DOI：10.20189/j.cnki.CN/61-1527/E.202504011

联邦标签噪声学习研究综述

焦怡博[1]，姚俊萍[1*]，李晓军[1]，苏　逸[1,2]，王　忠[1]，辜弘炀[1]
（1. 火箭军工程大学，西安 710025；
2. 国防科技大学 计算机学院，长沙 410000）

摘　要：针对私有分散数据中存在的标签噪声对模型训练效率和预测结果存在显著影响的问题，从构建标签噪声鲁棒的联邦学习系统角度出发，归纳总结了联邦标签噪声学习的相关研究成果。首先，阐述了联邦学习、标签噪声学习及联邦标签噪声学习的相关概念与定义；其次，根据是否使用显式的标签噪声检测机制将现有的联邦标签噪声学习方法分为基于噪声检测的训练方法、鲁棒的噪声容忍训练方法以及混合方法3类，并比较和分析了不同方法的执行机制、特征、关联性与优缺点；最后，对联邦标签噪声学习领域潜在的研究方向和发展趋势进行了展望，以期为相关研究提供参考和借鉴。

关键词：联邦学习；标签噪声；联邦标签噪声学习；标签噪声鲁棒；标签噪声检测

中图分类号：TP391.4　　**文献标志码**：A　　**开放科学标识码**（OSID）：

文章编号：2097-2741（2025）04-095-19

听语音　聊科研　与作者互动

Review on Federated Noisy Label Learning Researches

JIAO Yibo[1], YAO Junping[1*], LI Xiaojun[1], SU Yi[1,2],
WANG Zhong[1], GU Hongyang[1]
（1. Rocket Force University of Engineering, Xi'an 710025, China;
2. College of Computer Science and Technology, National University of Defense Technology, Changsha 410000, China）

ABSTRACT: Relevant research achievements in federated noisy label learning (FNLL) were synthesized and summarized from the perspective of building label-noise-robust federated learning systems. This work addressed the significant impact of noisy label in private and distributed data on model training efficiency and prediction performance. First, the key concepts and definitions related to federated learning, noisy label learning, and FNLL were elaborated. Second, based on the presence or absence of explicit noisy label detection mechanisms, existing FNLL methods were classified into three categories: noise-detection-based training methods, robust noise-tolerant training methods, and hybrid methods. The operational mechanisms, characteristics, interrela-

收稿日期：2025-03-23　　**修回日期**：2025-05-11　　**录用日期**：2025-05-30

基金项目：国家自然科学基金（62401609）；中国博士后科学基金（2024M754275）；陕西省自然科学基础研究计划（2025JC-YBMS-783）

第一作者：焦怡博（2000—），男，博士研究生，主要从事联邦学习、标签噪声处理研究。E-mail：email.jiaoyibo@163.com

***通信作者**：姚俊萍（1978—），女，教授，主要从事智能信息处理、装备数据工程领域研究。E-mail：junpingy200225@163.com

引用格式：焦怡博，姚俊萍，李晓军，等.联邦标签噪声学习研究综述[J].火箭军工程大学学报，2025，39（4）：95-113. *JIAO Yibo, YAO Junping, LI Xiaojun, et al. Review on federated noisy label learning researches[J]. Journal of Rocket Force University of Engineering, 2025, 39(4): 95-113.*

tionships, advantages, and disadvantages of each category were compared and analyzed. Finally, potential research directions and future development trends in FNLL were discussed, aiming to provide valuable insights and references for ongoing and future researches in this field.

KEYWORDS: federated learning; noisy label; federated noisy label learning; noisy label robust; noisy label detection

在数据赋能的时代背景下，数据已经成为一种战略性资源，足量的高质量数据对于模型训练、任务达成等至关重要。但高质量数据的获取往往伴随着高昂的成本，这已经成为制约各领域发展的一个关键要素。尤其是在深度学习领域中，其取得成功的核心因素之一在于高质量数据集的获取和构建[1]。

数据标签的正确与否是评价数据质量好坏的一个重要体现。一方面，随着自动化标注工具[2]等非专家标注方法的使用，标注数据获取的门槛和成本不断降低，但同时也带来了严重的标签噪声；另一方面，由于标注的主观性[3]、数据复杂[4]等原因，专家标注也存在一定程度的噪声。训练数据含噪，甚至是测试数据含噪，已经成为当下深度学习领域中的普遍问题和关键挑战。为了挖掘标签噪声数据潜在的训练价值，缓解标签噪声数据对模型训练产生的负面影响[5]，标签噪声学习[6]得到广泛关注并成为一个热点研究方向。

在实际的数据利用中，移动设备和物联网设备的普及贡献了全球半数以上的互联网流量[7]，为智能系统的应用发展提供了重要的数据基础。在过去的数年中，这些分散设备的数据通常直接传输到中心服务器，以中心式的训练范式获取特定的模型。但数据的集中往往面临着诸多困境，将大量分散设备的数据传输到中心服务器需要消耗大量的网络带宽和通信资源[8]，尤其在数据量较大的场景下将会导致通信成本显著增加。更重要的是，集中的数据存储或利用严重损害边端的隐私安全，会导致大量有价值的用户私有数据无法得到有效利用。因此，面对这些分散的数据，分布式处理利用成为当下一种有效的解决方案。而联邦学习[9]则是一种典型的隐私保护的分布式学习范式，其不分享原始数据，而是通过聚合模型参数或者梯度来共享节点知识。但是大部分联邦学习的研究均假设节点拥有正确标注的数据，从而限制了其实际应用的范围。因此，考虑到数据标注质量的联邦标签噪声学习[10-12]成为了适应当下高质量标注数据匮乏环境下的一种有效的分布式解决方案，有效权衡了数据标注成本、差错和模型训练成效。

现有未考虑标签噪声的联邦学习方法如果直接应用于噪声环境中，其性能表现往往会大幅降低，在高噪声率的场景下模型甚至无法收敛。而在中心式标签噪声学习中，不依赖于数据直接访问权的部分方法虽然能够较为便捷地集成在联邦框架内，但也面临分布式系统的固有异质性挑战，甚至相较于一般联邦学习方法的性能表现更差，因此需要探索更具有针对性的学习策略。现有联邦标签噪声学习方法可以分为基于噪声检测的训练方法、鲁棒的噪声容忍训练方法及混合方法。由于噪声场景的复杂性，单一的学习方法往往受到限制，而混合策略具有更强的噪声鲁棒性，代表了联邦标签噪声学习发展的趋势。

虽然对联邦标签噪声学习的研究已经取得了诸多成果，但面对极端数据异质、全局类不平衡、标签噪声异质、非随机噪声、多标签以及开集等复杂噪声场景，现有方法仍显不足。对此，本文对联邦标签噪声学习进行了系统性梳理，围绕联邦标签噪声学习的基本概念、挑战和发展前景，从模型训练的角度系统性地整理了现有的联邦标签噪声学习方法，比较和分析了不同方法的执行细节和特征，为研究者提供一种统一的分类框架和研究视角，为推动构建更加鲁棒的联邦标签噪声学习系统提供参考。

1 研究背景

1.1 联邦学习

联邦学习是一种隐私保护的分布式学习框架，允许分散的节点在不共享原始数据的前提下协同训练共享的模型。一方面，分布式架构可以有效打破“数据孤岛”的桎梏；另一方面，本地数据的不共享性赋予了联邦学习隐私友好的天然优势。在数据节点分散、隐私保护要求日益严格的时代环境下，联邦学习正在快速发展，其领域体系和行业标准被不断完善与统一。发展至今，联邦学

习的应用架构被极大丰富，根据与训节点网络拓扑结构的不同，可以将其划分为中心式联邦学习、去中心化联邦学习以及群簇联邦学习[13]。中心式联邦学习是联邦系统中最为普遍的拓扑架构，主要指具有一个中心服务器和多个客户端节点的传统联邦，本文仅关注中心式联邦学习。

联邦学习使用传播型的训练模式，训练以节点为单位，经过全局聚合形成新的全局模型，并进一步向与训节点或者客户端传播，不断迭代训练。联邦训练系统的数据总体D_g可以表示为

$$D_g=\bigcup_{k=1}^{K}D_k,\ D_k=\{(x_i^k,\ y_i^k)\mid i=1,\ 2,\ \cdots,\ N_k\} \tag{1}$$

式中：K为客户端总数；D_k为客户端k的本地数据；$(x_i^k,\ y_i^k)$为客户端k的第i个样本标签对，x_i^k为样本，y_i^k为对应标签；N_k为客户端k的本地数据数目。进一步可得全局的优化目标为

$$\begin{cases}\min S\left(F_1(\omega_1),\ F_2(\omega_2),\cdots,\ F_K(\omega_K)\right)\\ F_k(\omega_k)=E\left[l_k(y^k,f(x^k,\ \omega_k))\right],(x^k,\ y^k)\in D_k\end{cases} \tag{2}$$

式中：ω_k为客户端k的本地模型；$l_k(\cdot)$为客户端k的本地损失函数；$f(\cdot)$为模型对样本在标签空间的映射函数；$E(\cdot)$为模型在数据集上的经验损失；$S(\cdot)$为本地优化的联合函数。S一般通过全局聚合实现，例如使用FedAvg[9]的聚合，数学表达式为

$$\omega_g^{t+1}=\sum_{k=1}^{K}\frac{N_k}{\sum_{i=1}^{K}N_k}\omega_k^t \tag{3}$$

式中：t为全局通信的轮数；ω_g^{t+1}为第$t+1$轮的全局模型。

1.2 标签噪声学习

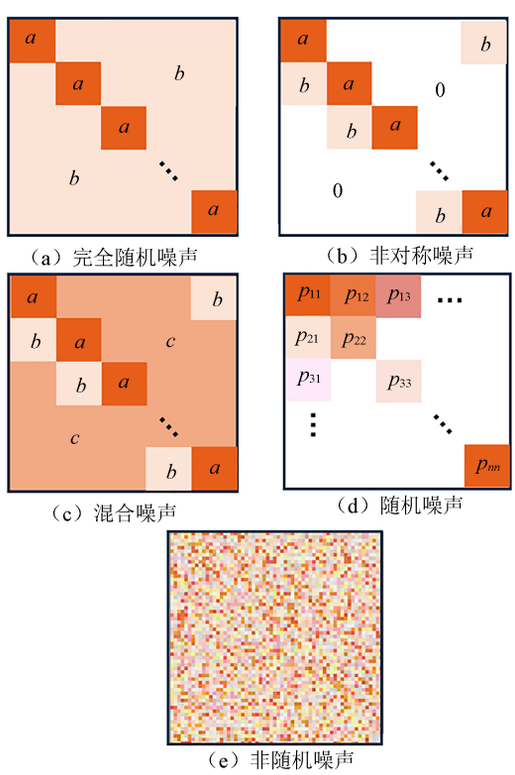
1.2.1 标签噪声概念、来源及负面影响

标签噪声主要是指在监督学习领域，和样本配对的标签与真实情况出现不一致的现象。标签噪声的产生存在多种因素，包括使用自动化标注系统[2]的非专家标注、专家标注的主观性[3]、数据自身的复杂性[4]等。例如，相比于一只猫的图片，医学CT图像具有更高的标注复杂性，对标注者的相关领域知识要求极高，因此更容易被标错。并且由于不同专家可能对同一图像的病变位置、大小或严重程度有不同的判断，所做的标注也存在差异，这在多类别问题上表现尤为明显。此外，恶意的标签翻转攻击会引入大量误导性强的标签噪声[14]，并且添加人工标签噪声也被用作一种隐私保护手段[15]。

标签噪声会干扰模型的训练进程，使其更加复杂且充满不确定性，从而显著降低模型的收敛速度，甚至会导致模型无法收敛[16]。尤其在深度神经网络中，由于其所具备的强大学习能力，往往容易过拟合训练集中的标签噪声，从而记忆这种复杂的数据分布[17]。因此，虽然模型在训练集上拟合程度较高，但在测试集上却表现一般，进一步限制了深度学习模型的泛化能力[3]。而在联邦环境下由于节点间数据异质和标签噪声异质的存在，模型在不同客户端上的更新可能会出现较大的差异，全局模型的性能会进一步受到影响[10]。

1.2.2 标签噪声分类

Frénay等[18]根据噪声标签的分布将其划分为完全随机噪声、随机噪声和非随机噪声。这种基于统计学的观点概括性强、易于理解，本文遵循相同的分类标准。图1展示了3种噪声模型的噪声转移矩阵，图1（e）为实例级，其余均为类级。



（a）完全随机噪声　（b）非对称噪声

（c）混合噪声　（d）随机噪声

（e）非随机噪声

图1 噪声转移矩阵示意

Fig. 1 Illustration of noise transition matrix

完全随机噪声的产生不依赖于任何类别或者样本特征。对于任意的噪声样本，其真实标签翻转为其他任意一个标签的概率相同。这种完全随机性，使得一个类别的样本被错误标记为另一个类别的概率与反向错误标记的概率相同，如图1（a）所示。因此，完全随机噪声也可以被称为对

称噪声。

随机噪声的产生依赖于其所处类别的特征。对于某一类别的任意噪声样本，其真实标签翻转为其他类别具有不同的概率，比如在医学胸部X光或CT图像中，由于影像特征相似，肺部结节可能被错误地标记为肺癌，但实际上可能是良性的肉芽肿或纤维瘤[19]。完全随机噪声或者对称噪声是随机噪声的一种特殊情况。如图1（b）所示，非对称噪声是指一个类别的样本被错误标记为另一个类别的概率与反向错误标记的概率不同[20]。如图1（c）所示，非对称噪声以及与对称噪声的混合情况同样属于随机噪声的范畴[21]。此外，如图1（d）所示，随机噪声还包含了更为复杂的转移矩阵概率分布。

非对称噪声的产生既依赖于样本所处类别，也受样本特征影响。对于某一类别的某个样本，其具有翻转为其他类别的某一样本的特定概率，比如由于个体毛色特征特殊，一张猫的图片可能被误分类为花豹。非对称噪声在实际应用中最为普遍，但也最为复杂，其噪声转移矩阵为实例级，如图1（e）所示。

1.2.3 标签噪声学习目标

标签噪声学习旨在有效处理标签噪声的负面影响，提高模型在真实数据上的泛化能力和准确性。一般的监督学习包含了由样本标签对组成的训练数据集合D，即

$$D=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\} \tag{4}$$

式中：N为样本总数目；x为样本；y为对应样本的标签。监督学习的目的在于找到由样本x到标签y的最佳映射函数。将y参数化为ω，监督学习目标可以表示为

$$\omega^*=\min E(f_\omega(x),y) \tag{5}$$

式中：f_ω将样本x向标签空间映射。在深度学习领域，f_ω可以通过深度神经网络训练得到。公式（5）代表了最小化经验损失的过程，即映射函数f_ω对样本x的映射结果与其对应标签y的距离最小。

在标签噪声存在时，训练数据集合$\tilde{D}$为

$$\tilde{D}=\{(x_1,\tilde{y}_1),(x_2,\tilde{y}_2),\cdots,(x_N,\tilde{y}_N)\} \tag{6}$$

式中：$\tilde{y}$为噪声标签。此时，通过最小化经验损失同样可以获取拟合训练集的模型，但却偏移了真实的数据分布，导致模型在测试集上的性能降低。标签噪声学习需要在标签噪声的干扰下学习到数据的真实分布，可以通过正则化项R实现，数学表达式为

$$\omega^*=\min[E(f_\omega(x),\tilde{y})+R] \tag{7}$$

1.3 联邦标签噪声学习

1.3.1 联邦标签噪声学习挑战

联邦标签噪声学习旨在分布式联邦框架下有效缓解客户端标签噪声的负面影响，提高全局模型的噪声鲁棒性和预测结果的准确性。联邦标签噪声学习结合隐私保护的联邦框架和训练数据含标签噪声的现实挑战，成为当下高质量标注数据匮乏环境下的一种有效的分布式学习方案，有效权衡了数据标注成本和模型训练成效，其面临的主要挑战如下。

1）标签噪声异质。由于客户端标注者之间往往具有不同的标注风格和经验水平，并且节点数据具有异质的特征，因此节点间噪声率或标签噪声模式也会存在一定的差异，这种现象被称为标签噪声异质[10]。异质的噪声率会导致通用的处理策略无法针对性地挖掘客户端知识。例如，低噪声率客户端上的标签矫正可能会引入额外的标签噪声，高噪声率的客户端上仅使用鲁棒优化而不矫正则无法充分挖掘噪声样本的潜在训练价值。另外，不同的噪声模式具有不同的特征，难以统一建模。例如，非对称标签噪声具有明显的标签误导性，对称标签噪声则是完全随机的干扰。这些差异使得单一的指标，如平均损失，在噪声客户端或者噪声样本的检测上失效，并且标签噪声异质会进一步影响模型的训练稳定性、准确性和泛化性等[22]，区分训练是一种常用的处理方法。

2）分布式系统中节点的异质性挑战以及与标签噪声负面影响的耦合。①数据异质性，即非独立同分布，将导致不同客户端的数据分布不一致，使得全局模型难以在所有客户端上达到最优解[23]。②由于客户端模型架构不同，直接聚合模型参数较为复杂，难以找到一个统一的全局模型来协调所有客户端的学习，并且节点的异构会导致本地模型对于噪声数据的学习拟合程度存在差异[24]。节点异构与标签噪声的耦合可能会进一步加剧训练偏差，降低协同学习的效率，影响模型泛化和模型训练精度。

3）与其他挑战的耦合，包括但不限于以下3个方面。①全局类不平衡[10]指某些类别的样本数量远多于其他类别。这种不平衡容易导致模型在训练过程中缺乏对少数类的学习，从而偏向于多数类。特征学习不足的少数类样本具有更大的损

失值，容易被检测为噪声样本，数据异质下具有更多少数类的客户端也容易被检测为噪声客户端，“小损失”原则面临失效的困境。②多标签问题[16]中，标签之间通常存在依赖关系，并且在多标签数据集中，不同类别的样本数量可能差异较大，这些会带来标签噪声与数据分布的复杂耦合，使得模型难以学习到真实的数据特征。③开集场景涉及未知类别的数据分布，需要模型能够识别并排除这些样本，同时避免对它们进行错误的分类。但开集噪声样本的损失值不一定符合“小损失原则”，模型对这些样本的识别变得更加困难[25]。

1.3.2 联邦标签噪声学习场景

区别于中心式训练，联邦框架下数据分散在不同的节点，且节点的噪声率往往未知。根据节点间标签噪声率、标签噪声类型的异同，联邦标签噪声学习包括标签噪声同质[26]和标签噪声异质[27]2种场景。这2种场景又可细分为是否含有干净节点或者高标注质量节点的情况，图2展示了几种不同的噪声场景。

在实际应用中，联邦节点间的噪声率往往呈现出异质的状态，是备受关注的噪声场景。相比之下，标签噪声同质的场景适用性较窄，且可以视为标签噪声异质的特殊情况。

1.3.3 联邦标签噪声训练数据生成

标签噪声数据生成是联邦标签噪声学习系统训练的基础和前提，主要包括基于人为注入的生成方法、人类标注错误的噪声数据及真实世界的噪声数据。对于基于人为注入的生成方法，噪声注入与数据划分的顺序并不固定。在客户端具有相同噪声特征的场景中，噪声注入可以在数据划分前进行。但一般考虑客户端具有不同的噪声特征，噪声注入应当在数据划分后，并对不同的客户端分别执行。对于人类标注错误和真实世界的噪声环境，数据中的噪声已经存在，只需要执行数据划分。而数据分布式划分的方式包括同质与异质2种情况。其中，异质的情况最为普遍，一般使用Dirichlet分布实现，可以通过调整Dirichlet分布参数来控制非独立同分布的程度。此外，病态非独立同分布划分将数据集划分为多个只包含部分类别的子集，其主要用于模拟极端的非独立同分布场景。

基于人为注入的生成方法通过人为控制监督数据集样本标签的翻转来形成噪声数据，主要包括基于类别添加的噪声和基于实例添加的噪声。1）基于类别添加的噪声普遍使用对称噪声、非对称噪声和混合噪声，一般基于噪声转移矩阵添加。为了保证设定的噪声率与实际添加的噪声率相同，标签翻转时一般先排除选定噪声样本的原有标签。对称噪声将原有标签等概率翻转为其他任何标签，而非对称噪声一般将原有标签翻转为特定其他标签，比如原类别的上一个类别[27]；混合噪声将对称噪声与非对称噪声进行混合，具有复杂的噪声转移矩阵[21]。2）基于实例添加的噪声可以通过深度神经网络的预测误差为数据集中的每个实例分配噪声标签。如Chen等[28]通过计算每个实例在训练过程中的预测概率分布，选择高噪声分数的实例并将其标签替换为预测概率最高的标签。Wu等[10]为每个噪声客户端训练一个本地神经网络并计算其对本地数据的分类概率，将高误差分类概率样本的标签替换为预测概率较高的错误标签。

人类标注错误的噪声数据是指在数据标注过程中由于人为因素导致的标签不准确。如CIFAR-

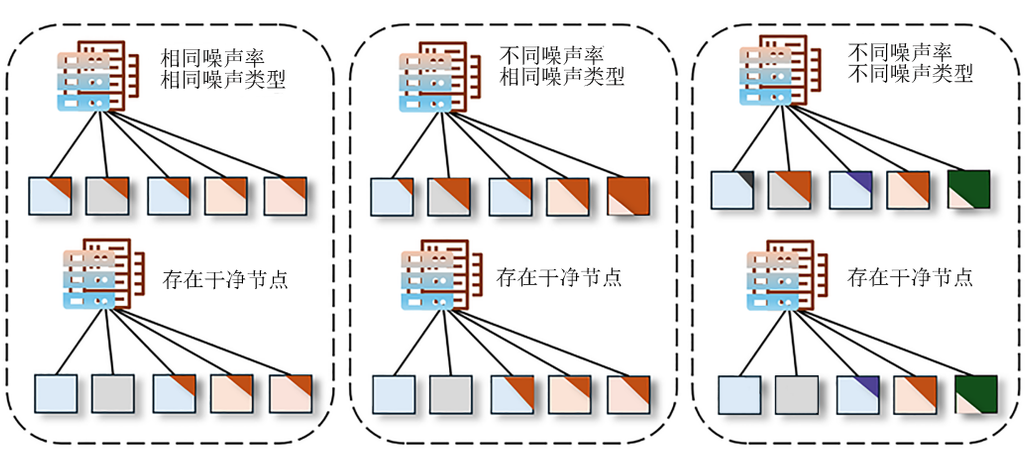


图2 联邦标签噪声学习场景

Fig. 2 Scenes of federated noisy label learning（FNLL）

10N、CIFAR-100N[29] 分别含有 10 个、100 个类别，均包含 50 000 张含噪的训练图像与 10 000 张干净的测试图像。其中，CIFAR-10N-aggregate 通过多数投票生成，若标签不一致则随机选择，噪声率为 9.03%；CIFAR-10N-random 随机选择单个标注者的标签，噪声率约为 18%；CIFAR-10N-worst 则在错误标签中随机选择[29]。CIFAR-100N-coarse 通过将 CIFAR-100 的 100 个精细类别重新分组为 20 个粗略类别，并收集人类对这些粗略类别的标注，噪声率为 25.60%；CIFAR-100N-fine 则直接收集人类对 CIFAR-100 原始的 100 个精细类别的标注，噪声率为 40.20%[29]。ANIMAL-10N[30] 包含 10 类动物的 55 000 张图像，分别为 50 000 张训练集图像和 5 000 张测试图像，模型验证的噪声率约为 8%。Red Mini-ImageNet[31] 拥有 100 个类别的 55 000 张图像，包括 50 000 个训练样本和 5 000 个测试样本，噪声率为 0～80%。

真实世界的噪声数据源于实际应用场景（如网络爬取数据、传感器采集数据等）中自然产生的标签噪声。如 Clothing1M[32] 包含 14 个类别的 100 万张服装图像，从多个在线购物网站收集，评估噪声率约为 38%。该数据集还分别包含 50 000、14 000 和 10 000 张带有干净标签的图像，用于训练、验证和测试。Food-101N[33] 包含 101 个类别的 101 000 张食物图像，噪声率约为 20%。WebVision[34] 是由 Google 和 Flickr 网络图像构建的大规模视觉数据集，拥有 1 000 个类别约 240 万张爬取图像，包括验证集和测试集各 5 万张人工精标图像，其噪声来源于网络文本描述与图像内容的语义偏差以及用户上传标签错误，包括约 5% 的类内噪声和 25% 的类外噪声。

2 联邦标签噪声学习方法

根据是否使用显式的标签噪声检测机制，现有联邦标签噪声学习方法可分为基于噪声检测的训练方法、鲁棒的噪声容忍训练方法和混合方法，图 3 展示了联邦标签噪声学习方法的分类细节。

2.1 基于噪声检测的训练方法

基于噪声检测的训练方法旨在识别出标签噪声样本，进而处理其负面影响。在联邦框架下，由于分散的客户端一般具有相异的数据质量，在

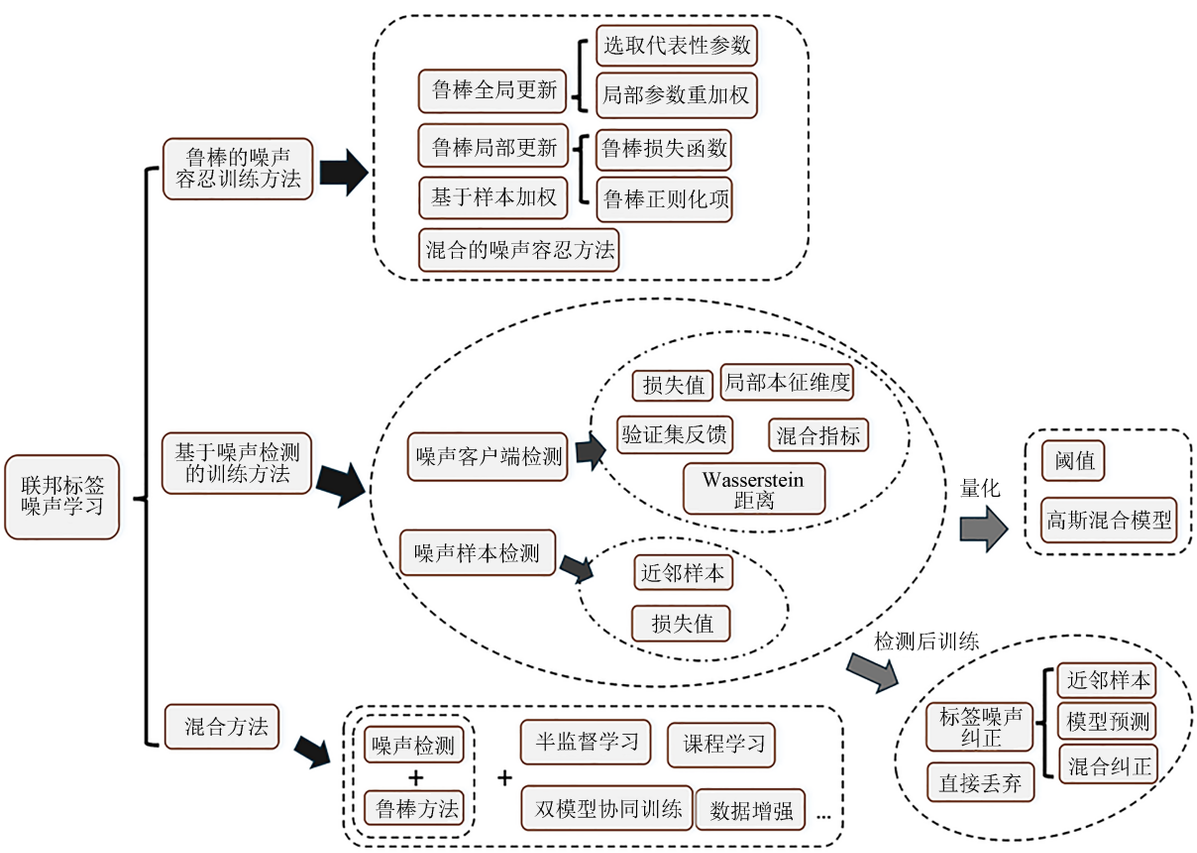


图 3 联邦标签噪声学习方法分类

Fig. 3 Classification of FNLL methods

进行标签噪声样本检测之前可以将其区分为干净客户端与噪声客户端，或者低噪声率客户端与高噪声率客户端。因此，基于噪声检测训练方法的流程一般包括噪声客户端检测、噪声样本检测以及检测后训练。图4展示了基于噪声检测训练方法的具体流程。

2.1.1 噪声客户端检测

对于高质量标注节点，其往往拥有大量干净或者低噪声率的数据，对于训练具有重要的引导作用。为防止高质量节点的数据被丢弃或被纠正而引入额外的标签噪声，需对其进行识别并与低质量节点进行区分训练。常见的噪声客户端或者高质量标注客户端的识别标准包括局部本征维度[35-37]、损失值[10,38]、Wasserstein距离[39]、验证集反馈[40-41]和混合指标[42]等。

局部本征维度是一种衡量高维数据在局部邻域内有效复杂度的指标。它基于数据点的邻域结构，通过计算目标数据点与其最近邻点之间的距离变化率来评估数据的局部复杂度。由于标签噪声往往会破坏数据的邻域结构，导致局部邻域内数据点分布更加复杂和不均匀，因此噪声数据一般会带来更高的局部本征维度分数[36]。

基于损失值的检测方法依托“小损失值”[43]假设。具体来讲，相比于噪声样本，深度神经网络在训练早期更容易拟合干净样本，因此将干净样本输入模型中的损失值更小。从客户端的角度出发，考虑到类别分布的差异，存在多种损失量化方法。例如，FedDC[42]使用客户端中所有样本的平均损失和作为评价指标，FedNoRo[10]通过计算客户端的类平均损失向量来去除全局类别数目不平衡的干扰。

Wasserstein距离是一种基于最优传输理论的概率分布度量工具，能够量化2个分布之间的差异，不仅考虑了概率质量的分布，还捕捉了数据点之间的几何关系[39]。通过Wasserstein距离可以计算目标客户端的数据分布与全局目标分布之间的距离，而噪声客户端的模型一般偏离全局模型的程度更大，对全局模型的贡献较小[39]。

基于验证集反馈的方法通过本地模型在服务器干净验证集上的表现来评估客户端的噪声程度。例如，FedPPO[40]在服务器端首先通过主成分分析降维模型参数，之后进行层次聚类并选择验证集上平均准确率最低的簇作为潜在的噪声客户端集合。FedMW-CSS[41]在服务器端验证集上测试本地模型对全局模型的损失优化效果，并将优化效果较差的节点识别为噪声客户端。

混合指标通过不同程度地结合单一检测指标来综合评估客户端的数据质量，可以更全面地识别噪声客户端。例如，基于损失值的检测方法能够有效检测新加入的客户端，而局部本征维度更

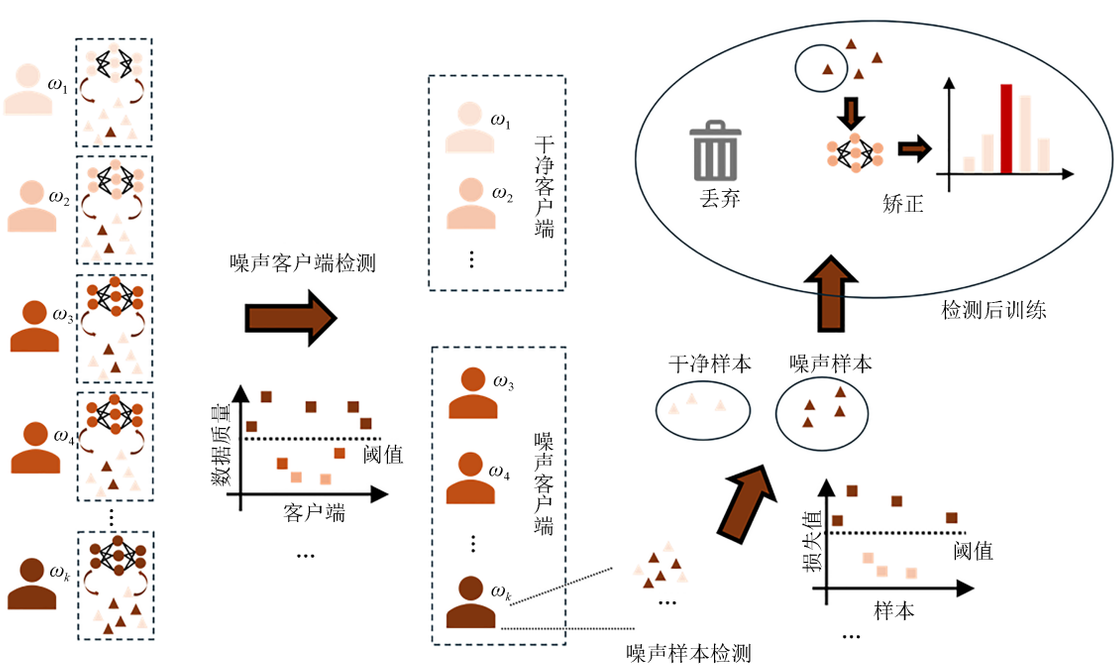


图4 基于噪声检测训练方法的流程

Fig. 4 Procedure of the training method based on noise detection

适合区分现有的客户端。因此，FedDC[42]结合这2种方法，提高了动态环境中噪声客户端检测的准确性。此外，Mishra等[44]通过比较预测标签与真实标签的一致性来筛选出可能的干净样本，进而计算噪声比例，最终根据估计的噪声比例，将所有客户端划分为若干组。

在量化客户端的含噪程度之后，需要进一步进行数值区分，主要采用设置阈值和混合高斯模型拟合的方法。阈值包括静态阈值与动态阈值，高于阈值则被识别为噪声客户端。而混合高斯模型是一种数据分布的概率模型，其假设数据点由多个高斯分布生成，在拟合量化数值的分布之后，具有较大均值的组分对应噪声客户端。FedDC[42]、TrustBCFL[37]、FedPPO[40]设定了区分噪声客户端和干净客户端的静态阈值，FedMIN[45]通过局部损失拟合高斯混合模型和全局聚合的高斯混合模型的距离实现了动态阈值设定，FedNoRo[10]、FedA3I[22]、DivideMix[46]、FedCNL[47]、DualOptim[38]等使用混合高斯模型来对噪声客户端进行区分。

以上5种噪声客户端的检测方法各有优劣。局部本征维度法的计算简单，但反复的距离计算将会耗费更多的计算资源；损失值直接来自模型的训练过程，不会增加额外的计算负担，但影响样本损失值变化的因素较多，如样本学习的难易程度、样本数量、数据的非独立同分布程度等，因此基于损失值的检测方法容易受到外界干扰；Wasserstein距离因其具有良好的数学性质，在处理概率分布时更加稳定可靠，但需要解决最优传输问题，计算复杂度较高，尤其是在大规模数据集中；验证集反馈直接反映了客户端数据对模型性能的影响，评估结果较为准确，但高质量验证集在实际中不容易被满足，并且需要保证其与训练集的数据分布较为接近；混合指标具有更广泛的适用范围和更强的检测鲁棒性，但不同指标间的协调处理会增加算法的实现难度与计算负担。在实际应用中，噪声客户端检测方法的选择需根据具体的学习场景、数据分布特征以及客户端计算资源限制等因素进行综合考虑。

2.1.2 噪声样本检测

对检测出的噪声客户端，需进一步检测其所包含的标签噪声样本，基于近邻样本[48-49]、损失值[43, 45, 47, 50]的方法被广泛使用，2种典型的标签噪声样本检测方法如图5所示。

基于近邻样本的检测方法依托同类别样本在特征空间中具有较高相似性[49]的假设，通过近邻样本的标签一致性检测标签噪声。Chen等[49]通过近邻样本标签的数目计算以及2个成分高斯混合模型来估计样本干净的概率。但基于近邻样本的检测方法普遍依赖高质量数据，在联邦框架下，尤其在面临数据异质的挑战时，节点的本地样本数目可能不足以提供可靠的近邻支撑，且进行节点间样本或者原型传播时可能会带来隐私泄露的风险。Jiang等[48]将与近邻样本多数标签不一致的目标样本视为噪声样本，通过计算样本标签与其近邻标签分布之间的Jensen-Shannon散度，选择与近邻标签分布具有高度一致性的样本作为干净样本。

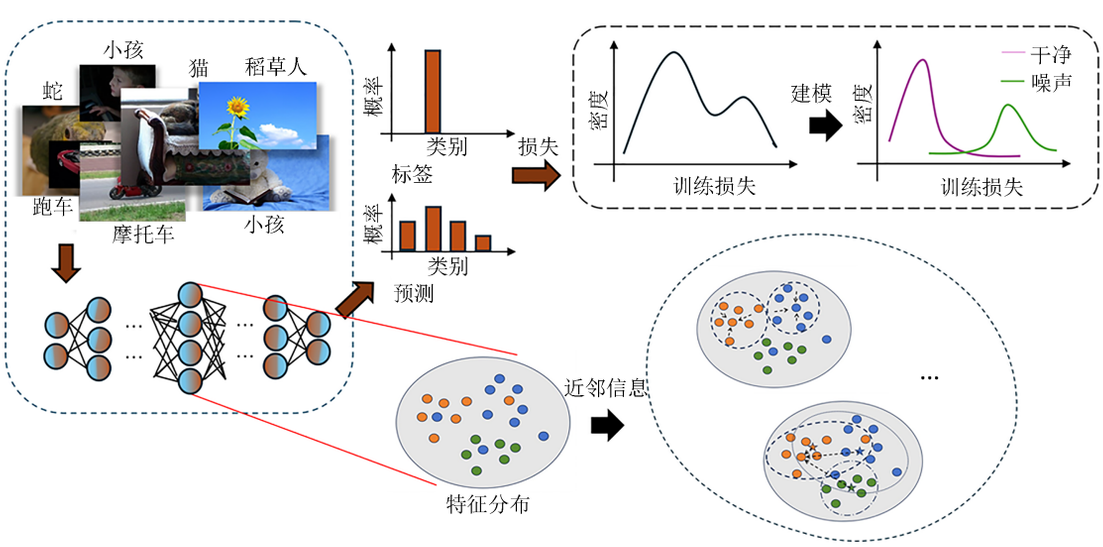


图 5 典型的标签噪声样本检测方法

Fig. 5 Typical detection methods of noisy label samples

基于损失值的标签噪声样本检测与基于损失值的噪声客户端检测一致，但并不存在复杂的量化形式，同样可通过设定阈值[50]和高斯混合模型[40, 45, 47]进行区分。其中，高斯混合模型中具有较大均值的成分对应噪声样本。但是小损失原则需要满足噪声转移矩阵对角占优，即对角线元素大于非对角线元素的最大值。从样本学习的角度而言，困难样本[51]和尾部样本[49]同样具有较大的损失值，对于数据长尾分布或者不平衡的场景，该原则需要针对性地进行优化。

此外，FedBary[39]通过Wasserstein距离的对偶问题来计算每个样本对Wasserstein距离的贡献梯度。负梯度值意味着将更多概率质量分配给该样本会减小客户端数据分布与目标分布之间的距离，正梯度值反之。因此，负梯度值样本具有更高的学习价值，而正梯度值样本可能为噪声样本。

2.1.3 检测后训练方法

对于检测出的标签噪声样本，进一步处理的方法包括直接丢弃与标签矫正。

标签噪声样本直接丢弃是最为简单易行的处理策略，仅使用检测出的干净样本训练能够直接去除标签噪声的干扰。例如，FedRN[52]基于其他客户端的数据质量和分布相似程度来定义目标客户端的k-可靠邻居，并通过高斯混合模型区分干净样本，然后在训练时排除噪声样本，但k-可靠邻居数目调整的灵活性较差。Co-teaching [53]和Co-teaching+[54]可以通过FedAvg[9]直接集成在联邦框架下，Co-teaching[53]同时训练2个深度神经网络，每个网络仅使用另一个网络筛选出的干净数据进行训练，但随着训练的进行，双网络会逐渐收敛到相似的状态，导致退化为单网络的训练。为此，Co-teaching+[54]在前者的基础上引入了“基于分歧更新”的策略。FedPPO[40]使用强化学习中的近端策略优化算法，联合评估客户端在全局和本地模型上的准确率来计算客户端贡献，并将其作为奖励信号，动态选择高贡献客户端参与训练。但该方法的实现复杂度较高，并且噪声率阈值的不准确估计很容易影响到模型的最终性能。Tuor等[55]通过训练一个基准模型来评估客户端之间数据的相关性，利用样本的损失值来确定其与基准模型的匹配程度，并通过计算阈值来过滤掉损失值较高的样本，但基准数据在实际中可能难以获取。

标签噪声矫正通过对检测出的噪声样本重新标注以进一步挖掘其潜在的训练价值，在高噪声率场景下的优势尤为明显[27]。需要注意的是，标签噪声矫正不一定完全依靠标签噪声样本检测，对于检测出的高噪声率客户端所有样本进行统一纠正是一种简单易行的处理方法[27]。大部分的矫正方法使用了信息量更充足的软标签，矫正方法包括基于近邻样本[48]、模型预测[36–37, 50, 56]以及混合策略[49]。基于近邻样本的矫正方法与检测方法一致，如FRCS[48]通过最近邻的多数标签来直接覆盖目标样本的标签。基于模型预测的矫正方法是指通过全局模型或本地模型的预测结果来辅助噪声样本标签的矫正，包括直接覆盖[56]、阈值法[36–37, 50]和基于标签可靠性量化的软标签矫正[49]，也存在多阶段逐步矫正的混合方案。例如，Yang等[56]直接使用全局模型预测覆盖噪声样本的标签，FedCorr[36]通过设定选择比例和置信度阈值来控制噪声标签的矫正。相比于单一的矫正方法，混合策略能够提供更加鲁棒的标签估计，在处理复杂噪声场景时更具优势。例如，Chen等[49]针对不平衡子群体，使用混合高斯模型拟合样本的近邻标签一致数目，将拟合概率作为目标样本的标签置信度，并结合原标签和预测的伪标签作为翻转标签。

2.2 鲁棒的噪声容忍训练方法

鲁棒的噪声容忍训练方法对正确标注样本与标签噪声样本不作训练上的显式区分，通过设计能够抵抗标签噪声干扰的训练策略来提升训练系统的标签噪声容忍度，可能涉及噪声客户端检测以进行针对性的区分训练。鲁棒的噪声容忍训练方法主要包括鲁棒全局更新、鲁棒局部更新、基于样本加权以及混合的方法。图6展示了不同的鲁棒噪声容忍训练方法在模型训练中的关联性。

2.2.1 鲁棒全局更新的方法

鲁棒全局更新是一种分布式噪声学习系统特有的鲁棒训练策略，旨在通过设计噪声容忍的聚合方法，有效降低标签噪声所引起的不良局部模型对全局模型聚合更新的干扰，同时增强低噪声率节点对于全局模型的贡献。鲁棒全局更新主要包括2种模式，分别为选取代表性参数和局部参数重加权。

选取代表性参数的方法是指在全局模型更新中仅选择具有一般代表性的局部参数用以全局模型更新。例如，Ditto[57]使用客户端模型的中位数来获取全局模型，Krum[58]选取与大多数正确节点最接近的梯度估计向量作为代表性局部参数。但上述方法虽然能够有效避免极端参数的负面影响，

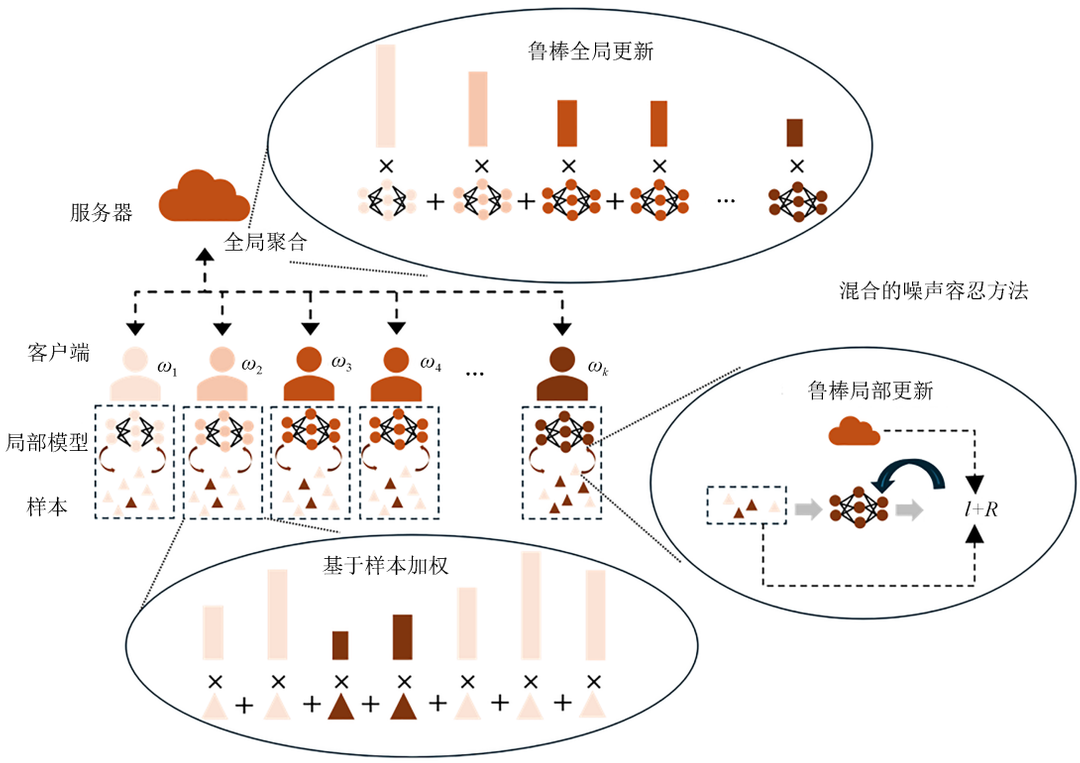


图 6　鲁棒的噪声容忍训练方法在联邦系统中的架构

Fig. 6　Architecture of robust noise-tolerant training methods in federated systems

但容易丢失大量的训练信息，导致全局模型的更新出现较大程度的偏移。而 Yin 等[59]首先去除所选客户模型中每个参数的最大值和最小值，并计算剩余参数的平均值作为全局模型。该方法虽然保证了训练信息的完整性，但其适用性较差。局部参数重加权法通过降低高噪声率客户端本地参数的权重，同时增强低噪声率客户端以及干净客户端本地参数的权重来缓解不利于本地参数的负面影响，并挖掘干净数据的训练价值。FedNoRo[10]引入一种基于模型参数感知的鲁棒训练策略，将客户端划分为高噪声率客户端和低噪声率客户端2组，并计算噪声率较高一组客户端模型参数对噪声率较低一组客户端模型参数的距离，选取最大值和最小值并进一步归一化来降低高噪声率客户端的权重。但该方法忽视了客户端之间的差异化，导致低噪声率客户端内部仍含有噪声率较高的客户端，但它们的权重却没有被减小。相比之下，RHFL[24]引入客户端置信度重加权的动态鲁棒聚合机制，综合评估客户端的标签质量和学习效率来计算客户端置信度，并进一步缩放权重。但该方法依赖于公共数据集来对齐不同客户端的知识分布，要求公共数据集与客户端的私有数据分布接近，并且该方法对超参数的敏感性较强，置信度对权重影响程度的控制参数需要进行适当调整。

2.2.2　鲁棒局部更新的方法

鲁棒局部更新通过在本地模型更新的过程中添加标签噪声鲁棒约束来减小标签噪声的影响，并且可以与噪声客户端检测结合以形成针对性更强的标签噪声鲁棒学习系统。具体约束方法包括添加标签噪声鲁棒的正则化项或使用标签噪声鲁棒的损失函数。

损失函数本质上是用于衡量模型预测值与真实值差异的一种表现形式，确定了模型训练过程中需要最小化的目标。传统的损失函数，如交叉熵损失函数，容易导致模型过拟合到噪声标签。而标签噪声鲁棒的损失函数通过引入特殊的机制来降低模型对于标签噪声的敏感性，并增强其对真实数据分布的学习。如 MAE[60]计算模型预测概率与样本真实标签分布之间的平均绝对差值，其满足对称性条件，能够容忍均匀噪声存在，但该方法的计算速度在大规模数据集上可能较为缓慢。GCE[61]是对 MAE 的泛化，通过调整特定超参数来获取标签噪声鲁棒和学习特征之间的动态平衡，取得了更好的学习效果。而 SCE[51]结合了传统的交叉熵损失和标签噪声鲁棒的逆交叉熵损失，具有一定标签噪声鲁棒性，同时也能促进模型对困

难类别的学习，扩展了之前的对称性思想，同时处理了欠拟合和过拟合的问题。

正则化项则是在损失函数的基础上，为了防止模型过拟合而添加的额外约束项。该方法通过特殊的机制约束模型更新方向，以防止其偏移真实的数据分布。正则化项的来源具有多种方式，通过模型预测的约束是一类较为普遍的方法。例如，FedELR[62]引入一种早期学习正则化的方式，通过最大化模型输出与早期预测之间的内积，强制模型在训练初期保持较高的准确性，但该方法也容易受到早期模型性能的影响。而FLR[5]通过结合全局模型平均预测和本地模型平均预测来生成伪标签，并通过进一步限制模型预测与伪标签之间的距离来防止噪声过拟合。另外，正则化项的控制参数可以进行调整以适应不同的训练阶段和噪声水平，但同时也带来了相应的计算负担。基于数据的约束方法同样也被广泛使用，如FedLSR[63]使用自蒸馏技术，利用数据增强的样本对训练模型，通过最小化原始样本和增强样本之间的输出差异来约束模型更新方向。此外，FedLNL[64]包含一种新的LDP正则化器，通过最大化分类器输出的多样性来提高模型的鲁棒性。NRL[44]引入噪声标签的对数损失，衡量模型预测为非噪声标签的概率与真实标签为非噪声标签概率之间的差异，进而限制模型对噪声标签的高置信度预测。

2.2.3 基于样本加权的方法

基于样本加权的方法关注于样本自身对训练的贡献或价值，通过为每个训练样本分配权重，动态调整训练过程以缓解标签噪声样本的负面影响，并使模型更加关注到正确标注的样本。例如，MentorNet[65]是一种辅助神经网络，通过学习数据驱动的课程方案来动态调整训练样本的权重，进而指导主网络训练。该方法适用于多种噪声类型，且不需要额外的干净数据。然而，该方法计算复杂度相对较高，对超参数较为敏感。Ren等[66]将样本权重调整转化为元学习问题，通过在小的干净验证集上评估模型性能来动态调整样本权重。具体来说，首先计算样本对验证集的损失梯度影响，然后通过最小化验证集损失来动态调整样本权重。这种方法不需要额外的超参数调整，简化了训练过程，但依赖于验证集的数据质量，因此在实际应用中可能受限。FedMW-CSS[41]利用服务器干净的验证集来计算全局模型的损失梯度，并反馈给客户端，客户端根据本地样本与验证集梯度的一致性来动态调整样本权重。

2.2.4 混合的噪声容忍方法

如图6所示，上述3种鲁棒的噪声容忍训练方法分别关注于不同的方面，并利用不同的机制来非显式地增强联邦标签噪声学习系统的噪声鲁棒性[10, 24, 38, 67-68]。单一方法具有较高的灵活性和集成性，而混合的噪声容忍方法能够结合每种方法各自的优点，从多个方面进行协同训练，从而获得更加鲁棒的联邦标签噪声学习系统。例如，RH-FL[24]使用一个公共数据集来对齐模型反馈，通过结合噪声容忍的对称交叉熵损失与客户端置信度重加权的鲁棒全局模型更新来提高系统学习的噪声鲁棒性，该方法适用于模型异质或同质的场景。RHFL+[67]进一步结合动态标签细化来减轻本地噪声样本对模型训练的负面影响。FedNoRo[10]通过类平均损失和混合高斯模型识别噪声客户端，并针对检测出的噪声客户端，结合普通交叉熵与预测蒸馏损失来对其进行联合优化，提出一种基于模型最短距离的鲁棒聚合策略。HDHRFL[68]通过计算客户端损失函数变化来评估其噪声程度，进而动态调整客户端权重，并通过公共数据集对不同架构的模型进行知识对齐以及使用SCE损失执行本地模型更新。

2.3 混合方法

面对复杂的标签噪声场景，单一的处理策略可能无法保证训练的稳定以及模型预测的精度，噪声检测、鲁棒策略的多机制融合成为更具针对性且可靠的选择，代表了联邦标签噪声学习发展的趋势。例如，FedELC[27]采用FedNoRo[10]的噪声客户端检测方法和鲁棒聚合策略，同时对于检测出的噪声客户端提出了一种新颖的端到端标签校正方法，通过联合优化分类损失、兼容性正则化损失和熵正则化损失来同时更新模型和样本软标签分布，最终获取到接近真实分布的软标签。但是该方法对于噪声客户端的区分较为粗糙，尤其在客户端噪声含量高度异质的场景中，容易弱化训练的针对性，难以发挥这些异质客户端各自的训练价值。而Yang等[56]通过全局平均类质心作为额外的全局监督来规范局部训练，并利用损失函数值的大小来估计样本的可靠性，同时使用全局模型的预测来辅助矫正低置信度样本的标签，但由于类质心携带了本地敏感信息，可能不利于客户端的隐私保护。Aorta[50]主要针对联邦设备异

构场景，通过标签校准、数据生成和质量感知的模型聚合，显著提高了模型的准确性和收敛速度。Chen 等[49]针对不平衡子群体存在的场景，提出一种基于样本标签噪声概率的软标签翻转方法。该方法能够更准确地处理标签噪声和不平衡子群体中的样本，避免硬标签翻转带来的信息丢失，并结合强-弱数据增强、协同训练和分布鲁棒优化，从多个方面提升模型在极端情况下的泛化能力，具有很强的噪声标签鲁棒性。

此外，在具体的训练过程中，半监督学习思路、课程学习、数据增强、双模型协同训练等方法也广被使用。半监督学习旨在充分利用有限的标注数据和大量未标注的数据来对模型进行训练，在降低对标注数据依赖的同时，提高模型的性能和泛化能力，适用于标注成本高或标注数据稀缺的场景，如 FRCS[48]与 DivideMix[46]算法。课程学习根据设定的样本复杂度，从简单的干净样本开始，逐步过渡到复杂的噪声样本，能够有效避免训练早期标签噪声对模型的干扰，使其能够快速收敛，更好地泛化到未训练的数据，减少对复杂任务中噪声数据的过拟合，如 FedCNL[47]。数据增强[46, 49]旨在生成输入数据的变体来增加训练数据的多样性，使模型能够从不同的角度理解样本，防止被噪声标签误导，有效提升了训练的鲁棒性和泛化性。双模型协同训练优化 2 个具有同样结构的神经网络，通过相异的初始化、数据分配等来保持网络的差异性，使得双网络之间相互监督和纠正，可以有效避免单一网络在训练过程中累积的错误，并且可以通过合理的设计来有效提高训练效率。

2.4 方法对比与分析

图 7 展示了基于噪声检测的训练方法、鲁棒的噪声容忍训练方法和混合方法之间的训练架构关系，不同类型的方法之间存在一定程度的交叉，每种方法关注的角度不同，但相互之间具有互补性。表 1 从训练的不同角度出发，更加清晰地对比几种具有代表性的联邦标签噪声学习方法的异同。

Jiang 等[27]在统一场景下对不同类型的联邦标签噪声学习方法进行了测试，包括 2 种数据非独立同分布、4 种标签噪声和多种噪声数据集。不同方法在人为注入标签噪声的合成数据集上的精确度如表 2 所示。其中，客户端的总数为 100，标签噪声的生成是通过噪声率范围为每个客户端添加噪声率线性增长的标签噪声。

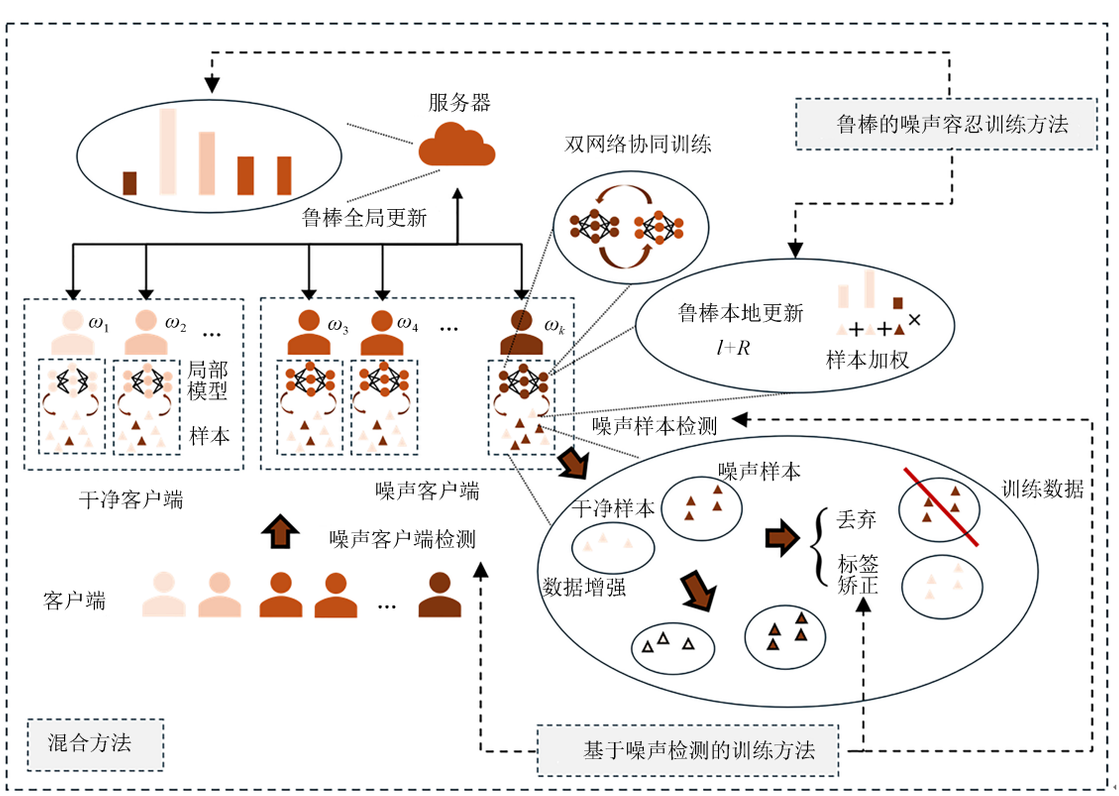


图 7 联邦标签噪声学习方法架构

Fig. 7 Architecture of FNLL methods

表1 代表性联邦标签噪声学习方法比较

Table 1 Comparison of representative FNLL methods

方法	鲁棒全局更新	鲁棒局部更新	样本加权	噪声客户端检测	噪声样本检测	噪声样本丢弃	标签噪声矫正	其他策略
鲁棒的噪声容忍训练方法								
TrimmedMean[59]	○	√	○	○	○	○	○	○
Krum[58]	○	√	○	○	○	○	○	○
Median[57]	○	√	○	○	○	○	○	○
Symmetric CE[51]	○	√	○	○	○	○	○	○
FedELR[62]	√	√	○	○	○	○	○	○
FedLSR[63]	○	√	○	○	○	○	○	数据增强
FedLNL[64]	√	√	○	○	○	○	○	○
FedMW-CSS[41]	√	○	√	○	○	○	○	○
FedPLL-LAAR[69]	√	√	○	○	○	○	○	数据增强
FedNoRo[10]	√	√	○	√	○	○	○	○
基于噪声检测的训练方法								
FedRN[52]	√	√	○	√	○	○	○	○
混合方法								
Co-teaching[53]	○	○	○	○	√	√	○	双模型协同训练
Co-teaching+[54]	○	○	○	○	√	√	○	双模型协同训练
JointOptim[70]	○	√	○	○	√	○	√	○
DivideMix[46]	○	√	○	○	√	○	√	半监督学习方法
RobustFL[56]	○	√	○	○	√	√	○	○
FedCorr[36]	○	√	○	√	√	○	√	三阶段训练框架
Aorta[50]	√	○	○	√	√	○	√	数据增强、分布感知的数据采样策略
FedELC[27]	√	√	○	√	○	○	√	数据增强
FedCNL[47]	○	○	○	√	√	○	√	课程学习、数据增强

注：“√”表示具有该模块，“○”反之。

由表1和表2可知，在场景1和场景5中，Symmetric CE的预测精确度与FedELC非常接近，但Symmetric CE模型却十分简单，仅仅使用了一种标签噪声鲁棒的损失函数，而FedELC作为一种混合方法使用了鲁棒全局更新、鲁棒局部更新、标签噪声矫正等多种优化策略，并且其在场景2和场景6中的预测精确度还要低于Symmetric CE。另外，使用简单鲁棒全局更新的TrimmedMean、Median在场景1、3、4等多个场景下的分类性能表现反而优于更为复杂的Co-teaching+、JointOptim和DivideMix等标签噪声学习方法。另外，即使使用单一鲁棒全局更新的TrimmedMean、Krum和Median在不同的场景下也具有较大的性能差异。因此，虽然不同方法之间存在一定程度的交叉与互补，但由于不同机制间的耦合作用非常复杂，即使联合多种机制共同优化，也不能保证该方法最终能够取得较好的性能表现。不同方法的取舍还需要考虑包括分类性能在内的多方面因素，如考虑方法的复杂性和灵活性等。例如，面对数据集复杂度的上升，部分方法出现了异常的性能降低，如FedLSR与RobustFL，表明这类方法相对不适用于多类别的复杂数据场景中。

表 2 不同方法在CIFAR-10和CIFAR-100数据集人工注入噪声标签上的精确度[27]

Table 2 Precisions of different methods on CIFAR-10 and CIFAR-100 with manually-injected noisy labels[27] %

数据异质	Dirichlet(1.0)								Dirichlet(0.5)							
噪声类型	对称噪声		对称噪声		非对称噪声		混合噪声		对称噪声		对称噪声		非对称噪声		混合噪声	
噪声率	0～0.4		0～0.8		0～0.4		0～0.4		0～0.4		0～0.8		0～0.4		0～0.4	
场景编号	1	9	2	10	3	11	4	12	5	13	6	14	7	15	8	16
数据集	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
鲁棒的噪声容忍训练方法																
TrimmedMean[59]	71.21	41.44	47.93	28.33	73.06	44.46	70.77	41.48	68.94	41.56	49.97	22.95	69.44	44.16	71.86	42.45
Krum[58]	51.90	22.99	39.60	17.36	48.29	24.88	43.28	24.32	33.49	14.40	25.19	7.95	18.54	13.12	25.12	12.29
Median[57]	72.52	43.90	58.45	31.01	75.56	43.20	73.60	43.44	68.45	44.59	52.30	30.05	72.06	50.47	72.09	40.31
Symmetric CE[51]	76.32	42.35	**70.96**	32.11	74.46	42.44	76.13	42.41	72.05	40.99	**62.39**	31.64	70.93	40.90	71.48	42.05
FedLSR[63]	75.16	3.08	70.42	2.80	75.11	3.14	73.34	3.55	64.20	3.84	62.35	3.18	66.46	3.56	60.04	3.66
FedNoRo[10]	73.67	44.70	66.50	32.98	77.88	44.51	75.38	43.32	71.09	45.09	59.11	**33.79**	73.63	43.97	73.51	43.54
基于噪声检测的训练方法																
FedRN[52]	72.65	32.48	62.89	26.35	69.02	35.22	72.23	35.72	58.72	38.04	54.47	29.78	54.14	36.73	55.76	34.39
混合方法																
Co-teaching[53]	75.79	44.74	57.30	32.72	76.88	44.78	75.37	**45.64**	72.23	**47.83**	48.23	30.04	74.95	45.52	74.22	**45.37**
Co-teaching+[54]	69.70	31.03	64.71	24.09	67.60	31.35	63.70	31.91	58.09	28.77	57.01	22.80	46.79	30.35	55.72	30.26
JointOptim[70]	64.74	27.66	59.78	22.67	65.02	27.84	64.53	27.83	57.76	26.39	52.44	21.95	55.21	26.20	57.84	26.30
DivideMix[46]	63.28	37.42	60.19	30.74	65.36	36.78	66.78	38.59	57.36	37.23	55.09	31.79	63.03	36.98	62.23	36.86
RobustFL[56]	63.17	17.41	51.34	3.73	63.26	14.67	62.65	17.59	54.05	11.23	55.45	5.03	54.08	10.50	53.94	13.54
SELFIE[30]	76.70	44.62	65.30	32.66	77.84	**44.90**	77.55	42.77	72.60	45.06	60.08	31.82	73.36	**46.36**	74.26	42.20
FedELC[27]	**76.81**	**44.83**	68.22	**33.07**	**77.97**	44.01	**77.98**	43.87	**73.13**	45.22	60.31	32.98	**75.78**	44.97	**74.55**	43.59

注：①A和B分别代表CIFAR-10和CIFAR-100数据集；

②粗体表示每列的最大值，下划线表示每列的次优值。

对于鲁棒的噪声容忍训练方法，可以从多个角度来提升学习系统的标签噪声鲁棒性。例如，FedNoRo作为一种典型的鲁棒噪声容忍训练方法，包括鲁棒全局更新、鲁棒局部检测、噪声客户端检测等多种机制，在所有的实验场景下均取得了较为突出的表现。例如，在场景14中FedNoRo具有最高的预测精确度，而在场景3和10中取得了次优的表现，并且在大部分场景中其分类性能非常接近于性能表现最好的FedELC，但同时FedNoRo的模型复杂度更高、灵活性较差。

单一角度优化的方法具有更好的集成性。例如，Symmetric CE容易与其他方法结合以构建更加鲁棒的标签噪声学习系统。而对于混合方法，不同机制间的耦合关系非常复杂，使用相同机制的方法彼此间也具有较大差异。例如，RobustFL在更加复杂的CIFAR-100数据集上的性能异常降低，而FedELC始终保持了较为突出的分类能力；又如Co-teaching与Co-teaching+的核心框架均为双模型协同训练，但由于不同的数据筛选策略，前者始终保持了较为突出的分类性能，甚至在场景12、13和16中均取得了最优的精确度指标，而后者的分类性能明显较差。

总体而言，完全基于显式的标签噪声检测机制过于依赖噪声检测的准确性，而在多种复杂的噪声场景下，标签噪声检测的质量难以始终满足需求。正确标注样本被错误识别为噪声可能浪费

其训练价值，甚至由于标签纠正的质量问题而引入额外的标签噪声，而那些没有被识别的噪声样本则不断地干扰系统训练。因此，显式的标签噪声检测机制一般与标签噪声鲁棒的训练方法相结合以提高训练系统的噪声容忍性。而不使用显式标签噪声检测机制的鲁棒噪声容忍训练方法可以通过评估数据质量进行差异化处理，这种量化的机制具有显著的效果，但处理的范围包含全部样本，难免引入额外的训练误差。最后，虽然混合方法的整体复杂度相对较高，但能够最充分地发挥不同样本的训练价值且减少标签噪声的影响，代表了联邦标签噪声学习发展的趋势。

3 展望

当前，联邦标签噪声学习的相关研究已经充分探索了鲁棒的噪声容忍训练方法、基于噪声检测的训练方法以及相关混合方法，并在多种噪声场景中均取得了显著的成效，如考虑非独立同分布、标签噪声异质、极端噪声率等问题，但在应用中仍有一些挑战亟待解决。

1）全局类不平衡在现实世界中非常普遍，例如在医学图像分析中，某些疾病较为常见，而另一些罕见疾病只拥有少量样本；在自然语言处理中，某些语言或方言拥有大量数据，而另一些语言数据较为稀少。模型对全局类不平衡中少数类别的学习较为困难，往往呈现出较高的损失值，因而影响噪声样本的检测。并且不平衡子群体自身会使模型容易过拟合头子群的模式，而忽略尾子群。尤其在类间噪声异质的情况下，如少数子群体拥有更高的噪声率，会进一步加剧子群体间的不平衡，降低模型的泛化能力。因此，解决全局类不平衡与标签噪声耦合带来的噪声检测精度降低以及模型在少数类或者子群体影响下的性能下降问题具有重要的现实意义。

2）目前的联邦标签噪声学习研究主要集中在特定的噪声类型，如对称噪声、非对称噪声和真实噪声，对于非随机噪声、具有复杂转移矩阵的随机噪声关注较少。这些噪声拥有复杂的特征分布，干扰噪声样本的检测，对训练的负面影响可能更加隐蔽。另外，复杂的分布式噪声场景同样受关注较少，而在现实环境中，由于参加训练的客户端标注主体不同、标注环境与方法差异等，客户端之间往往呈现出复杂的噪声异质现象。相对于较为简单的标签噪声率异质、噪声类型同质的情况，噪声率异质和噪声类型异质共同存在的场景对于噪声客户端、噪声样本的检测以及模型的优化训练均提出了更大的挑战。单一的检测方法和学习策略难以满足训练的准确性和泛化性需求，探索解决复杂分布式噪声场景的鲁棒学习方法具有重要的应用价值。

3）现实世界的数据分布呈现出多样复杂的特征，但目前的联邦标签噪声学习研究主要集中在特定的数据分布，如Dirichlet分布，对于某些极端情况，如客户端可能只包含少数几个类别，甚至某些类别完全缺失等，现有方法难以有效适用。为了更为广泛地利用这些数据的潜在训练价值，需要探索如何在极端数据分布条件下更好地融合节点间差异巨大的本地知识。极端的数据分布异质容易造成节点间的知识消融，降低模型准确性，甚至导致模型无法收敛。而在标签噪声的干扰下，模型的学习过程存在更大的混肴情况，分布式学习面临更严峻的聚合困境，即全局模型的准确性弱于本地模型。因此，需要探索更为广泛数据分布下的联邦标签噪声系统。

4）现实世界的数据来源具有复杂性和多样性，并且是一种动态持续的过程。因此，样本分布往往呈现出高度的不确定性。开集场景考虑了现实世界的开放问题，相比于闭集场景具有更大的应用价值和适用范围。闭集场景中，模型只需要学习如何纠正已知类别的错误标签，或者具有已知类别分布的噪声鲁棒性，而开集场景的标签噪声面临着数据来源的多样性和不确定性，噪声样本的分布呈现出丰富和复杂的特征，极大增加了噪声处理的难度。开集场景下，模型要学习已知类别的数据分布特征，将已知类别和未知类别进行区分。但噪声样本干扰了这个过程，一方面会造成已知类别数据特征分布学习的偏移，另一方面容易将未知类别数据聚类到已知类别，严重干扰模型的学习过程。探索建设适用于开集场景的联邦标签噪声学习系统具有重要的现实意义。

参考文献

[1] KIM S Y，LEE D H，KANG S K，et al. Learning discriminative dynamics with label corruption for noisy label detection[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition（CVPR）. Piscataway，NJ，USA：IEEE，2024：22477-22487.

[2] CHOI J H, YUN J M, JIN K H, et al. Multi-news+: cost-efficient dataset cleansing via LLM-based data annotation[C]//Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA, USA: ACL, 2024: 15-29.

[3] SHI J L, ZHANG K L, GUO C Y, et al. A survey of label-noise deep learning for medical image analysis[J]. Medical Image Analysis, 2024, 95: 103166.

[4] CHEN Z, LI W Y, XING X H, et al. Medical federated learning with joint graph purification for noisy label learning[J]. Medical Image Analysis, 2023, 90: 102976.

[5] KIM T H, KIM D G, YUN S Y. Revisiting early-learning regularization when federated learning meets noisy labels[EB/OL].(2024-02-08) [2025-05-10]. http://arxiv.org/abs/2402.05353v1.

[6] GORNITZ N, PORBADNIGK A, BINDER A, et al. Learning and evaluation in presence of non-IID label noise[C]//Proceedings of the 17th International Conference on Artificial Intelligence and Statistics. Cambridge, MA, USA: PMLR, 2014: 293-302.

[7] WANG K B, HE Q, CHEN F F, et al. FlexiFed: personalized federated learning for edge clients with heterogeneous model architectures[C]//Proceedings of the ACM Web Conference 2023. New York, USA: ACM, 2023: 2979-2990.

[8] QIAO C, LI M J, LIU Y, et al. Transitioning from federated learning to quantum federated learning in internet of things: a comprehensive survey [J]. IEEE Communications Surveys & Tutorials, 2025, 27(1): 509-545.

[9] MCMANHAN B, MOORE E, RAMAGE D, et al. Communication-efficient learning of deep networks from decentralized data[C]//Proceedings of the 20th International Conference on Artificial Intelligence and Statistics. Cambridge, MA, USA: PMLR, 2017: 1273-1282.

[10] WU N N, YU L, JIANG X F, et al. FedNoRo: towards federated learning by addressing class imbalance and label noise heterogeneity[C]//Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence. Red Hook, NY, USA: Curran Associates, Inc., 2023: 4424-4432.

[11] WANG Y W, DING W L, LIN W Y, et al. FedHNR: federated hierarchical resilient learning for echocardiogram segmentation with annotation noise[J]. Expert Systems with Applications, 2025, 273: 126841.

[12] SHI S P, GUO Y Y, WANG D, et al. Distributionally robust federated learning for network traffic classification with noisy labels[J]. IEEE Transactions on Mobile Computing, 2024, 23(5): 6212-6226.

[13] CHEN H M, WANG H D, LONG Q Y, et al. Advancements in federated learning: models, methods, and privacy[J]. ACM Computing Surveys, 2025, 57(2): 1-39.

[14] LYU L J, YU H, ZHAO J, et al. Threats to federated learning[C]//YANG Q, FAN L X, YU H, et al. Federated Learning: Privacy and Incentive. Switzerland: Springer Publishing, 2020: 3-16.

[15] WU X L, CHEN Y W, YU H Y, et al. Privacy-preserving federated learning based on noise addition[J]. Expert Systems with Applications, 2025, 267: 126228.

[16] CHEN J Y, LI S Y, HUANG S J, et al. UNM: a universal approach for noisy multi-label learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2024, 36(9): 4968-4980.

[17] ZHANG C Y, BENGIO S, HARDT M, et al. Understanding deep learning (still) requires rethinking generalization[J]. Communications of the ACM, 2021, 64(3): 107-115.

[18] FRÉNAY B, VERLEYSEN M. Classification in the presence of label noise: a survey[J]. IEEE Transactions on Neural Networks and Learning Systems, 2014, 25(5): 845-869.

[19] JU L, WANG X, WANG L, et al. Improving medical images classification with label noise using dual-uncertainty estimation[J]. IEEE Transactions on Medical Imaging, 2022, 41(6): 1533-1546.

[20] DU J, CAI Z H. Modelling class noise with symmetric and asymmetric distributions[C]//BONET B, KOENIG S. Proceedings of the 29th AAAI Conference on Artificial Intelligence. Washington, DC, USA: AAAI Press, 2015, 29(1): 2589-2595.

[21] JIANG X F, LI J, WU N N, et al. FNBench: benchmarking robust federated learning against noisy labels[EB/OL].(2025-05-10) [2025-05-11]. https://arxiv.org/abs/2505.06684v1.

[22] WU N N, SUN Z B, YAN Z Q, et al. FedA3I: annotation quality-aware aggregation for federated medical image segmentation against heterogeneous annotation noise[C]//WOOLDRIDGE M J, DY J G,

NATARAJAN S. Proceedings of the 38th AAAI Conference on Artificial Intelligence. Washington, DC, USA: AAAI Press, 2024: 15943-15951.

[23] DUBEY P, KUMAR M. Integrating explainable AI with federated learning for next-generation IoT: a comprehensive review and prospective insights[J]. Computer Science Review, 2025, 56: 100697.

[24] FANG X W, YE M. Robust federated learning with noisy and heterogeneous clients[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2022: 10062-10071.

[25] SACHDEVA R, CORDEIRO F R, BELAGIANNIS V, et al. EvidentialMix: learning with combined open-set and closed-set noisy labels[EB/OL]. (2020-11-11)[2025-05-10]. http://arxiv.org/abs/2011.05704v1.

[26] LI J C, LI G B, CHENG H, et al. FedDiv: collaborative noise filtering for federated learning with noisy labels[C]//WOOLDRIDGE M J, DY J G, NATARAJAN S. Proceedings of the 38th AAAI Conference on Artificial Intelligence. Washington, DC, USA: AAAI Press, 2024: 3118-3126.

[27] JIANG X F, SUN S, LI J, et al. Tackling noisy clients in federated learning with end-to-end label correction[C]//SERRA E, SPEZZANO F. Proceedings of the 33rd ACM International Conference on Information and Knowledge Management. New York, USA: ACM, 2024: 1015-1026.

[28] CHEN P F, YE J J, CHEN G Y, et al. Beyond class-conditional assumption: a primary attempt to combat instance-dependent label noise[C]//LEYTON-BROWN K, FREUND Y. Advances in Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates, Inc., 2021: 11442-11450.

[29] WEI J H, ZHU Z W, CHENG H, et al. Learning with noisy labels revisited: a study using real-world human annotations[EB/OL]. (2022-03-27)[2025-05-10]. https://arxiv.org/abs/2110.12088.

[30] SONG H J, KIM M S, LEE J G. SELFIE: refurbishing unclean samples for robust deep learning[C]//Proceedings of the 36th International Conference on Machine Learning. Cambridge, MA, USA: PMLR, 2019: 5907-5915.

[31] JIANG L, HUANG D, LIU M, et al. Beyond synthetic noise: deep learning on controlled noisy labels[C]//Proceedings of the 37th International Conference on Machine Learning. Cambridge, MA, USA: PMLR, 2020: 4804-4815.

[32] XIAO T, XIA T, YANG Y, et al. Learning from massive noisy labeled data for image classification[C]//2015 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2015: 2691-2699.

[33] LEE K H, HE X D, ZHANG L, et al. CleanNet: transfer learning for scalable image classifier training with label noise[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 5447-5456.

[34] LI W, WANG L M, LI W, et al. WebVision database: visual learning and understanding from web data[EB/OL]. (2017-08-09)[2025-05-10]. http://arxiv.org/abs/1708.02862v1.

[35] WANG H D, JIANG T Y, GUO Y, et al. Label noise correction for federated learning: a secure, efficient and reliable realization[C]//2024 IEEE 40th International Conference on Data Engineering (ICDE). Piscataway, NJ, USA: IEEE, 2024: 3600-3612.

[36] XU J Y, CHEN Z H, QUEK T Q S, et al. FedCorr: multi-stage federated learning for label noise correction[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2022: 10174-10183.

[37] ZHOU S S, LI K, CHEN Y X, et al. TrustBCFL: mitigating data bias in IoT through blockchain-enabled federated learning[J]. IEEE Internet of Things Journal, 2024, 11(15): 25648-25662.

[38] JIANG X F, LI P, SUN S, et al. Refining distributed noisy clients: an end-to-end dual optimization framework[EB/OL]. (2025-01-17)[2025-04-16]. https://www.authorea.com/doi/full/10.36227/techrxiv.173707406.66001019/v1.

[39] LI W Q, FU S R, ZHANG F R, et al. Data valuation and detections in federated learning[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2024: 12027-12036.

[40] ZHAO Z Y, LI A R, LI R D, et al. FedPPO: reinforcement learning-based client selection for federated learning with heterogeneous data[J]. IEEE Transactions on Cognitive Communications and Networking, 2025, 18(9): 1-13.

[41] LI A R, WANG G J, HU M, et al. Joint client-and-sample selection for federated learning via bi-level optimization[J]. IEEE Transactions on Mo-

bile Computing, 2024, 23(12): 15196-15209.

[42] GIAP T T, KIEU T D, LE T L, et al. FedDC: label noise correction with dynamic clients for federated learning[J]. IEEE Internet of Things Journal, 2025, 12(8): 10266-10277.

[43] GUI X J, WANG W, TIAN Z H, et al. Towards understanding deep learning from noisy labels with small-loss criterion[EB/OL].(2021-06-17)[2025-05-11]. https://arxiv.org/abs/2106.09291v1.

[44] MISHRA R, GUPTA H P, DUTTA T. Noise-resilient federated learning: suppressing noisy labels in the local datasets of participants[C]//IEEE Conference on Computer Communications Workshops. Piscataway, NJ, USA: IEEE, 2022: 1-2.

[45] ZENG B X, YANG X D, CHEN Y Q, et al. Federated data quality assessment approach: robust learning with mixed label noise[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(12): 17620-17634.

[46] LI J N, SOCHER R, HOI S C H. DivideMix: learning with noisy labels as semi-supervised learning[EB/OL].(2020-02-18)[2025-05-10]. http://arxiv.org/abs/2002.07394v1.

[47] SUN W J, YAN R Q, JIN R B, et al. Curriculum-based federated learning for machine fault diagnosis with noisy labels[J]. IEEE Transactions on Industrial Informatics, 2024, 20(12): 13820-13830.

[48] JIANG H, CHEN Y X, LIU L, et al. Leveraging noisy labels of nearest neighbors for label correction and sample selection[C]//2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway, NJ, USA: IEEE, 2024: 7720-7724.

[49] CHEN M C, ZHAO Y, HE B, et al. Learning with noisy labels over imbalanced subpopulations [J]. IEEE Transactions on Neural Networks and Learning Systems, 2025, 36(4): 6544-6555.

[50] XU Y, LIAO Y M, WANG L, et al. Overcoming noisy labels and non-IID data in edge federated learning[J]. IEEE Transactions on Mobile Computing, 2024, 23(12): 11406-11421.

[51] WANG Y S, MA X J, CHEN Z Y, et al. Symmetric cross entropy for robust learning with noisy labels[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2019: 322-330.

[52] KIM S, SHIN W, JANG S, et al. FedRN: exploiting k-reliable neighbors towards robust federated learning[C]//Proceedings of the 31st ACM International Conference on Information & Knowledge Management. New York, USA: ACM, 2022: 972-981.

[53] HAN B, YAO Q M, YU X R, et al. Co-teaching: robust training of deep neural networks with extremely noisy labels[J]. Advances in Neural Information Processing Systems, 2018, 31: 8536-8546.

[54] YU X R, HAN B, YAO J C, et al. How does disagreement help generalization against label corruption? [C]//Proceedings of the 36th International Conference on Machine Learning. Cambridge, MA, USA: PMLR, 2019: 7164-7173.

[55] TUOR T, WANG S Q, KO B J, et al. Overcoming noisy and irrelevant data in federated learning[C]//25th International Conference on Pattern Recognition(ICPR). Piscataway, NJ, USA: IEEE, 2020: 5020-5027.

[56] YANG S, PARK H, BYUN J, et al. Robust federated learning with noisy labels[J]. IEEE Intelligent Systems, 2022, 37(2): 35-43.

[57] LI T, HU S Y, BEIRAMI A, et al. Ditto: fair and robust federated learning through personalization [C]//Proceedings of the 38th International Conference on Machine Learning. Cambridge, MA, USA: PMLR, 2021: 6357-6368.

[58] BLANCHARD P, EL MHAMDI E M, GUERRAOUI R, et al. Machine learning with adversaries: byzantine tolerant gradient descent[C]//GUYON I, LUXBURG U V, BENGIO S, et al. Advances in Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates, Inc., 2017: 119-129.

[59] YIN D, CHEN Y D, KANNAN R, et al. Byzantine-robust distributed learning: towards optimal statistical rates[C]//Proceedings of the 35th International Conference on Machine Learning. Cambridge, MA, USA: PMLR, 2018: 5650-5659.

[60] GHOSH A, KUMAR H, SASTRY P S. Robust loss functions under label noise for deep neural networks[EB/OL].(2017-12-27)[2025-05-11]. https://arxiv.org/abs/1712.09482v1.

[61] ZHANG Z L, SABUNCU M R. Generalized cross entropy loss for training deep neural networks with noisy labels[EB/OL].(2018-11-29)[2025-05-11]. https://arxiv.org/abs/1805.07836v1.

[62] PU R Z, YU L X, ZHAN S J, et al. FedELR:

when federated learning meets learning with noisy labels[J]. Neural Networks，2025，187：107275.

[63] JIANG X F，SUN S，WANG Y W，et al. Towards federated learning against noisy labels via local self-regularization[C]//Proceedings of the 31st ACM International Conference on Information & Knowledge Management. New York，USA：ACM，2022：862-873.

[64] ZHOU X C，WANG X D. Federated label-noise learning with local diversity product regularization [C]//WOOLDRIDGE M J，DY J G，NATARAJAN S. Proceedings of the 38th AAAI Conference on Artificial Intelligence. Washington，DC，USA：AAAI Press，2024：17141-17149.

[65] JIANG L，ZHOU Z Y，LEUNG T，et al. MentorNet：learning data-driven curriculum for very deep neural networks on corrupted labels[EB/OL].(2018-08-13)[2025-05-10]. http：//arxiv.org/abs/1712.05055v2.

[66] REN M Y，ZENG W Y，YANG B，et al. Learning to reweight examples for robust deep learning [C]//Proceedings of the 35th International Conference on Machine Learning. Cambridge，MA，USA：PMLR，2018：4334-4343.

[67] FANG X W，YE M. Noise-robust federated learning with model heterogeneous clients[J]. IEEE Transactions on Mobile Computing，2025，24(5)：4053-4071.

[68] JIANG Y L，WANG D，SONG B，et al. HDH-RFL：a hierarchical robust federated learning framework for dual-heterogeneous and noisy clients[J]. Future Generation Computer Systems，2024，160：185-196.

[69] YAN Y，GUO Y H，WOOLDRIDGE M J，et al. Federated partial label learning with local-adaptive augmentation and regularization[C]//WOOLDRIDGE M J，DY J G，NATARAJAN S. Proceedings of the 38th AAAI Conference on Artificial Intelligence. Washington，DC，USA：AAAI Press，2024：16272-16280.

[70] TANAKA D，IKAMI D，YAMASAKI T，et al. Joint optimization framework for learning with noisy labels[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Piscataway，NJ，USA：IEEE，2018：5552-5560.

（编辑：刘建民）