FedLE：面向边缘物联网网络的联邦学习客户端选择与寿命延长

Jiajun Wu电气与软件工程 卡尔加里大学加拿大卡尔加里
jiajun.wu1@ucalgary.ca

Steve Drew电气与软件工程 卡尔加里大学加拿大卡尔加里
steve.drew@ucalgary.ca

Jiayu Zhou计算机科学与工程 密歇根州立大学美国东兰辛
jiayuz@msu.edu

arXiv:2302.07305v1 [cs.LG] 14 Feb 2023

摘要——**联邦学习（FL）是一种分布式、保护隐私的预测建模学习框架，能够处理物联网（IoT）设备在边缘生成的大量数据。阻碍联邦学习在物联网中广泛应用的一大挑战，是物联网设备普遍面临电源限制，因为电池供电客户端进行本地训练和模型更新时能耗巨大。客户端电池电量过低最终会导致它们提前退出边缘网络，造成训练数据流失，损害联邦学习的性能，并使这些设备无法执行其他指定任务。本文提出 FedLE，这是一种节能客户端选择框架，可延长边缘物联网网络的寿命。在 FedLE 中，客户端首先运行最小训练轮次，以生成本地模型更新。随后，模型会部分上传至服务器，用于计算每对客户端之间的相似度。通过对这些客户端对进行聚类，找出模型分布相似的客户端。每一轮中，低功耗客户端被选中的概率较低，从而延缓其电池耗尽。实证研究表明，在基准数据集上，FedLE 的表现优于基线方法；在电池电量限制下，其持续的训练轮次也多于 FedAvg。**

索引词——**联邦学习、客户端选择、边缘计算、物联网（IoT）、能源效率**

I. 引言

物联网（IoT）设备的广泛部署，使网络边缘产生了海量个人数据 [1]。出于隐私顾虑和资源限制，边缘侧的物联网数据无法上传至集中式服务器进行机器学习。联邦学习（FL）允许分布式客户端在不上传本地训练数据的情况下协同学习模型，为边缘计算中的隐私保护型分布式学习提供了理想的解决方案。实际上，要有效开展联邦学习仍面临诸多挑战，因为物联网网络大多依靠电池供电，且具有异构性，电池容量各不相同，寿命也可能较短。更糟的是，物联网设备在联邦学习训练过程中会消耗大量能源。部分设备会比其他设备更早耗尽能源并退出网络。这些退出会减少可用客户端的数量，进而限制对训练数据的访问。原始 FedAvg [2] 以均匀随机的方式执行客户端选择，并未考虑客户端的能源状况和状态。在新一轮数据聚合中，剩余能源较低的关键客户端与非关键客户端被选中的概率相同。这些因素可能导致关键客户端过早退出，降低整体网络可用性，进而造成模型性能下降和收敛时间延长。

由于关键客户端可能流失，客户端选择成为一项主要关注点。现有研究主要聚焦于通过无线网络中的带宽分配来优化能源效率。Yu 等人 [3] 研究了最大化所选客户端数量与最小化总能耗之间的权衡，并探讨了如何选择客户端，为联邦学习本地训练优化分配 CPU 资源，并为模型更新优化传输功率。 [4]讨论了时序客户端选择模式变化对学习性能的影响。随后，有研究构建了一个随机优化问题，以联合优化带宽分配和客户端选择，并提供长期性能保证。FedMCCS[5] 旨在考虑客户端资源预算及其 CPU、内存和能源等能力，同时最大化每轮所选客户端的数量。尽管在某些场景下，客户端选择率和测试准确率有所提升，但现有研究未考虑不可靠的边缘计算环境，因而无法在客户端数量更充足的情况下最大限度地延长整个边缘网络的寿命。

本文提出了 FedLE，这是一种用于联邦学习的基于聚类的客户端选择方法，可加速异构数据场景下的训练。FedLE 的设计思路源于模型相似性：在联邦学习过程中，数据相似的客户端往往会提供相似的模型。因此，优先选择模型差异显著的客户端作为一种采样策略，可以高效地减少潜在冗余的模型上传，从而加速联邦学习过程。基于这一思路，FedLE 按照图 1 所示的步骤，根据客户端的本地模型权重形成簇：服务器计算客户端模型权重的相似度矩阵，并对客户端进行聚类。客户端被选中的概率取决于其所属簇的大小。采用这种方法，我们不会损害客户端隐私，不会直接访问客户端数据，也不会从客户端获取新信息

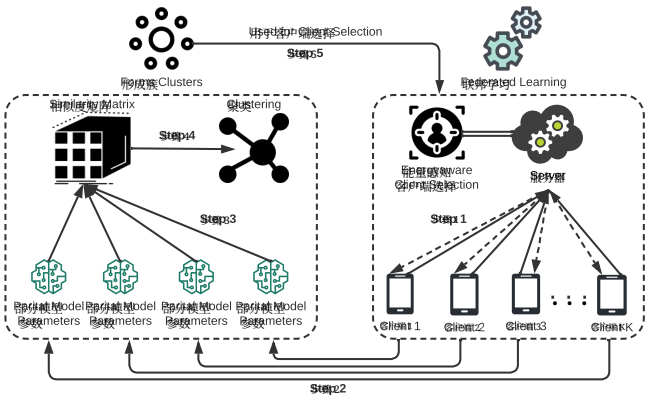

图 1. FedLE 客户端选择框架概览。1）在初始训练轮次之前，所有客户端运行至少一个轮次，并将部分模型权重上传至服务器。2）服务器计算相似度矩阵。3）通过在权重空间中将每个客户端与相似度最低的一对客户端进行比较，来表示每个客户端。4）根据客户端的权重对其进行聚类。5）根据簇大小动态设置被选中的概率。

。与现有研究相比，FedLE 具有以下贡献：

- FedLE 平衡能耗，并延长所有客户端的运行时间。据我们所知，这是首项针对 FL 考虑边缘网络寿命延长的研究。
- 我们提出的方法使用部分模型层构建用于聚类的相似度矩阵，事实证明，这比使用完整模型的能效高得多。
- 实证研究表明，在大多数情况下，相似度矩阵的创建和聚类只需执行一次，因此仅增加少量开销，同时仍优于基线方法。

II. 相关工作

FedAvg [2]证明了当所有客户端均可用且接受了充分训练时，随机客户端选择是有效的。为提高异构数据下联邦学习的性能，FedProx 被提出 [6], ，通过引入近端项限制本地更新，使其更接近全局模型。
节能客户端选择 已成为联邦学习中的热门主题之一。power-of-choice [7]客户端选择框架优先考虑通信和计算效率。Nishio和Yonetani [8] 提出了一种名为FedCS的联邦学习客户端选择协议，该协议同时考虑了移动边缘计算（MEC）环境中的能源资源和异构数据。尽管现有方法在某些场景下提高了能源效率，但它们没有考虑延长边缘网络和物联网网络的寿命，以平衡客户端的能耗。相比之下，我们提出的FedLE框架考虑了联邦学习的总轮次数，旨在延长整个边缘网络的有效寿命，从而保留数据源并帮助物联网设备执行主要任务。在FedSim[9], 中，客户端两两相似度也被用于聚类，但该方法充分利用了客户端的本地模型参数的客户端，并将聚类作为一种联邦学习后处理技术。作为我们所提方法的一部分，我们仅以一次性的方式利用部分本地模型参数进行客户端选择。Wang等人 [10] 提出了一种联邦学习方案，该方案使用余弦相似度移除无效模型，并通过将本地模型参数和全局模型参数向量化来对设备进行聚类。该方案还要求使用完整的模型参数，且未考虑能源因素。我们提出的方法仅使用本地模型参数计算相似度，并且需要更少的计算量和通信功耗。

III. FedLE客户端选择框架

本节将介绍 FedLE 客户端选择框架及其问题表述。

A. 受限电池电量

我们首先定义一个基于边缘的联邦学习场景，其中包含一组电池供电客户端 $\mathbb{C}$，其中 $|\mathbb{C}| = K$。所有客户端都与聚合服务器 S通信。我们使用联邦学习中的通信轮次来表示系统中的当前时间。具体而言，对于第 k个客户端 $c_k \in \mathbb{C}$，将其电池电量定义为第 t轮时可供主动使用的剩余电量，记为 $b_t^k \in [0,1]$，其中 0 表示电池电量完全耗尽，1 表示电量充满。只要客户端处于开机状态，即使闲置也会消耗设备电量。我们将 c_k 的待机电池放电率定义为每个联邦学习训练轮次中减少的 b_t^k 电量，记为 r_t^k。对于被选中进行模型更新的客户端，向服务器传输会额外消耗电量。我们将每次模型更新传输造成的电池放电量定义为 s_t^k。对于其他通信造成的任何电池放电，我们将 c_k 的放电量记为 a_t^k。第 $t+1$ 轮时 c_k 的总电量为：

$$b_{t+1}^k = b_t^k - r_t^k - s_t^k - a_t^k.$$

边缘设备通常除了参与联邦学习外，还承担着主要任务。我们设置了一个临界电池电量 δ，以避免联邦学习过程将设备电池电量完全耗尽。如果客户端的电池电量低于 δ，联邦学习过程将暂停，以便边缘物联网设备完成其主要任务。

B. 选择统计多样化客户端

在联邦学习中，大量客户端可能携带数据分布高度相似的数据。典型例子包括距离较近的工业环境中的物联网传感器、同一条街道上的监控摄像头，以及偏好相同的用户所使用的智能手机。在联邦学习过程中，有必要避免在连续轮次中选择此类客户端。当多数客户端共享分布相似的数据时，训练结果很可能会导致模型对多数类过拟合。另一方面，客户端的能源限制会限制其训练总轮次：剩余能源有限的客户端可参与的轮次数量有限。因此，参与训练的机会十分宝贵。选择合适的客户端非常重要

，应选择数据分布多样的客户端，以促进全局模型的收敛。我们计划比较每个客户端的本地模型权重，构建相似度矩阵，并根据权重对客户端进行聚类。

算法 1 构建相似度矩阵的算法

如果初始轮次 **then**
  for 每个客户端 $c_k \in \mathbb{C}$ **执行**
    在本地训练客户端 c_k，训练 1 轮
    c_k 向服务器发送模型更新 w_k
    服务器将**部分** w_k展平
  结束循环
  服务器比较每对
K 客户端的点积，以构建 $K \times K$ 相似度矩阵
  服务器将 $K \times K$中每一行的值相加
矩阵转换为大小为 K的一维列表
结束 if

C. 高效构建相似度矩阵

在实践中构建相似度矩阵的主要挑战包括：比较每两个客户端的梯度所带来的开销。创建相似度矩阵需要进行总计 $\binom{K}{2}/2$ 次完整模型更新比较。尽管此过程在服务器端完成，但会带来显著的计算成本和延迟。当 K 和模型梯度矩阵都很大时，问题会更加严重。为应对构建相似度矩阵时面临的这些挑战，我们在更新矩阵时仅将部分模型参数加载到服务器。根据我们的观察，本地客户端模型在模型权重和偏置方面存在显著差异，尤其是在卷积神经网络（CNN）模型的第一层和最后一层。因此，我们不使用整个模型来构建矩阵，而仅使用第一卷积层和最后一个全连接层的模型参数。假设对模型进行裁剪后，对于 c_k 的每个梯度矩阵，我们得到一个 $D \times D$ 梯度矩阵。

为了证明使用部分模型的有效性，我们设计了一项实验，展示在联邦学习过程中，每个客户端使用部分模型参数时的余弦相似度。在 FedAvg 设置下，我们使用 MNIST [11] 训练卷积神经网络模型，共进行 30 轮；使用 CIFAR-10 [12] 数据集共进行 20 轮，每轮均包含 15 个客户端。每轮中，每个客户端都会计算其本地模型权重与所有其他客户端之间的余弦相似度，从而得到 k 个余弦相似度值。随后，我们构造一个 $K \times K$ 矩阵，其中每个单元格对应两个客户端之间的相似度分数，计算方式如下：

$$sim(c_i, c_j) = \mathbf{w_i} \cdot \mathbf{w_j} = \sum_{d_i=1}^{D} \sum_{d_j=1}^{D} w_i^{d_i} w_j^{d_j}, \qquad (1)$$

其中，$\mathbf{w_i}$ 和$\mathbf{w_j}$ 是客户端的梯度矩阵 c_i 和c_j，分别。变量 $w_i^{d_i}$和 $w_j^{d_j}$ 表示单元格

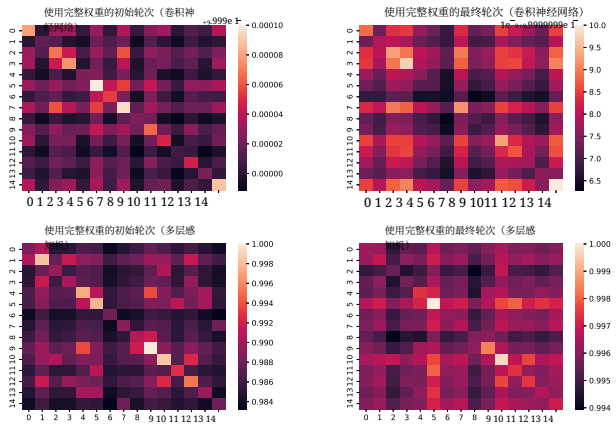

图 2. 使用完整模型权重得到的相似度矩阵。（上排：卷积神经网络模型的初始轮次和最终轮次矩阵。下排：多层感知机模型的初始轮次和最终轮次矩阵。）

值在$\mathbf{w_i}$ 和 $\mathbf{w_j}$的梯度中。应用公式（1），我们将每一轮中所有矩阵的相似度矩阵可视化为热力图，涵盖完整权重和部分权重。结果如图 2 所示。

1）初始相似度矩阵的重要性：观察图 2 中的热力图可知，模式并没有显著变化。我们还发现，随着轮次增加，相似度矩阵的模式仍相对稳定。连续矩阵之间乃至后续矩阵之间都没有显著差异。我们认为，这是因为每个客户端本地数据的分布通常较为稳定，不太可能发生剧烈变化。从该实验可以明显看出，使用相似度矩阵是有效的，而且不需要频繁更新矩阵。

2）不同非独立同分布程度及不平衡数据下的鲁棒性：我们设计了一项简单实验，使用具有不同异质性程度和类别不平衡的数据，测试相似度矩阵的鲁棒性。为便于展示，我们设置了两个异质性等级。对于使用卷积神经网络的 MNIST、使用多层感知机（MLP）的 MNIST，以及使用卷积神经网络模型的 CIFAR-10，我们采用了 40 个客户端，并为每个客户端分配 2 类或 3 类数据，其中 50% 的客户端共享来自同一类别的数据。图 3 中的权重空间图表明，无论数据异质性程度如何，其聚类模式都很清晰。

根据这些信息，我们认为相似度矩阵的构建过程可以一次性完成。因此，通过限制矩阵的创建和更新次数，并缩减模型参数的规模，可以大幅降低客户端在构建相似度矩阵时与服务器之间的通信成本。在每轮训练中，客户端选择都基于该矩阵及其簇。我们可以在单轮中选择数据更多样化的客户端，从而为全局模型作出更大贡献。

根据每个客户端数据变化的频率和幅度，可能需要重构矩阵。如果客户端数据是动态的且不断变化，

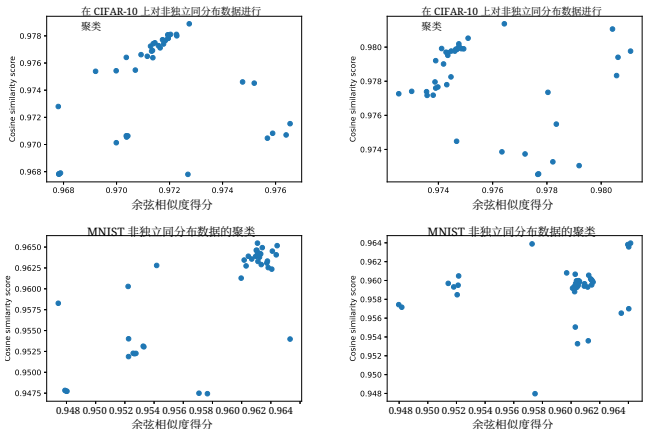

图 3. 使用卷积神经网络和多层感知机时，FedLE 在不同异质性程度下的表现。第一行：使用 CIFAR-10 的卷积神经网络模型，每个客户端包含 2 类和 3 类数据。第二行：使用 MNIST 的卷积神经网络模型，每个客户端包含 2 类和 3 类数据。第三行：使用 MNIST 的多层感知机模型，每个客户端包含 2 类和 3 类数据。

那么就需要更新权重矩阵。不过，这并非本文的重点。

D. 形成簇

为最大限度地提高能源效率，整个训练过程中仅执行一次聚类操作。所有客户端都会先进行一轮本地训练，然后上传其本地模型更新。根据本地模型权重，将每个客户端的权重与所有其他客户端的权重进行比较。构建相似度矩阵后，FedLE 从矩阵中选取最低的 N_s 相似度值。选择相似度最低的客户端进行聚类，旨在找出多样性更高的客户端，并避免反复选择相似度较低的客户端。在我们的实验中，我们选取最低相似度值进行分析，该值对应一对客户端 c_α 和 c_β。这两个客户端被认为彼此相距最远。其他客户端的相似度得分则介于两者之间。为了应对大多数客户端数据相似所导致的过拟合问题，我们根据各簇的大小为其设置不同的概率。对于客户端数量最多的簇，我们将其归类为多数簇。多数簇被选中的概率要低得多，其余簇被选中的概率则相同。这样一来，包含统计上相似数据的簇被选中的概率就会较低。客户端较少的簇被选中参与训练的概率则较高。

E. 提出的 *FedLE* 客户端选择方法

本节提出一种基于梯度相似度的全新客户端选择方法，用于在能源受限场景下优化联邦学习的训练。我们介绍了两种基线方法和一种优化方法。根据第 III-B 节中的讨论，在 FedLE 中，我们在更新矩阵时仅将部分模型参数加载到服务器。各本地客户端模型在模型权重和偏置方面存在显著差异，尤其是在卷积神经网络模型的第一层和最后一层。我们不使用整个模型来构建矩阵，而仅使用第一卷积层和最后一个全连接层的模型参数。

算法 2 FedLE 算法。K 表示参与的客户端数量。C $\in$ (0,1]是一个超参数，表示每轮联邦学习训练中参与客户端的比例。

1: **初始化：** 服务器将全局模型发送给所有客户端。2: **构建相似度矩阵：** 见算法 1。3: **K 均值聚类：** 服务器根据每个客户端之间的成对相似度，使用 K 均值聚类对所有客户端进行聚类。4: **客户端选择：** 服务器根据簇大小选择$K \times C$ 个客户端。5: **分发：** 服务器在本轮将最新的全局模型分发给选定的客户端。6: **更新和上传：** 选定的客户端使用全局模型和本地数据进行本地训练后，更新其本地模型参数。7: **聚合：** 服务器对上传的参数取平均，并将其聚合为下一轮的新全局模型。8: 从**客户端选择** 开始重复所有步骤，直至模型达到预期效果。

IV. 仿真设置

本节中，我们使用 MNIST 和 CIFAR-10，在模拟的能源受限环境中评估 FedLE，并将其与两种基线方法 FedBO 和 FedAvg-B 进行比较。

A. 基线方法

第一种基线方法是在考虑电池约束的情况下采用 FedAvg，称为 *FedAvg-B*。在该基线方法中，客户端选择与 FedAvg 一样采用均匀随机方式，但额外对所有客户端施加电池电量约束。训练期间，电池电量低于临界电量 δ 的客户端将被剔除。

第二种基线方法称为 仅电池（*FedBO*）。在 FedBO 中，客户端选择优先考虑电池电量较高的客户端。FedBO 根据每个客户端的初始电池电量计算其最大训练能力。随后，根据每个客户端的最大能力确定其被选中的概率。从本质上说，低功耗客户端被选中的概率低于高功耗客户端。

对于我们的电池优化方法，在训练开始之前，所有可用客户端都使用各自的本地数据进行 1 轮次本地训练，并将部分本地模型参数上传至服务器。中央服务器计算相似度矩阵并使用 K均值聚类。新的矩阵将用于客户端选择。我们根据以下标准选择客户端：

- FedLE 为客户端较少的簇赋予更高的概率，为客户端较多的簇赋予更低的概率。

- 在每个簇中，电池电量较高的客户端被选中的概率也更高，以优化能源使用

B. 模拟环境

考虑一个包含 K 个客户端和一台中央服务器的典型联邦学习场景。每个客户端根据服务器维护的全局模型，在本地数据上进行训练。我们使用以下变量，模拟一个类别不平衡且受能源约束的联邦学习环境。训练和通信成本体现为客户端剩余的电池电量。每个客户端的电池电量b 决定了何时将该客户端排除在联邦学习训练之外。每轮结束后，即使客户端未被选中进行训练，电池放电r 也会应用于所有客户端。客户端与服务器之间的任何其他通信都会产生额外的通信成本a。

每个客户端的初始电池电量均为B。低功耗客户端的 $b \in [0.245, 0.265]$，高功耗客户端的$b \in [0.895, 0.905]$。我们将50%的客户端初始化为低功耗客户端，另外50%的客户端初始化为高功耗客户端。当客户端的电池电量 $b \leq \delta$，其中 $\delta = 0.2$时，将认为其已达到电池的临界电量，并排除其参与后续任何训练。所有客户端均具有电池待机成本 $r \in [0.1, 0.15]$。我们还设置了额外的电池成本 $s \in [0.4, 0.5]$ 以及额外的通信成本 $a \in [0.24, 0.25]$。在模拟环境中，我们假设所有客户端拥有相同数量的数据和相同的计算能力。其中50%的客户端被分配了来自同一类别的数据点，从而形成类别不平衡场景。我们还为实验设置了一个提前停止点。当50%的客户端电池电量低于 δ 时，整个联邦学习训练过程停止。

C. 非独立同分布与不平衡数据及全局模型

评估中采用卷积神经网络和多层感知机模型。三种方法在联邦学习环境中均采用相同的设置。MNIST 和 CIFAR-10 的联邦学习设置如下，并依据 [2] 中的定义。我们使用 $K = 40$ 和 $C = 0.05$，因此每个训练轮次恰好有 2 个客户端参与。为了创建一个模拟非独立同分布联邦学习和类别不平衡的环境，我们将 MNIST 训练集划分为 20 个大小为 3000 的数据分片，将 CIFAR-10 训练集划分为 80 个大小为 625 的数据分片。然后，我们为前 20 个客户端各分配一个数据分片，并将其余 20 个客户端分配到同一类别。

V. 性能评估

在本节中，我们评估 FedLE 的性能，第 IV 节中的另外两种基线方法。

A. 测试准确率

如图 4 所示，采用卷积神经网络模型的 FedLE 在测试准确率方面大幅优于另外两种基线方法（约高出 10%）。并列比较结果表明，当客户端的数据为非独立同分布且类别不平衡时，FedLE 仍能发挥良好效果。结果显示，仅优化客户端选择

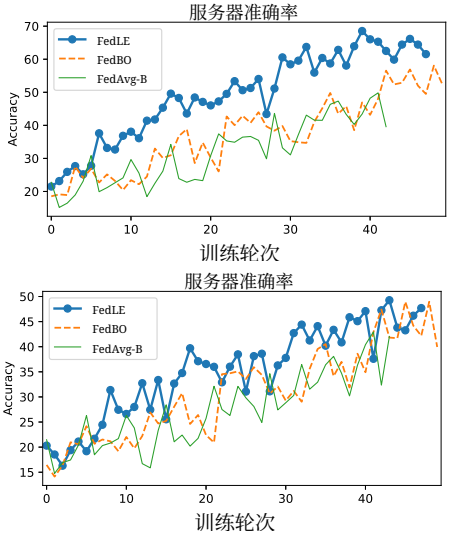

图 4. 使用 FedLE、FedBO 和 FedAvg-B 在 MNIST 上进行的卷积神经网络和多层感知机实验。

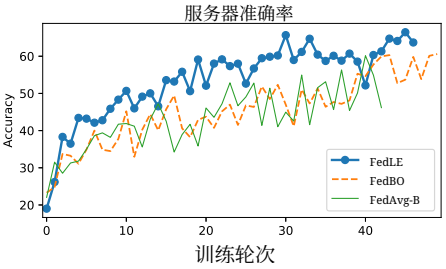

图 5. 使用 FedLE、FedBO 和 FedAvg-B 在 CIFAR-10 上进行的卷积神经网络实验。

就能显著影响准确率。我们可以看到，与 FedAvg 相比，FedLE 能够延长训练时间。我们还使用另一种模型测试了 FedLE 算法。采用多层感知机模型时，FedLE 在前几轮收敛更快，但在训练后期仅略微优于另外两种基线方法。不过，我们认为这是由于多层感知机模型较为简单。在极端的非独立同分布且类别不平衡场景下，简单的多层感知机模型难以收敛，而卷积神经网络模型则具有更强的鲁棒性。

B. 簇数量和多数类

我们进行了另一项实验，以测试簇的数量如何影响 FedLE 的性能。如图 6 所示，增加簇的数量对性能没有显著影响。我们认为，无论簇的数量是多少，FedLE 都会将多数类数据聚集在一起。增加簇的数量只会进一步划分簇中的少数类数据。因此，增加簇的数量对准确率的影响较小。从结果可以看出，最优簇数量应为多数类数据的数量与每轮训练客户端数量之和。

实验采用与第 IV 节相同的设置（对 5 个不同随机种子的结果取平均）。在 MNIST 实验中，我们提出的方法采用不同数量的簇时，结果表明，当簇大小超过训练

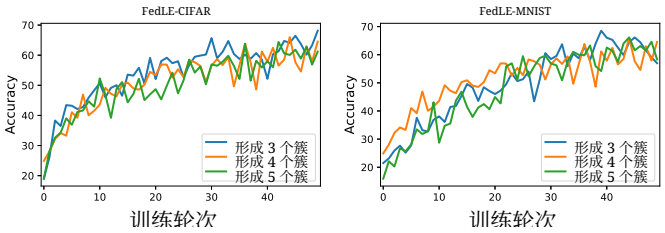

图 6. FedLE 在 MNIST 和 CIFAR-10 上使用不同簇大小的情况。

表 I 达到临界电池电量前持续的轮次数 δ

	FedLE	FedAvg-B	FedBO
70% 低功耗客户端			
MNIST 卷积神经网络	49	45	50
MNIST 多层感知机	48	44	50
CIFAR-10 卷积神经网络	48	45	50
50% 低功耗客户端			
MNIST 卷积神经网络	48	43	50
MNIST 多层感知机	49	44	50
CIFAR-10 卷积神经网络	48	43	50
30% 低功耗客户端			
MNIST 卷积神经网络	49	45	50
MNIST 多层感知机	48	44	50
CIFAR-10 卷积神经网络	49	45	50

客户端数量（本例中为 2）时，无论簇的数量是多少，结果都相近，因为我们的方法能够正确分类大多数类别数据。FedLE 在 CIFAR-10 上采用不同数量的簇时也呈现出类似趋势。

C. 节能结果

在表 I 中，我们展示了边缘物联网网络在能源保护方面的表现，衡量指标为网络持续运行的总轮次数。当 50 % 的客户端因低电量而无法使用时，我们便认为网络不可用，并停止联邦学习训练。可以看到，虽然 FedLE 的表现不及仅以能源为优先考量的 FedBO，但仅少了 2 轮，差距很小。与经典的 FedAvg-B 相比，FedLE 能将训练时间延长约 5 轮，即 11.6%。通过将数据分布相似的客户端聚为一簇，我们降低低功耗客户端被选中的概率，并以同一簇中的高功耗客户端替代它们，从而保护低功耗客户端。由于低功耗客户端仍有小概率被选中，我们认为这就是 FedBO 略胜一筹的原因，因为它对高功耗客户端的优先级要高得多。

六、结论

本文提出了 FedLE，这是一种基于客户端模型相似度矩阵的聚类方法，旨在平衡电池供电的边缘物联网客户端的能耗。实验结果验证了 FedLE 优于基线方法，同时通过防止低功耗客户端耗尽电池，延长了边缘物联网网络的寿命。未来的工作将研究其他深度学习模型的影响，因为该矩阵是基于模型参数计算得出的。

致谢

本材料部分基于以下项目的资助：卡尔加里大学启动经费 10032260、美国国家科学基金会资助项目 IIS-2212174 和 IIS-1749940、美国海军研究办公室项目 N00014-20-1-2382，以及美国国家老龄研究所（NIA）项目 RF1AG072449。

参考文献

[1] J. Wu、S. Drew、F. Dong、Z. Zhu 和 J. Zhou，“边缘计算中的拓扑感知联邦学习：综合综述”，*arXiv* 预印本 *arXiv:2302.02573*，2023。[2] B. McMahan、E. Moore、D. Ramage、S. Hampson 和 B. A. y Arcas，“基于去中心化数据的深度网络通信高效学习”，载于 人工智能与统计，2017 年，第 1273–1282 页。[3] L. Yu、R. Albelaihi、X. Sun、N. Ansari 和 M. Devet-sikiotis，“面向物联网的无线联邦学习中客户端选择与资源管理的联合优化”，*IEEE* 物联网期刊，第 9 卷，第 6 期，第 4385–4395 页，2022 年。[4]J. Xu 和 H. Wang，“无线联邦学习网络中的客户端选择与带宽分配：长期视角”，*IEEE* 无线通信汇刊，第 20 卷，第 2 期，第 1188–1200 页，2021 年。[5]S. Abdulrahman、H. Tout、A. Mourad 和 C. Talhi，“Fedmccs：用于优化物联网联邦学习的多准则客户端选择模型”，*IEEE* 物联网期刊，第 8 卷，第 6 期，第 4723–4735 页，2021 年。[6] T. Li、A. K. Sahu、M. Zaheer、M. Sanjabi、A. Talwalkar 和 V. Smith，“异构网络中的联邦优化”，机器学习与系统会议论文集，第 2 卷，第 429–450 页，2020 年。[7]Y. Jee Cho、J. Wang 和 G. Joshi，“理解联邦学习中的有偏客户端选择”，载于 第 *25* 届人工智能与统计国际会议论文集，2022 年。[8] T. Nishio 和 R. Yonetani，“移动边缘中面向异构资源的联邦学习客户端选择”，载于 *2019 IEEE* 国际通信会议（*ICC*）。IEEE，2019 年，第 1–7 页。[9] C. Palihawadana、N. Wiratunga、A. Wijekoon 和 H. Kalutarage，“FedSim：面向联邦学习的相似度引导模型聚合”，*Neurocomputing*，第 483 卷，第 432–445 页，2022 年。[10]T. Wang、Y. Liu、X. Zheng、H.-N. Dai、W. Jia 和 M. Xie，“分布式联邦学习的基于边缘的通信优化”，*IEEE* 网络科学与工程汇刊，2021 年。[11] Y. LeCun，“手写数字 MNIST 数据库”，*http://yann.lecun.com/exdb/mnist/*，1998 年。[12]A. Krizhevsky、G. Hinton 等，“从微小图像中学习多层特征”，2009 年。