication cost through methods such as communication compression, and realize the high-performance FL model.

3.4.2. Knowledge distillation

The federated distillation algorithm [16] was proposed under the influence of the mutual distillation algorithm [89], which aims to reduce the communication overhead of the model and reduce the damage to the model caused by the problem of data heterogeneity. Inspired by knowledge distillation, Itahara [43] proposes a distillation-based semi-supervised federated learning algorithm (DS-FL) using unlabeled open datasets. The DS-FL algorithm exchanges local model outputs rather than model parameters between participating devices, allowing control of communication costs without scaling with model size. Jiang [90] proposed a distributed federated training method based on knowledge distillation. Each device introduces a personalization model to adapt to local data, improving local performance. Among them, the performance improvement of the local device benefits from knowledge distillation, so that the knowledge transfer between heterogeneous networks can guide the improvement of the global model.

As a conclusion, the scheme based on knowledge distillation aims at mutual learning and transfer of data knowledge between the client and the server. The advantage of the scheme is that data knowledge is trained and learned by all parties in the state of data protection, which can effectively reduce the model communication overhead and improve the model convergence rate.

3.4.3. Base + Personalization layer

Aiming at the problem of the decrease of adaptive accuracy in heterogeneous data domains, [91] proposed a hierarchical heterogeneous horizontal federated learning method. Knowledge is accumulated and transferred across all clients. Structured data is mapped into a common embedding space. Afterwards, all clients build an EEG classifier to classify homogeneous features in a common embedding space. In the medical field where the heterogeneity of multi-center data is common, Andreux [92] proposed an isolated federated learning method, that is, the introduction of a local statistical batch normalization layer to achieve a depth robust to multi-center data Learning model variability. Furthermore, Chai [93] proposes an adaptive tier selection strategy that divides customers into tiers based on their training performance. Among them, the layers are updated in real-time over time based

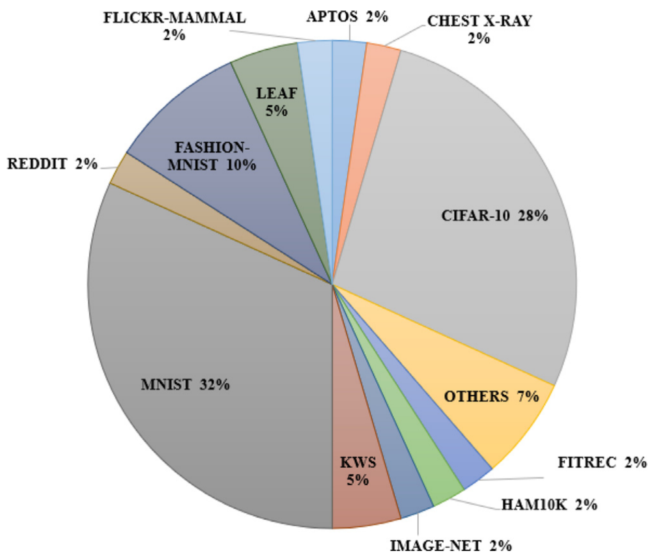


Fig. 9. Commonly used datasets types in FL.

on the observed training performance and accuracy, and customers are selected from the same layer in each round of training. Additionally, [94] proposes a novel hierarchical personalized federated learning framework for the statistical heterogeneity challenge, which consists of client–server. The framework divides data precisely into public and private information. For public information, the client uploads, and the server-side weighted fusion updates the global model to obtain global public information. For private information, the raw data is kept locally, and only the universe of local drafts is uploaded and aggregated to generate private data for the global model. This fine-grained personalized update strategy is used to accommodate statistical heterogeneity.

3.5. Datasets and data-setting

As shown in Fig. 9, among the datasets used by FL at present, the MNIST datasets is the most commonly used evaluation datasets, while graph data and sequential data are not widely used because of their nonlinear data structure characteristics. The MNIST database is a large handwritten digit database, which is widely used to train various image machine learning algorithms. The datasets contains 60,000 training images, 10,000 test images and corresponding labels [69,95,96]. The second popular CIFAR-10 datasets contains 60,000 32 × 32 color image datasets of 10 different categories, which are widely used in machine learning algorithms for object recognition [95,97]. In addition to the above two datasets, the FASHION-MNIST data [98] has the third highest application frequency. The FASHION-MNIST dataset has the same image size and category as the MNIST data and is often used for extended applications. In addition, more and more reality datasets are widely used, such as LEAF [99] and Flickr-Mammal [100]. Other datasets are relatively niche.

Based on these datasets, FL has been developed into various applications. This article takes FL's most commonly used MNIST dataset as an example, and counts the datasettings of the FL experimental environment in existing work, and summarizes the following three types:

- Class1 non-IID: Researchers process the data by classifying it, then distribute it to each client and allow it to store a single category of datasetting method. For example, [9] sorts the data according to the number label, and divides it into 200 pieces of data with a size of 300, and 2 pieces for every 100 clients. [16] randomly selects 2000 sample sets, divides the data into 10 subsets according to the real labels, and selects a predetermined number of target labels on each device. [84] partitions the data and distributes labels. The quantity received by the client only corresponds to 2 tags, and each client has 600 samples. In addition, there are some researchers who have similar datasetting methods [36,51].
- Class2 non-IID: Researchers distributes the data to each client device, and then classifies the data according to the datasetting method. For example, [95] sorts the dataset by category and divides it into slices. Each client allocates two slices. Similarly, [36] sorts the data by category and divides it into 20 partitions. Each client randomly allocates 2 partitions from the 2 categories. [84] shuffles the data and divides it into four groups, swapping the two-digit labels of each group.
- IID: Researchers distributes the data randomly and evenly to each client device's datasetting method. For example, [95] randomly divides the data sample into 50 datasets of equal size and scrambles them, and each client gets the same amount of data. [51] randomly allocates 60,000 samples to 100 devices, each with 600 samples. [36] distributes the 10 classes of the dataset uniformly and randomly to the client.

4. Trends and future directions

As a new distributed machine learning paradigm, FL has been successfully applied to computer vision, finance, healthcare, education, smart cities, blockchain, and other related applications. However, non-IID data and unbalanced data are not conducive to the performance of the FL model and become one of the urgent challenges to be solved. In recent years, although researchers have made great efforts to solve the non-IID data in FL, FL still faces a series of challenges. Through research, we discussed and summarized the following issues that may become future research trends:

4.1. Communication costs

The communication cost issue has always been an important factor restricting the development of federal learning. Under the influence of increasingly complex neural network models and non-IID data, FL communication costs are often more expensive. The level of communication costs has also become an important indicator for evaluating the effectiveness of non-IID solutions. The communication reduction technology adopted in the following research is an interesting topic to further improve FL communication efficiency.

- Communication compression: Existing compression schemes only compress the upstream communication from the client to the server, and they all focus on the sparseness of the weights or the weighting of the typical 32-bit floating-point numbers. For example, [13] uses Adam optimization and sparse compression methods to adjust FL.
  The Sparse Ternary compression scheme proposed by [38] uses top-k gradient sparsity technology to compress downstream communications. Adding other SGD optimization algorithms based on communication compression or compressing the downstream communication of the client from the server will be a possible research hotspot in the future.
- Model Pruning: The communication challenge in FL is that participating devices usually have the lower computing power and communication bandwidth than data center servers. Training large deep neural networks in this FL environment may consume a lot of time and other resources.

In this regard, some researchers [101] have proposed the FL paradigm that combines distributed model pruning, so that the model parameters have a high degree of freedom and can be adjusted freely according to the different calculation and communication capabilities of the clients in the FL process. The concept of pruning was first proposed by [102] for model pruning to remove redundant parameters. [103] proposed a model pruning method for centralized machine learning settings, but as the problem of [102], it requires training data to be provided on the central server and does not apply to federated learning. Adding different model optimization algorithms to model pruning to see its FL performance becomes a possible future development direction.

- Update strategy: Due to serious data differences between FL clients, the learning accuracy rate determined by local training and aggregation strategies is low. Reducing communication costs and improving the accuracy of FL models are the top priorities. [104] proposes an asynchronous strategy for partial model update and aggregation to improve the communication efficiency of each round. In the asynchronous learning strategy, the different layers of the deep neural network are divided into shallow and deep layers, and the parameter update frequency of the deep layer is set lower than the update frequency of the shallow layer to reduce communication. [105] proposes a control algorithm that realizes the best compromise between local update and global aggregation by analyzing the FL convergence range under non-IID data distribution. In the future, how to effectively use heterogeneous resources for distributed learning and the theoretical convergence analysis of the non-convex loss function of deep neural networks will be a direction worth exploring.

4.2. Privacy protection

Although FL can ensure that users do not expose local data privacy for collaborative training, the current privacy security of FL is still one of the challenges that FL needs to solve urgently. The privacy protection technology used below adds privacy and security assurance to the FL process and is an interesting topic for future research.

- Secure Multiparty Computation: Secure multiparty computing is a privacy protection scheme that aims to collaboratively calculate the result of a function from each party's private input without the need to make these inputs public [9]. For example, [5] proposes that the client device of FL can act as an secure multiparty computing participant to protect user update privacy. However, adding secure multiparty computing in the FL aggregation process will increase communication volume and computational complexity. Therefore, how to balance privacy protection and communication costs is also a hot topic in future research.
- Differential Privacy: Differential privacy is the earliest method used to promote the security analysis of sensitive data, and it has subsequently developed into a hot research topic in the field of machine learning. Differential privacy provides a security guarantee for information theory and is used to resist inference attacks by attackers. Differential privacy in FL is naturally applicable to aggregation algorithms such as FedAvg. Compared with secure multiparty computing, differential privacy weighs practicality to increase data privacy and can be applied to a variety of situations to alleviate the issue of issuing privacy weight updates during the communication round in the FL environment [16,106,107].
- Data redundancy: In addition to the above two traditional privacy protection methods, there are more commonly used data redundancy methods to protect data privacy. For example, [108] introduces data redundancy into FL through data exchange or overlapping data collection. Experiments have shown that doubling data storage under a strict energy budget can improve accuracy by 9.8%. How to quantify the impact of data redundancy and combine gradient compression or error accumulation to further reduce communication costs is a possible future research direction.

4.3. Federated learning framework

The FL framework is a popular paradigm that enables all participants to collaboratively train machine learning models while ensuring privacy and security. However, the traditional FL framework is weak in the face of challenges such as communication costs, data heterogeneity, and malicious attacks. Many researchers have begun to propose a federal learning framework that can solve these challenges.

- Federal Integration Framework: To overcome the influence of the data heterogeneity of the traditional FL framework, the researchers proposed the concept of merged FL [109]. Different from the traditional FL, the integrated FL framework divides the client group into alternatives with a data distribution that can be jointly trained, which has a high degree of error for the transformed client group. The inconsistency is that the program may increase additional communication costs. How to further analyze the similarity estimation of weight update and how to weigh communication cost and learning effect is the main research direction in the future.
- Asynchronous Online Federated Learning: Traditional FL's assumptions on data and equipment are too ideal, ignoring the heterogeneity of data and equipment in reality [110]. To solve the FL problem on distributed edge devices, an asynchronous online FL learning framework is proposed to solve the problems of edge device load, lag, or withdrawal. Similar research solves the problem of device heterogeneity through active device selection [51]. Some researchers have proposed a new federated aggregation scheme, which uses scaling the corresponding aggregation coefficients to achieve incomplete local updates, but this scheme has not been verified and its effectiveness can be verified in the future [111].
- Federal Extreme Learning Machine System: [112] adds the idea of extreme learning machine algorithm to FL on edge devices. The ELM algorithm randomly generates its input weights and hidden biases instead of iteratively adjusting each parameter through a gradient-based algorithm. In addition, ELM also directly calculates its output weight through the Moore–Penrose pseudo-inverse. This solution can greatly improve the model training speed, has excellent generalization performance and efficient computing power. The extreme learning machine algorithm provides greater development space for the edge FL.

4.4. Personalized learning method

To balance the impact of non-IID data differences on the global sharing model, researchers have proposed a variety of personalized optimization techniques to suit each client in FL. The article investigates the following popular methods, summarizes and analyzes the future development trends.

- Adaptive optimization: As we mentioned earlier, adaptive optimization can improve the convergence speed of FL and effectively deal with the problem of non-IID data distribution. For example, [113] proposed to add ADAM and YOGI to FedAvg to implement adaptive federated learning. In addition, [67] also added the adaptive optimizer ADAGRAD to FL. Some researchers have proposed a meta-information-based client adaptive weighting method to deal with the imbalance and non-IID data [69]. In the future, the field of adaptive optimization should focus on exploring the use of momentum and adaptive client optimizers and how to heuristically set learning rates and other hyperparameters.
- Multitask Learning: Limited by privacy, communication, and data heterogeneity, the performance of the global model obtained by FL training may be worse than that of the local model. The vast majority of existing aggregation algorithms aim to improve the robustness and accuracy of the global federation model, while [114] focuses on the robustness and accuracy of the participant federation model. They evaluated the federation in the article three techniques for model local adaptation: fine-tuning, multi-task learning, and knowledge distillation. To alleviate over-adaptation, the article treats local adaptation as a multi-task learning problem. For example, a federated model optimized for task A is used to create an adaptive model optimized for task B.
- Semi-supervised Learning: A series of model output exchange operations in the FL framework will generate a lot of communication overhead, and the model output dimension is the key to determining the size of the communication overhead. [43] proposes a semi-supervised FL algorithm, the core of which is to exchange model outputs of unlabeled open datasets, rather than exchange model parameters. Because the model output is exchanged instead of the model parameters, the communication overhead does not increase in proportion to the model size. In the absence of a fully annotated training datasets, semi-supervised FL will play a huge potential.

In general, researchers are committed to solving non-IID problems in FL to ensure data privacy and security, and to improve the performance of the model in terms of communication rounds, convergence speed, and model accuracy [115,116]. As mentioned earlier, the difficulty in solving non-IID data in FL lies in the degree of data heterogeneity. Therefore, the important future direction is to design effective algorithms to quickly calculate the degree of statistical heterogeneity. Besides, the system heterogeneity that accompanies data heterogeneity will also affect FL performance. Therefore, when solving non-IID data problems, attention should be paid to whether hardware and software limitations will affect the overall effect. This will be another problem to be solved in the future. Finally, FL needs to solve the problem of fault tolerance to ensure that every device participating in the FL can be guaranteed and improve the practical application capabilities of the FL.

5. Conclusion

Federated learning, as a new paradigm of distributed machine learning, has gained wide attention in academia and industry. In this paper, we identify a potentially important challenge in federated learning: non-IID distribution of data In FL. We introduce the specific challenges brought by non-IID data to FL, sort out the existing solutions, and discuss and point out the prospects of FL on non-IID data. To the best of our knowledge, this is the latest authoritative review article on addressing the problem of non-IID data in FL. The comprehensive systematic literature review provided in this paper are expected to guide researchers and practitioners to understand the state-of-the-art and conduct future studies in FL with non-IID data.

CRediT authorship contribution statement

Xiaodong Ma: Conceptualization, Methodology, Data curation, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. **Jia Zhu:** Conceptualization, Supervision, Project administration, Funding acquisition. **Zhihao Lin:** Software, Validation, Investigation, Data curation. **Shanxuan Chen:** Validation, Formal analysis. **Yangjie Qin:** Data curation, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Key Laboratory of Intelligent Education Technology and Application of Zhejiang Province and the National Natural Science Foundation of China under Grant 62077015.

References

[1] P. Voigt, A. Von dem Bussche, The eu general data protection regulation (gdpr), in: A Practical Guide, St Ed., Springer International Publishing, Cham, 2017.
[2] M. Abadi, A. Chu, I. Goodfellow, H.B. McMahan, I. Mironov, K. Talwar, L. Zhang, Deep learning with differential privacy, in: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 2016, pp. 308–318.
[3] Q. Yang, Y. Liu, Y. Cheng, Y. Kang, T. Chen, H. Yu, Federated learning, Synth. Lect. Artif. Intell. Mach. Learn. 13 (3) (2019) 1–207.
[4] H.B. McMahan, E. Moore, D. Ramage, B.A. y Arcas, Federated learning of deep networks using model averaging, 2016.
[5] K. Bonawitz, V. Ivanov, B. Kreuter, A. Marcedone, H.B. McMahan, S. Patel, D. Ramage, A. Segal, K. Seth, Practical secure aggregation for privacy-preserving machine learning, in: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, 2017, pp. 1175–1191.
[6] M. Aledhari, R. Razzak, R.M. Parizi, F. Saeed, Federated learning: A survey on enabling technologies, protocols, and applications, IEEE Access 8 (2020) 40699–40725.
[7] Z. Li, V. Sharma, S. P. Mohanty, Preserving data privacy via federated learning: Challenges and solutions, IEEE Consum. Electron. Mag. 9 (3) (2020) 8–16.
[8] W.Y.B. Lim, N.C. Luong, D.T. Hoang, Y. Jiao, Y.-C. Liang, Q. Yang, D. Niyato, C. Miao, Federated learning in mobile edge networks: A comprehensive survey, IEEE Commun. Surv. Tutor. 22 (3) (2020) 2031–2063.
[9] B. McMahan, E. Moore, D. Ramage, S. Hampson, B.A. y Arcas, Communication-efficient learning of deep networks from decentralized data, in: Artificial Intelligence and Statistics, PMLR, 2017, pp. 273–282.
[10] H.B. McMahan, D. Ramage, K. Talwar, L. Zhang, Learning differentially private recurrent language models, 2017, arXiv preprint arXiv:70.06963.
[11] Y. Liu, A. Huang, Y. Luo, H. Huang, Y. Liu, Y. Chen, L. Feng, T. Chen, H. Yu, Q. Yang, FedVision: An online visual object detection platform powered by federated learning, in: AAAI, 2020, pp. 372–379.
[12] H. Yu, Z. Liu, Y. Liu, T. Chen, M. Cong, X. Weng, D. Niyato, Q. Yang, A sustainable incentive scheme for federated learning, IEEE Intell. Syst. (2020).
[13] J. Mills, J. Hu, G. Min, Communication-efficient federated learning for wireless edge intelligence in IoT, IEEE Internet Things J. 7 (7) (2020) 5986–5994.
[14] X. Wang, Y. Han, C. Wang, Q. Zhao, X. Chen, M. Chen, In-edge ai: Intelligentizing mobile edge computing, caching and communication by federated learning, IEEE Netw. 33 (5) (2019) 156–165.
[15] K. Kopparapu, E. Lin, J. Zhao, FedCD: Improving performance in non-IID federated learning, 2020, arXiv preprint arXiv:2006.09637.
[16] E. Jeong, S. Oh, H. Kim, J. Park, M. Bennis, S.-L. Kim, Communication-efficient on-device machine learning: Federated distillation and augmentation under non-iid private data, 2018, arXiv preprint arXiv:8.479.

[17] N. Shoham, T. Avidor, A. Keren, N. Israel, D. Benditkis, L. Mor-Yosef, I. Zeitak, Overcoming forgetting in federated learning on non-IID data, 2019, arXiv preprint arXiv:90.07796.
[18] H. Zhu, J. Xu, S. Liu, Y. Jin, Federated learning on non-IID data: A survey, Neurocomputing 465 (2021) 371–390.
[19] Q. Yang, Y. Liu, T. Chen, Y. Tong, Federated machine learning: Concept and applications, ACM Trans. Intell. Syst. Technol. (TIST) 10 (2) (2019) 1–19.
[20] J. Xu, B.S. Glicksberg, C. Su, P. Walker, J. Bian, F. Wang, Federated learning for healthcare informatics, J. Healthc. Inform. Res. 5 (1) (2021) 1–19.
[21] V. Kulkarni, M. Kulkarni, A. Pant, Survey of personalization techniques for federated learning, in: 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4), IEEE, 2020, pp. 794–797.
[22] Z. Du, C. Wu, T. Yoshinaga, K.-L.A. Yau, Y. Ji, J. Li, Federated learning for vehicular internet of things: Recent advances and open issues, IEEE Open J. Comput. Soc. 1 (2020) 45–61.
[23] S. AbdulRahman, H. Tout, H. Ould-Slimane, A. Mourad, C. Talhi, M. Guizani, A survey on federated learning: The journey from centralized to distributed on-site learning and beyond, IEEE Internet Things J. 8 (7) (2020) 5476–5497.
[24] C. Zhang, Y. Xie, H. Bai, B. Yu, W. Li, Y. Gao, A survey on federated learning, Knowl.-Based Syst. 216 (2021) 106775.
[25] O.A. Wahab, A. Mourad, H. Otrok, T. Taleb, Federated machine learning: Survey, multi-level classification, desirable criteria and future directions in communication and networking systems, IEEE Commun. Surv. Tutor. 23 (2) (2021) 1342–1397.
[26] D. Jiang, C. Shan, Z. Zhang, Federated learning algorithm based on knowledge distillation, in: 2020 International Conference on Artificial Intelligence and Computer Engineering (ICAICE), IEEE, 2020, pp. 163–167.
[27] D.C. Nguyen, M. Ding, P.N. Pathirana, A. Seneviratne, J. Li, H.V. Poor, Federated learning for internet of things: A comprehensive survey, IEEE Commun. Surv. Tutor. (2021).
[28] H. Zhu, H. Zhang, Y. Jin, From federated learning to federated neural architecture search: a survey, Complex Intell. Syst. 7 (2) (2021) 639–657.
[29] A. Imteaj, U. Thakker, S. Wang, J. Li, M.H. Amini, A survey on federated learning for resource-constrained iot devices, IEEE Internet Things J. 9 (1) (2021) 1–24.
[30] X. Yin, Y. Zhu, J. Hu, A comprehensive survey of privacy-preserving federated learning: A taxonomy, review, and future directions, ACM Comput. Surv. 54 (6) (2021) 1–36.
[31] N. Bouacida, P. Mohapatra, Vulnerabilities in federated learning, IEEE Access 9 (2021) 63229–63249.
[32] C. Briggs, Z. Fan, P. Andras, A review of privacy preserving federated learning for private IoT analytics, 2020, arXiv preprint arXiv:2004.794.
[33] K. Hsieh, A. Phanishayee, O. Mutlu, P.B. Gibbons, The non-IID data quagmire of decentralized machine learning, 2019, arXiv preprint arXiv:90.0089.
[34] T.T. Phuong, L.T. Phong, Distributed SGD with flexible gradient compression, 8 (2020) 64707–64717.
[35] K. Kopparapu, E. Lin, FedFMC: Sequential efficient federated learning on non-iid data, 2020, arXiv preprint arXiv:2006.0937.
[36] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, V. Chandra, Federated learning with non-iid data, 2018, arXiv preprint arXiv:806.00582.
[37] S.K. Lo, Q. Lu, C. Wang, H. Paik, L. Zhu, A systematic literature review on federated machine learning: From a software engineering perspective, 2020, arXiv preprint arXiv:2007.354.
[38] F. Sattler, S. Wiedemann, K.-R. Müller, W. Samek, Robust and communication-efficient federated learning from non-iid data, IEEE Trans. Neural Netw. Learn. Syst. (2019).
[39] X. Li, K. Huang, W. Yang, S. Wang, Z. Zhang, On the convergence of fedavg on non-iid data, 2019, arXiv preprint arXiv:907.0289.
[40] F. Sattler, K.-R. Müller, W. Samek, Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints, IEEE Trans. Neural Netw. Learn. Syst. (2020).
[41] F. Hanzely, P. Richtárik, Federated learning of a mixture of global and local models, 2020, arXiv preprint arXiv:2002.0556.
[42] T. Yu, E. Bagdasaryan, V. Shmatikov, Salvaging federated learning by local adaptation, 2020, arXiv preprint arXiv:2002.04758.
[43] S. Itahara, T. Nishio, Y. Koda, M. Morikura, K. Yamamoto, Distillation-based semi-supervised federated learning for communication-efficient collaborative training with non-IID private data, 2020, arXiv preprint arXiv:2008.0680.
[44] L. Collins, H. Hassani, A. Mokhtari, S. Shakkottai, Exploiting shared representations for personalized federated learning, in: International Conference on Machine Learning, PMLR, 2021, pp. 2089–2099.
[45] X.-C. Li, D.-C. Zhan, Y. Shao, B. Li, S. Song, FedPHP: Federated personalization with inherited private models, in: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer, 2021, pp. 587–602.
[46] P. Tian, Z. Chen, W. Yu, W. Liao, Towards asynchronous federated learning based threat detection: A DC-Adam approach, Comput. Secur. 108 (2021) 102344.
[47] M. Shin, C. Hwang, J. Kim, J. Park, M. Bennis, S.-L. Kim, XOR mixup: Privacy-preserving data augmentation for one-shot federated learning, 2020, arXiv preprint arXiv:2006.0548.
[48] W. Hao, M. El-Khamy, J. Lee, J. Zhang, K.J. Liang, C. Chen, L.C. Duke, Towards fair federated learning with zero-shot data augmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 3310–3319.
[49] L. Cai, D. Lin, J. Zhang, S. Yu, Dynamic sample selection for federated learning with heterogeneous data in fog computing, in: ICC 2020-2020 IEEE International Conference on Communications (ICC), IEEE, 2020, pp. 1–6.
[50] C. Wang, Y. Yang, P. Zhou, Towards efficient scheduling of federated mobile devices under computational and statistical heterogeneity, IEEE Trans. Parallel Distrib. Syst. 32 (2) (2020) 394–410.
[51] H. Wang, Z. Kaplan, D. Niu, B. Li, Optimizing federated learning on non-IID data with reinforcement learning, in: IEEE INFOCOM 2020-IEEE Conference on Computer Communications, IEEE, 2020, pp. 698–707.
[52] X. Xu, S. Duan, J. Zhang, Y. Luo, D. Zhang, Optimizing federated learning on device heterogeneity with a sampling strategy, in: 2021 IEEE/ACM 29th International Symposium on Quality of Service (IWQOS), IEEE, 2021, pp. 1–10.
[53] Z. Wang, Y. Zhu, D. Wang, Z. Han, FedACS: Federated skewness analytics in heterogeneous decentralized data environments, in: 2021 IEEE/ACM 29th International Symposium on Quality of Service (IWQOS), IEEE, 2021, pp. 1–10.
[54] A. Taïk, H. Moudoud, S. Cherkaoui, Data-quality based scheduling for federated edge learning, in: 2021 IEEE 46th Conference on Local Computer Networks (LCN), IEEE, 2021, pp. 17–23.
[55] Y. Zeng, X. Wang, J. Yuan, J. Zhang, J. Wan, Local epochs inefficiency caused by device heterogeneity in federated learning, Wirel. Commun. Mob. Comput. 2022 (2022).
[56] E. Sannara, F. Portet, P. Lalanda, V. German, A federated learning aggregation algorithm for pervasive computing: Evaluation and comparison, in: 2021 IEEE International Conference on Pervasive Computing and Communications (PerCom), IEEE, 2021, pp. 1–10.
[57] S. Kwatra, V. Torra, A k-anonymised federated learning framework with decision trees, in: Data Privacy Management, Cryptocurrencies and Blockchain Technology, Springer, 2021, pp. 106–120.
[58] Y. Qin, M. Kondo, MLMG: Multi-local and multi-global model aggregation for federated learning, in: 2021 IEEE International Conference on Pervasive Computing and Communications Workshops and Other Affiliated Events (PerCom Workshops), IEEE, 2021, pp. 565–571.
[59] A. Abay, Y. Zhou, N. Baracaldo, S. Rajamoni, E. Chuba, H. Ludwig, Mitigating bias in federated learning, 2020, arXiv preprint arXiv:2012.02447.
[60] S.P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich, A.T. Suresh, Scaffold: Stochastic controlled averaging for federated learning, in: International Conference on Machine Learning, PMLR, 2020, pp. 5132–5143.
[61] Z. Chai, Y. Chen, A. Anwar, L. Zhao, Y. Cheng, H. Rangwala, FedAT: a high-performance and communication-efficient federated learning system with asynchronous tiers, in: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, 2021, pp. 1–16.
[62] Q. Ma, Y. Xu, H. Xu, Z. Jiang, L. Huang, H. Huang, FedSA: A semi-asynchronous federated learning mechanism in heterogeneous edge computing, IEEE J. Sel. Areas Commun. 39 (12) (2021) 3654–3672.
[63] Y. Chen, Y. Ning, M. Slawski, H. Rangwala, Asynchronous online federated learning for edge devices with non-iid data, in: 2020 IEEE International Conference on Big Data (Big Data), IEEE, 2020, pp. 15–24.
[64] Y. Tan, G. Long, L. Liu, T. Zhou, Q. Lu, J. Jiang, C. Zhang, FedProto: Federated prototype learning across heterogeneous clients, in: AAAI Conference on Artificial Intelligence, Vol. 1, 2022.
[65] D. Leroy, A. Coucke, T. Lavril, T. Gisselbrecht, J. Dureau, Federated learning for keyword spotting, in: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2019, pp. 6341–6345.
[66] J. Zhang, S.P. Karimireddy, A. Veit, S. Kim, S.J. Reddi, S. Kumar, S. Sra, Why adam beats sgd for attention models, 2019.
[67] S. Reddi, Z. Charles, M. Zaheer, Z. Garrett, K. Rush, J. Konečný, S. Kumar, H.B. McMahan, Adaptive federated optimization, 2020, arXiv preprint arXiv:2003.00295.
[68] M. Zaheer, S.J. Reddi, D.S. Sachan, S. Kale, S. Kumar, Adaptive methods for nonconvex optimization, in: NeurIPS, 2018, pp. 9815–9825.
[69] Y. Yeganeh, A. Farshad, N. Navab, S. Albarqouni, Inverse distance aggregation for federated learning with non-iid data, in: Domain Adaptation and Representation Transfer, and Distributed and Collaborative Learning, Springer, 2020, pp. 150–159.

[70] L. Li, M. Duan, D. Liu, Y. Zhang, A. Ren, X. Chen, Y. Tan, C. Wang, FedSAE: A novel self-adaptive federated learning framework in heterogeneous systems, in: 2021 International Joint Conference on Neural Networks (IJCNN), IEEE, 2021, pp. 1–10.

[71] J. Zhang, X. Cheng, C. Wang, Y. Wang, Z. Shi, J. Jin, A. Song, W. Zhao, L. Wen, T. Zhang, FedAda: Fast-convergent adaptive federated learning in heterogeneous mobile edge computing environment, World Wide Web (2022) 1–28.

[72] A.K. Sahu, T. Li, M. Sanjabi, M. Zaheer, A. Talwalkar, V. Smith, On the convergence of federated optimization in heterogeneous networks, 3 (2018). arXiv preprint arXiv:82.0627.

[73] T. Li, A.K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, V. Smith, Federated optimization in heterogeneous networks, in: MLSys, mlsys.org, 2020.

[74] X. Li, N. Liu, C. Chen, Z. Zheng, H. Li, Q. Yan, Communication-efficient collaborative learning of geo-distributed JointCloud from heterogeneous datasets, in: 2020 IEEE International Conference on Joint Cloud Computing, IEEE, 2020, pp. 22–29.

[75] Y. Jiang, J. Konečnỳ, K. Rush, S. Kannan, Improving federated learning personalization via model agnostic meta learning, 2019, arXiv preprint arXiv:1909.12488.

[76] X. Zhang, M. Hong, S. Dhople, W. Yin, Y. Liu, Fedpd: A federated learning framework with adaptivity to non-iid data, IEEE Trans. Signal Process. 69 (2021) 6055–6070.

[77] W. Li, S. Wang, Federated meta-learning for spatial-temporal prediction, Neural Comput. Appl. (2022) 1–20.

[78] V. Smith, C.-K. Chiang, M. Sanjabi, A.S. Talwalkar, Federated multi-task learning, Adv. Neural Inf. Process. Syst. 30 (2017).

[79] M. Jaggi, V. Smith, M. Takác, J. Terhorst, S. Krishnan, T. Hofmann, M.I. Jordan, Communication-efficient distributed dual coordinate ascent, Adv. Neural Inf. Process. Syst. 27 (2014).

[80] L. Corinzia, A. Beuret, J.M. Buhmann, Variational federated multi-task learning, 2019, arXiv preprint arXiv:1906.06268.

[81] Y. Chen, Y. Ning, Z. Chai, H. Rangwala, Federated multi-task learning with hierarchical attention for sensor data analytics, in: 2020 International Joint Conference on Neural Networks (IJCNN), IEEE, 2020, pp. 1–8.

[82] S. Kim, R. Mowakeaa, S.-J. Kim, H. Lim, Building energy management for demand response using kernel lifelong learning, IEEE Access 8 (2020) 82131–82141.

[83] A. Ghosh, J. Hong, D. Yin, K. Ramchandran, Robust federated learning in a heterogeneous environment, 2019, arXiv preprint arXiv:906.06629.

[84] C. Briggs, Z. Fan, P. Andras, Federated learning with hierarchical clustering of local updates to improve training on non-IID data, 2020, arXiv preprint arXiv:2004.79.

[85] D.K. Dennis, T. Li, V. Smith, Heterogeneity for the win: One-shot federated clustering, in: International Conference on Machine Learning, PMLR, 2021, pp. 2611–2620.

[86] A. Ghosh, J. Chung, D. Yin, K. Ramchandran, An efficient framework for clustered federated learning, in: NeurIPS, 2020.

[87] Y. Luo, X. Liu, J. Xiu, Energy-efficient clustering to address data heterogeneity in federated learning, in: ICC 2021-IEEE International Conference on Communications, IEEE, 2021, pp. 1–6.

[88] D. Caldarola, M. Mancini, F. Galasso, M. Ciccone, E. Rodolà, B. Caputo, Cluster-driven graph federated learning over multiple domains, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 2749–2758.

[89] R. Anil, G. Pereyra, A. Passos, R. Ormandi, G.E. Dahl, G.E. Hinton, Large scale distributed neural network training through online distillation, 2018, arXiv preprint arXiv:1804.03235.

[90] J.C. Jiang, B. Kantarci, S. Oktug, T. Soyata, Federated learning in smart city sensing: Challenges and opportunities, Sensors 20 (21) (2020) 6230.

[91] D. Gao, C. Ju, X. Wei, Y. Liu, T. Chen, Q. Yang, Hhhfl: Hierarchical heterogeneous horizontal federated learning for electroencephalography, 2019, arXiv preprint arXiv:1909.05784.

[92] M. Andreux, J.O.d. Terrail, C. Beguier, E.W. Tramel, Siloed federated learning for multi-centric histopathology datasets, in: Domain Adaptation and Representation Transfer, and Distributed and Collaborative Learning, Springer, 2020, pp. 129–139.

[93] Z. Chai, A. Ali, S. Zawad, S. Truex, A. Anwar, N. Baracaldo, Y. Zhou, H. Ludwig, F. Yan, Y. Cheng, Tifl: A tier-based federated learning system, in: Proceedings of the 29th International Symposium on High-Performance Parallel and Distributed Computing, 2020, pp. 125–136.

[94] J. Wu, Q. Liu, Z. Huang, Y. Ning, H. Wang, E. Chen, J. Yi, B. Zhou, Hierarchical personalized federated learning for user modeling, in: Proceedings of the Web Conference 2021, 2021, pp. 957–968.

[95] J. Mills, J. Hu, G. Min, Communication-efficient federated learning for wireless edge intelligence in iot, IEEE Internet Things J. (2019).

[96] Y. Sun, S. Zhou, D. Gündüz, Energy-aware analog aggregation for federated learning with redundant data, 2019, arXiv preprint arXiv:9.0088.

[97] A. Li, J. Sun, B. Wang, L. Duan, S. Li, Y. Chen, H. Li, LotteryFL: Personalized and communication-efficient federated learning with lottery ticket hypothesis on non-IID datasets, 2020, arXiv preprint arXiv:2008.0337.

[98] H. Xiao, K. Rasul, R. Vollgraf, Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms, 2017, CoRR abs/1708.07747.

[99] S. Caldas, P. Wu, T. Li, J. Konečný, H.B. McMahan, V. Smith, A. Talwalkar, LEAF: a benchmark for federated settings, 2018, CoRR abs/1812.01097.

[100] K. Hsieh, A. Phanishayee, O. Mutlu, P. Gibbons, The non-IID data quagmire of decentralized machine learning, in: International Conference on Machine Learning (ICML), 2020, pp. 5819–5830.

[101] Y. Jiang, S. Wang, B. Ko, W. Lee, L. Tassiulas, Model pruning enables efficient federated learning on edge devices, 2019, CoRR abs/1909.12326.

[102] Y. LeCun, J.S. Denker, S.A. Solla, Optimal brain damage, in: D.S. Touretzky (Ed.), Advances in Neural Information Processing Systems 2, [NIPS Conference, Denver, Colorado, USA, November 27-30, 1989], 1989, pp. 598–605.

[103] S. Han, J. Pool, J. Tran, W.J. Dally, Learning both weights and connections for efficient neural network, in: C. Cortes, N.D. Lawrence, D.D. Lee, M. Sugiyama, R. Garnett (Eds.), Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada, 2015, pp. 1135–1143.

[104] Y. Chen, X. Sun, Y. Jin, Communication-efficient federated deep learning with asynchronous model update and temporally weighted aggregation, 2019, arXiv preprint arXiv:903.07424.

[105] S. Wang, T. Tuor, T. Salonidis, K.K. Leung, C. Makaya, T. He, K. Chan, Adaptive federated learning in resource constrained edge computing systems, IEEE J. Sel. Areas Commun. 37 (6) (2019) 1205–1221.

[106] H. Zheng, H. Hu, Z. Han, Preserving user privacy for machine learning: Local differential privacy or federated machine learning? IEEE Intell. Syst. 35 (4) (2020) 5–14.

[107] R. Hu, Y. Guo, H. Li, Q. Pei, Y. Gong, Personalized federated learning with differential privacy, IEEE Internet Things J. 7 (10) (2020) 9530–9539.

[108] Y. Sun, S. Zhou, D. Gündüz, Energy-aware analog aggregation for federated learning with redundant data, in: 2020 IEEE International Conference on Communications, ICC 2020, Dublin, Ireland, June 7-11, 2020, IEEE, 2020, pp. 1–7.

[109] F. Sattler, K. Muller, W. Samek, Clustered federated learning: Model-agnostic distributed multi-task optimization under privacy constraints, Learning (2019) ArXiv:.

[110] Y. Chen, Y. Ning, H. Rangwala, Asynchronous online federated learning for edge devices, 2019, CoRR abs/1911.02134.

[111] Y. Ruan, X. Zhang, S. Liang, C. Joe-Wong, Towards flexible device participation in federated learning for non-IID data, 2020, CoRR abs/2006.06954.

[112] Y.-T. Chen, Y.-C. Chuang, A.-Y.A. Wu, Online extreme learning machine design for the application of federated learning, in: 2020 2nd IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS), IEEE, 2020, pp. 188–192.

[113] A. Hard, K. Partridge, C. Nguyen, N. Subrahmanya, A. Shah, P. Zhu, I.L. Moreno, R. Mathews, Training keyword spotting models on non-IID data with federated learning, 2020, arXiv preprint arXiv:2005.0406.

[114] T. Yu, E. Bagdasaryan, V. Shmatikov, Salvaging federated learning by local adaptation, 2020, CoRR abs/2002.04758.

[115] X. Yao, T. Huang, C. Wu, R.-X. Zhang, L. Sun, Federated learning with additional mechanisms on clients to reduce communication costs, 2019, arXiv preprint arXiv:908.0589.

[116] W. Chen, K. Bhardwaj, R. Marculescu, Fedmax: mitigating activation divergence for accurate and communication-efficient federated learning, in: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer, 2020, pp. 348–363.

Xiaodong Ma holds a bachelor's degree in computer science and technology and is currently a postgraduate student at South China Normal University. She plans to pursue a Ph.D. degree in intelligent education at Zhejiang Normal University. Her research interests are primarily in the area of federated learning and blockchain.

Organization:Zhejiang Normal University
Email: felicity@m.scnu.edu.cn

JIa Zhu is a Distinguished Professor at the School of Teacher Education, Zhejiang Normal University, and the Deputy Director of the Key Laboratory of Intelligent Education Technology and Application of Zhejiang Province. He received his Ph.D. degree from the University of Queensland, Australia. His research interests include intelligent education, theoretical algorithms for database and data mining, federated learning, and blockchain with AI.

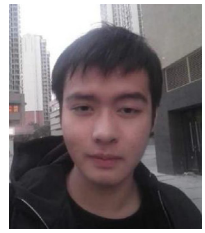
Organization: Zhejiang Normal University
Email: jiazhu@zjnu.edu.cn

Zhihao Lin holds a bachelor's degree in computer science and technology and is currently a postgraduate student at South China Normal University. His research interests are primarily in the area of federated learning.

Organization: South China Normal University
Email: zhihaolin@m.scnu.edu.cn

Yangjie Qin holds a bachelor's degree in mathematics and is currently a postgraduate student at South China Normal University. His research interests are primarily in the area of federated learning.

Organization: South China Normal University
Email: eurusyoung@m.scnu.edu.cn



Shanxuan Chen holds a bachelor's degree in computer science and technology and is currently a postgraduate student at South China Normal University. His research interests are primarily in the area of federated learning and blockchain.

Organization: South China Normal University
Email: shanxuanchen@m.scnu.edu.cn