文章编号:1007-5321(2025)05-0040-08 DOI:10.13190/j.jbupt.2024-122

基于半监督联邦学习的高效物联网入侵检测方法

韩昊霖, 王小娟
(北京邮电大学 电子工程学院, 北京 100876)

摘要: 在物联网(IoT)入侵检测领域中,联邦学习已成为实现模型权重集成更新的有效解决方案。这种分布式学习方法允许设备在本地训练模型,并将更新后的参数传输到中央服务器进行聚合。然而,现有基于联邦学习的入侵检测方法仍然存在局限性,在非独立同分布的数据及客户端模型异构的场景下,全局模型的入侵检测性能会受到严重影响。同时,传输模型参数导致的大量通信开销也阻碍了联邦学习方案的实际部署。为了解决上述问题,提出了一种基于半监督联邦学习的高效物联网入侵检测方法。通过利用未标记的公开数据增强模型对数据的理解能力,不断提高客户端分类器的性能,同时加入鉴别器模块提高客户端预测标签的质量,并通过硬标签策略和投票机制的结合有效降低通信开销。实验结果表明,在非独立同分布数据和客户端模型异构场景下,实现了 86.97% 的准确率,优于典型的联邦学习方法,同时实现了更低的通信开销。

关 键 词: 联邦学习; 入侵检测; 半监督学习; 知识蒸馏
中图分类号: TN918.91 **文献标志码**: A

An Efficient IoT Intrusion Detection Method Based on Semi Supervised Federated Learning

HAN Haolin, WANG Xiaojuan
(School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: In the field of Internet of things (IoT) intrusion detection, federated learning has become an effective solution for implementing model weight integration updates. This distributed learning method allows devices to train models locally and transmit updated parameters to a central server for aggregation. However, existing intrusion detection methods based on federated learning still have limitations. In scenarios with non-independent and identically distributed data and heterogeneous client models, the intrusion detection performance of the global model will be severely affected. The significant communication overhead caused by simultaneously transmitting model parameters also hinders the actual deployment of federated learning schemes. To address the aforementioned issues, an efficient IoT intrusion detection method based on semi supervised federated learning is proposed. By utilizing unlabeled public data to enhance the model's understanding of the data, the performance of the client classifier is continuously improved. At the same time, a discriminator module is added to improve the quality of the client's predicted labels, and the combination of hard label strategy and voting mechanism effectively reduces communication overhead. The experimental results show that an accuracy of 86.97% is achieved in non-independent and identically distributed data and heterogeneous client model scenarios, which is superior to typical federated learning methods and achieves lower communication overhead.

收稿日期: 2024-06-01
作者简介: 韩昊霖(2000—), 男, 硕士生。
通信作者: 王小娟(1985—), 女, 副教授, 博士生导师, 邮箱:wj2718@163.com。

Key words: federated learning; intrusion detection; semi supervised learning; knowledge distillation

物联网(IoT,Internet of things)作为物理对象和网络有效的连接方式之一,涉及嵌入式系统、无线网络、机器学习、自动化等许多领域的应用。与此同时,大量的物联网设备缺乏足够的安全防御,成为入侵者的目标,因此在物联网设备中应用入侵检测系统至关重要。随着深度学习技术的发展,基于深度学习的方法在流量分析和入侵检测领域得到广泛的应用[1],通常采用集中式深度学习技术,从各种物联网设备收集大量的流量数据,并将这些数据上传到中央服务器,训练用于入侵检测的深度神经网络模型。尽管这类方法可以在入侵检测任务中实现高精度,但存在着信息泄露的风险,即客户端的私有数据被上传到中央服务器,并可能被服务器滥用。为了解决上述问题,在入侵检测系统中开始采用联邦学习的框架[2]。在联邦学习框架中,物联网设备通过在中央服务器定期交换和聚合深度学习模型的参数和梯度,取代上传其原始数据,完成协同训练。然而由于传统的联邦学习方法需要客户端频繁上传其本地模型的参数和梯度,产生了较大的通信开销,严重阻碍了实际部署。同时,物联网设备通常规模较大,数量较多,意味着在应用联邦学习框架时,大量的物联网设备需要上传参数和梯度等信息,导致了更多的通信开销。此外,目前已经有可以从联邦学习上传的模型参数中恢复客户端原始数据的方法[3],这意味着联邦学习方法仍存在信息泄露的安全风险。

为了降低上述由于模型参数交换而产生的安全风险和通信开销,Chen 等[4]提出了联邦蒸馏方案,其核心思想是客户端通过将深度学习模型的输出替代模型参数上传到中央服务器,然后在中央服务器聚合后形成全局 logits,并将其分发到各个客户端,在下一轮联邦训练中使用。尽管联邦蒸馏方案降低了通信开销,在数据独立同分布(IID, independent and identically distributed)的场景下,实现了与传统联邦学习方法相近的精度,但在数据非独立同分布 non-IID 及客户端模型异构的场景下,其性能较差。

为了解决上述挑战,笔者提出了一种用于物联网入侵检测的半监督联邦学习方法。笔者的主要贡献可概括如下:

1)提出一种用于物联网入侵检测的高效半监督联邦学习方法,在 non-IID 数据和模型异构场景下有较好的入侵检测性能;

2)在联邦学习的客户端中加入鉴别器模块,用于判断对流量数据是否熟悉,有效提高了每个客户端预测标签的质量;

3)在联邦学习的客户端将预测标签上传时,采用硬标签和投票机制结合的方法,替代软标签和直接聚合,有效降低了通信开销。

1　相关工作

随着对入侵检测方法研究的深入,为了更好地分析复杂的网络流量,基于机器学习的方法首先被应用于入侵检测系统,尽管可以有效地基于流量的统计特征来检测未知入侵,但其通常严重依赖于特征工程,并且只能提取浅层特征。近年来,基于深度学习的入侵检测方法被提出且拥有更好的性能,然而这些方法依赖大量的流量样本来训练具有优异检测性能的模型,同时获取大量用户隐私敏感数据的难度也较高。

针对上述问题,相较于集中式深度学习需要收集用户的私有数据,基于联邦学习的入侵检测方法能够有效解决收集客户端隐私敏感数据的难题。近年来,由于具有保护隐私的优势,联邦学习方法在入侵检测系统和物联网中得到了广泛应用。Tran 等[5]通过并行计算加速计算性能,实现了一种具有强隐私保护的高效跨孤岛联邦学习方案,同时允许客户端在训练过程中退出和重新加入。Zhang 等[6]提出了一种半监督联邦学习方法,利用未标记的数据来训练基于一致性正则化的无监督模型,然后将无监督模型、有监督模型和全局模型聚合为一个新的全局模型。Lazzarini 等[7]提出了一种在物联网环境中实现入侵检测的方法,使用浅层人工神经网络作为共享模型,并使用联邦平均作为聚合算法,实现了较高的精度。以上研究表明,基于联邦学习的入侵检测方法不仅能够实现较高的检测精度,还能有效保护用户隐私,因此被视为集中式深度学习的有效替代方法。

然而,基于深度学习的模型具有大量的参数,这使客户端在联邦训练的过程中产生较大的通信开销。为了降低基于联邦学习的入侵检测方法的通信开销,研究人员提出了一些方法。Chai 等[8]提出了一种高效通信的联邦学习方法,通过基于分解的多目标优化算法优化全局模型的结构,使用高度可扩

展的编码方法,在不严重降低全局模型精度的情况下降低了通信开销。Jahani 等[9]提出了应用于联邦学习系统的新的安全聚合协议 SwiftAgg ++,中央服务器以保护隐私的方式聚合客户端的本地模型,在不影响安全性的情况下显著降低通信开销,实现最佳通信负载。尽管上述研究都在一定程度上降低了通信开销,但其实际效果仍然受到训练模型大小等因素的影响。在实际训练过程中,需考虑不同客户端模型异构对性能的影响。Jiang 等[10]通过自适应模型的修剪,提出了一种高效通信的联邦学习框架,从理论上分析了修剪率对训练性能的影响,采用在线学习算法自适应地确定异构客户端的不同修剪率,有效提高了异构客户端的计算和通信效率。

受基于蒸馏的联邦学习方法的启发,笔者提出了一种用于入侵检测的半监督联邦学习方法,该方法在每轮训练中通过传输未标记数据的预测标签,取代传统方法中的模型参数。此外,增加了鉴别器模块来提高每个客户端预测标签的质量。实验结果表明,笔者提出的方法在现实 non-IID 数据及客户端模型异构场景下的效果优于现有方法。

2 方法

图 1 和图 2 分别展示了笔者提出的基于半监督联邦学习的高效物联网入侵检测方法的客户端和中央服务器框架,由多个客户端和 1 个中央服务器组成。客户端首先从中央服务器下载未标记的公开数据,然后客户端使用本地标记的 non-IID 数据训练不同的分类器网络,由于分类器无法学习不在本地标记数据集中的攻击流量类别,因此它可能会对不熟悉的未标记流量数据作出错误的预测。因此,笔者提出的方法引入 1 个鉴别器,来区分客户端对未标记的流量数据是否熟悉,将不熟悉的流量数据标记,以提高客户端预测的软标签质量。同时,为了进一步降低通信开销,客户端将软标签转换为硬标签后,上传到中央服务器。在中央服务器收集了所有客户端的预测标签之后,通过投票机制来确定未标记流量数据的全局标签。最后,将全局硬标签分发给每个客户端,用于客户端训练下一轮的分类器网络。

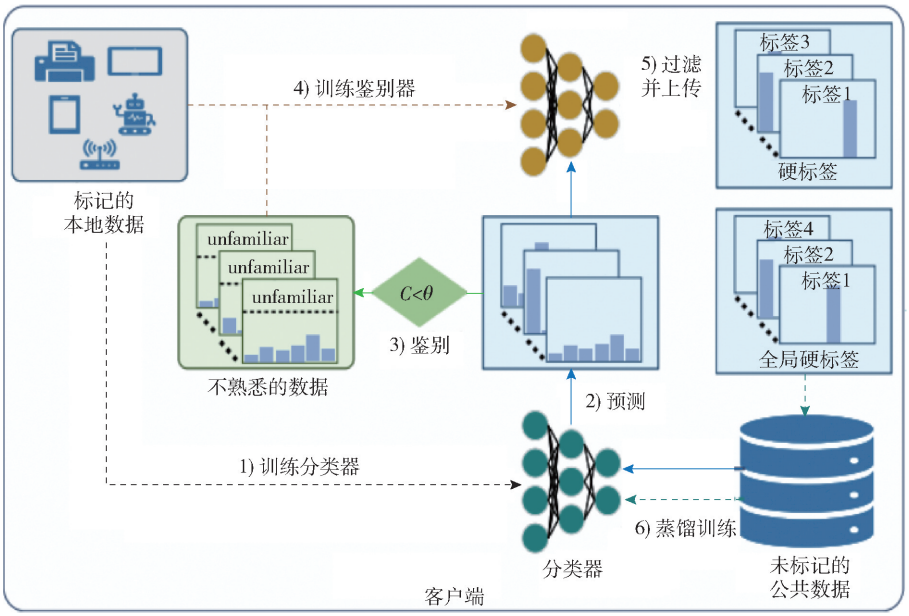

图 1 客户端框架

联邦蒸馏方法通过聚合每个类的 logits 来利用客户端的知识,因此只能增强每个客户端中相应类的训练效果,这意味着当客户端缺乏某些类的样本时,无法实现对所有类的有效分类。笔者通过加入 1 个未标记的公开数据集来解决这个问题,可以利用全局 logits 来识别未标记数据集中的每个样本属于哪个类。此外,由于在 non-IID 的数据场景下,客户端很难在未标记的公开数据集上作出正确预测,笔者通过多种机制结合来共同提高预测标签的质量。

在笔者提出的方法中,假设有 K 个客户端,每个客户端 $k \in \{1,2,\cdots,K\}$ 有以下 2 个数据集:

本地私有标记的数据集为

$$D^{k,c} = \{(x_i^{k,c}, y_i^{k,c}) \mid i = 1,2,\cdots,n\}$$

在所有客户端上共享的未标记的公开数据集为

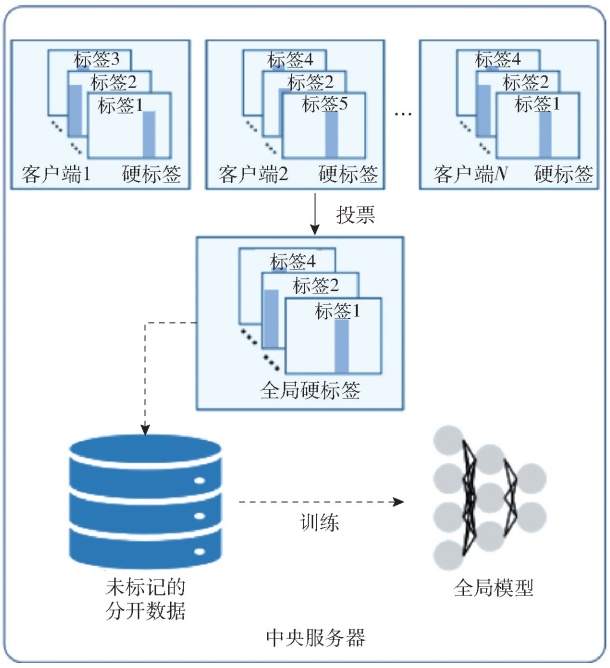

图 2　中央服务器框架

$$D^o = \{x_j^o \mid j=1,2,\cdots,N^o\}$$

对于 L 分类任务，$y_i^{k,c}$ 为独热编码。在每个客户端都有 1 个用本地私有数据集训练的分类器模型 $w^{k,c}$ 和鉴别器 $w^{k,d}$，笔者提出的方法具体步骤如下。

1）训练分类器：在每个客户端 k 上，用其本地标记的私有数据集训练客户端分类器模型 $w^{k,c}$，此过程表示为

$$w^{k,c} \leftarrow w^{k,c} - \gamma \nabla \varphi(\hat{Y}^{k,c}, Y^{k,c}) \tag{1}$$

其中：$\varphi(\cdot,\cdot)$ 表示损失函数，该损失函数在分类器训练时被最小化，$\hat{Y}^{k,c}$ 代表分类器模型函数 F 的输出，即 $\hat{Y}^{k,c} = F(X^{k,c} \mid w^{k,c})$，$\gamma$ 表示学习率。对于多分类任务，损失函数可以是交叉熵函数，然后将 D^o 中的每个样本都通过分类器，分类器对其进行预测，得到样本的置信度得分，表示为

$$c_j^{k,o} = \max(\hat{p}_j^k) = \max(F(x_j^o \mid w^{k,c})) \tag{2}$$

2）训练鉴别器：首先创建 1 个空集 $D^{k,d} = \phi$，如果 x_j^o 的置信度得分小于阈值 θ，则在集合中添加该样本，并且该样本将被鉴别为客户端 k 的“不熟悉”样本，此过程表示为

$$D^{k,d} = D^{k,d} \cup \{(x_j^o, [0,1]^{\mathrm{T}}) \mid c_j^{k,o} < \theta\} \tag{3}$$

其中：使用 $[0,1]^{\mathrm{T}}$ 表示“不熟悉”标签，本地私有数据集中的每个样本对客户端 k 来说都是熟悉的，其所有样本都被添加到 $D^{k,d}$ 中，并且标签为“熟悉”，此过程表示为

$$D^{k,d} = D^{k,d} \cup \{(x_i^{k,c}, [1,0]^{\mathrm{T}}) \mid i=1,2,\cdots,N^{k,c}\} \tag{4}$$

其中：使用 $[1,0]^{\mathrm{T}}$ 表示“熟悉”标签，在该步骤完成后，$D^{k,d}$ 数据集包含 2 个类别的数据，然后用 $D^{k,d}$ 数据集来训练鉴别器 $w^{k,d}$。

3）过滤和上传：利用第 2）步经过训练的鉴别器，将未标记的公开数据集中样本的预测进行过滤，首先计算鉴别结果为

$$d_j^{k,o} = \operatorname{argmax}(F(x_j^o \mid w^{k,d})) \tag{5}$$

其中：$F(\cdot \mid w^{k,d})$ 的输出是 1 个 2 维向量，将 $\operatorname{argmax}(\cdot)$ 看作向量最大值的索引，如第 2）步所述，如果 $d_j^{k,o} = 0$，则客户端 k 熟悉流量样本 x_j^o，否则不熟悉该流量样本。对于 D^o 中的每个流量样本 x_j^o，按如下方式对预测结果进行过滤：

$$\hat{p}_j^k = \begin{cases} \operatorname{argmax}(F(x_j^o \mid w^{k,c})), & d_j^{k,o} = 0 \\ -1, & d_j^{k,o} = 1 \end{cases} \tag{6}$$

其中：使用“-1”表示对流量样本不熟悉，然后每个客户端将未标记的公开数据的硬标签，即 $\{\hat{p}_j^k \mid j=1,2,\cdots,N^o\}$ 上传到中央服务器。

4）投票和分发：对于 1 个流量样本 x_j^o，有 L 个投票集合，即 $\{V^{j,0}, V^{j,1}, \cdots, V^{j,L-1}\}$，“$-1$”类别代表不熟悉的样本，因此不包括在 L 个投票集合中。对于样本的预测标签 $\hat{p}_j^k$，如果预测结果 $\hat{p}_j^k \neq -1$，则该预测标签被添加到相应的投票集中。在所有客户端将对熟悉样本的预测标签输入到相应的投票集合中后，中央服务器根据每个投票集合中的投票数，确定该样本的全局硬标签，此过程表示为

$$\hat{p}_j^s = \operatorname{argmax}([\,|V^{j,0}|, |V^{j,1}|, \cdots, |V^{j,L-1}|\,]) \tag{7}$$

其中：矩阵 $\hat{\boldsymbol{P}}^s$ 代表连接的 $\{\hat{p}_j^s \mid j=1,2,\cdots,N^o\}$，最后将全局硬标签 $\hat{\boldsymbol{P}}^s$ 分发给每个客户端。

5）蒸馏：每个客户端将自己视为“学生”，接收来自中央服务器的全局硬标签，将其作为“教师”，每个本地客户端模型使用带有全局硬标签的公开数据进行蒸馏训练，此过程表示为

$$w^{k,c} \leftarrow w^{k,c} - \gamma \nabla \varphi(\hat{P}^{k,c}, \hat{\boldsymbol{P}}^s) \tag{8}$$

其中：$\hat{P}^{k,c} = F(X^0 \mid w^{k,c})$，中央服务器的分类器模型也使用 $\hat{\boldsymbol{P}}^s$ 进行训练并被用于评估。

在模型异构的设置上，采用如表 1 所示的卷积神经网络模型、ResNet 模型、MobileNet 模型、EfficientNet 模型。

在联邦学习训练中，域偏移导致同一标签 Y 在不同域中具有不同的特征 X。因此，logits 输出在不

表1 基于卷积神经网络的检测模型

层	单元	输出大小
输入层	输入	(80,23,5)
卷积层1~4	1维卷积层(64,3,1)	(80,64,5)
卷积层5~6	1维卷积层(128,3,1)	(80,128,5)
卷积层7	1维卷积层(128,3,2)	(80,128,3)
卷积层8	1维卷积层(128,3,2)	(80,128,2)
全连接层	线性层	(80,128)
输出层	输出	(80,11/2)

注:实验中batch大小设置为80,学习率设置为0.000 1,epoch设置为5,分类器网络和鉴别器网络最后1层分别具有11个和2个神经元。

同域上沿批处理维度的分布并不相同,logits输出的不同维度对应于不同的类,需要保持相同维度的相关性和不同维度的去相关性,因此构造互相关矩阵至关重要,如图3所示。具体来说,得到第 i 个参与客户端的logits输出为

$$Z_i = f(\theta_i, X_0) \in R^{N_0 \times C} \tag{9}$$

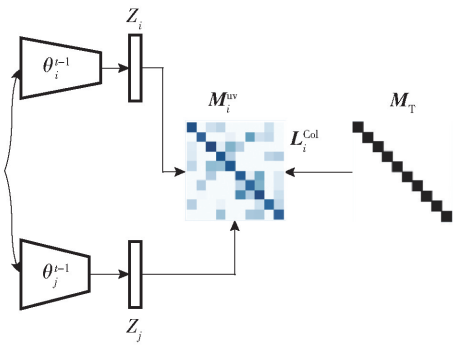

图3 互相关矩阵构造

对于第 i 个和第 j 个参与客户端,未标记公共数据上的logits输出为 Z_i 和 Z_j。考虑到中央服务器的计算负担,计算平均logits输出为

$$\overline{Z} = \frac{1}{K} \sum_i Z_i \tag{10}$$

然后计算第 i 个参与客户端的互相关矩阵 $\boldsymbol{M}_i$,其平均logits输出为

$$\boldsymbol{M}_i^{uv} \triangleq \frac{\sum_b \| Z_i^{b,u} \| \, \| \overline{Z}^{b,v} \|}{\sqrt{\sum_b \| Z_i^{b,u} \|^2} \sqrt{\sum_b \| \overline{Z}^{b,v} \|^2}} \tag{11}$$

其中:b 表示批样本,u,v 表示logits输出的维度,$\| \cdot \|$ 是沿批维度的归一化操作,$\boldsymbol{M}_i$ 是1个输出维数为 C 的矩阵,值在 -1 和1之间。第 i 个参与客户端的协作损失定义为

$$L_i^{\mathrm{Col}} \triangleq \sum_u (1 - \boldsymbol{M}_i^{uv})^2 + \lambda_{\mathrm{Col}} \sum_u \sum_{u \neq v} (1 + \boldsymbol{M}_i^{uv})^2 \tag{12}$$

其中:λ_{Col} 是1个用于权衡第1项和第2项损失重要性的常数,当互相关矩阵的对角项取值为+1时,鼓励不同参与客户端的logits输出相似;当互相关矩阵的非对角项取值为-1时,因为这些logits输出在不同维度上彼此不相关,鼓励logits输出不同,从而实现相同维度的相关性和不同维度的去相关性。

3 实验

3.1 数据集介绍及预处理

本实验在基于网络的物联网检测系统(N-BaIoT,network based detection of IoT)公开数据集(引用网址 https://archive.ics.uci.edu/ml/datasets/detection_of_IoT_botnet_attacks_N_BaIoT)上进行,包含来自9个物联网设备的9个子数据集,其中7个物联网设备有11类流量(1类正常流量和10类攻击流量),2个物联网设备有6类流量(1类正常流量和5类攻击流量)。每个流量包含从最近5个不同的时间窗口中(100 ms,500 ms,1.5 s,10 s,1 min)提取的115个特征。N-BaIoT数据集具有全面的流量类别和大量的流量记录,被广泛应用于入侵检测领域。

本实验分3个步骤对数据进行预处理,包括数据集划分、特征值归一化、二维化。

1)数据划分:在现实世界中,通常有大量拥有non-IID数据的客户端参与到联邦训练中。因此,本实验对N-BaIoT原始数据集进行划分,以更符合现实数据分布情况。

使用原始数据集进行训练需要大量的计算资源和基础设施,本实验使用从N-BaIoT数据集提取的子集mini-N-BaIoT,由9个物联网设备的11类流量组成。首先,将原始数据集中9个物联网设备的子数据集划分为 $D_{d1}, D_{d2}, \cdots, D_{d9}$,$D_{di}$ 根据业务类别划分为 L_{di} 个子集,每个子集 $D_{di,l}$ 只包含类别为 l 的流量数据,在每个子集 $D_{di,l}$ 中选择1 000条流量数据作为mini-N-BaIoT数据集的子集 $D_{di,l}^{\mathrm{mini}}$,按7:1:2的比例划分为私有数据集 D^p、公开数据集 D^o 和测试数据集 D^{test}。3个数据集互不相交,且 D^o 没有标签。

接下来,将 D^p 分发给 K_{di} 客户端。D^p 按照标签进行分类,并被划分为大小为 $|D^p|/(2K_{di})$ 的 $2K_{di}$ 碎片,每个客户端分配其中的2个碎片。在本实验中,

K_{di}的值与 L_{di}相等，如果 D_{di}包含 11 个类别的流量，K_{di}为 11；如果 D_{di}包含 6 个类别的流量，K_{di}为 6。

在完成数据划分后，本实验中共划分了 89 个客户端，每个客户端的私有数据只来自 1 个设备，这与现实世界的物联网设备分布是一致的。且不同客户端拥有的数据量和数据类别有很大不同，满足 non-IID 的数据分布条件。

2）特征值归一化：由于流量数据具有不同维度的特征，数据集间维度的值差异很大，为了更有效地训练模型，使用 min-max 归一化将所有特征值都缩放到[0,1]区间。

3）二维化：每个流量样本 $\boldsymbol{x}_i$ 有 115 个特征，根据时间窗口分为 5 个部分，可以将样本的特征向量转移到 1 个 5 列 23 行的矩阵中，作为分类器模型的输入。

$$\boldsymbol{x}_i=\begin{bmatrix} x_{i,0} & x_{i,23} & x_{i,46} & x_{i,69} & x_{i,92} \\ x_{i,1} & x_{i,24} & x_{i,47} & x_{i,70} & x_{i,93} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{i,22} & x_{i,45} & x_{i,68} & x_{i,91} & x_{i,114} \end{bmatrix} \tag{13}$$

3.2　实验设置

在实验环境方面，所有实验都是基于 Python 3.8.10，PyTorch 框架版本为 1.9.0，在带有 Intel(R) Xeon(R) Silver 4216 中央处理器@2.10 GHz、GeForce RTX2080 图形处理器的服务器上运行。使用 Adam 优化器训练第 2 节中的分类器模型，在每个客户端上判断不熟悉样本的阈值 θ，被设置为客户端预测概率的中值。

为了评估本实验的性能，统计真阳性 T_{p}、真阴性 T_{n}、假阳性 F_{p}、假阴性 F_{n} 的数量，从而进一步计算准确率 A_{Accuracy}、精确率 $P_{\mathrm{Precision}}$、召回率 R_{Recall}和 F_1 值，计算公式为

$$A_{\mathrm{Accuracy}}=\frac{T_{\mathrm{p}}+T_{\mathrm{n}}}{T_{\mathrm{p}}+F_{\mathrm{p}}+F_{\mathrm{n}}+T_{\mathrm{n}}} \tag{14}$$

$$P_{\mathrm{Precision}}=\frac{T_{\mathrm{p}}}{T_{\mathrm{p}}+F_{\mathrm{p}}} \tag{15}$$

$$R_{\mathrm{Recall}}=\frac{T_{\mathrm{p}}}{T_{\mathrm{p}}+F_{\mathrm{n}}} \tag{16}$$

$$F_1=2\frac{P_{\mathrm{Precision}}R_{\mathrm{Recall}}}{P_{\mathrm{Precision}}+R_{\mathrm{Recall}}} \tag{17}$$

3.3　入侵检测性能实验

在入侵检测性能方面，将笔者提出的方法与以下几种基线方法进行比较。联邦平均算法(FedAvg，federated averaging algorithm)是联邦学习中最常见的方法，直接平均聚合不同客户端上传的模型参数。在联邦蒸馏(FD，federated distillation)方法中，客户端将模型参数的共享替换为每类数据的 logits 共享。Itahara 等[11]提出了一种基于蒸馏的半监督联邦学习方法(DSFL，distillationbased semisupervised federated learning)，通过交换模型的输出对未标记的公开数据进行标记。与 DSFL 方法相比，笔者提出的方法有 3 个关键区别，一是通过加入鉴别器模块提高上传标签的质量，二是通过硬标签策略和全局投票机制相结合有效降低通信开销，三是通过互相关学习提高模型异构场景下的检测性能。

入侵检测性能实验结果如表 2 所示，从表 2 中可以观察到，笔者提出的方法明显优于其他基线方法。在联邦训练框架中，FD 和 DSFL 都没有实现高检测精度。具体来说，FD 检测准确率为 47.84%，主要是因为当数据分布是 non-IID 时，客户端本地数据不能代表全局数据，每个客户端的流量类别较少，FD 的最佳检测性能是由单个客户端训练获得，可以认为 FD 方法在 non-IID 的数据分布下是无效的。DSFL 的准确率高于 FD，表明在联邦学习框架中引入未标记的公开数据，提高了在 non-IID 的数据分布下基于蒸馏的联邦学习方法的可行性。但本实验场景中更不均匀的数据分布和客户端模型的异构，使得 DSFL 无法准确预测未标记的公开数据的标签，导致最终模型的检测性能还有提升的空间。作为一种传统的联邦训练方法，FedAvg 实现了良好的检测性能，但巨大的通信开销不可忽视，这将在 3.4 小节详细讨论。同时，FedAvg 方法也存在梯度泄露的风险，即攻击者可以通过客户端上传的模型参数恢复原始数据[3]，意味着该方法并不安全。笔者提出的方法在通信时从每个客户端上传硬标签，而不是梯度或参数，这使得从上传的梯度或参数中恢复客户端本地数据的攻击方法变得不可行。

表 2　不同方法入侵检测性能比较　　%

方法	准确率	精确率	召回率	F_1值
FedAvg	79.21	78.13	76.26	77.18
FD	47.84	44.25	45.57	44.90
DSFL	59.28	62.98	61.42	62.19
笔者方法	86.97	88.24	84.31	86.23

笔者提出的方法实现了更有效的联邦训练，主要有以下原因：1）与 DSFL 类似，笔者方法利用未标记的公开数据，能够更好地适应 non-IID 的数据分

布情况;2)每个客户端上的鉴别器可以显著提高预测标签的质量;3)中央服务器的投票机制可以有效地聚合来自不同客户端的训练结果;4)联邦互相关学习可以有效解决客户端模型异构的问题。

通过上述分析可以得出结论,笔者提出的方法具有良好的检测性能,适用于物联网中的联邦训练和入侵检测。

3.4 通信效率实验

在通信效率方面,表3和表4与3.3小节中的基线方法进行了训练速度和通信开销的比较。总体来说,笔者方法实现了更快的训练过程和更低的通信开销。

表3 不同方法在不同通信轮数下准确率比较 %

方法	不同通信轮数下的 A_{Accuracy}				
	10	50	100	150	200
FedAvg	15.27	26.22	38.47	47.85	56.26
FD	45.74	46.35	47.21	47.84	47.84
DSFL	47.76	59.28	59.28	59.28	59.28
笔者方法	74.25	82.84	85.21	86.63	86.97

表4 不同方法通信开销和准确率的比较

方法	$C@D^o$	通信开销/MB@A_{Accuracy}/%			
		$C@50$	$C@75$	$C@$Top-Acc	Top-Acc
FedAvg	–	137.00	772.00	1 353.00	79.21
FD	–	–	–	0.04	47.84
DSFL	1.20	17.30	–	22.60	59.28
笔者方法	1.20	0.10	0.20	0.70	86.97

注:$C@D^o$ 是下载公开数据集需要的通信开销,$C@A$ 是达到 A 准确率所需要的累计通信开销。

在训练速度方面,从表3中可以看到,笔者提出的方法在10轮通信后就可以实现较高的准确率,并在180轮左右达到收敛。FedAvg方法的准确率在200轮通信时仍未收敛,这表明通过直接平均聚合参数进行联邦训练的速度较慢。此外,FD和DSFL很快达到检测性能的上限,无法通过增加通信轮数进一步提高检测性能。

表4列出了每个方法的通信开销。在FedAvg和FD方法中,不涉及将公开数据集分发给每个客户端。由于FedAvg方法在每一轮中传递模型的参数,因此较大的模型会导致巨大的通信开销。其他3种基于蒸馏的方法的通信开销仅取决于模型输出的维度,而不根据模型大小进行扩展,因此这些方法的通信成本相对较低。此外,在笔者提出的方法中,客户端将本地预测转为硬标签后上传,在DSFL方法中客户端需要上传本地预测向量,因此能有效降低通信开销。由于缺乏处理non-IID数据分布的有效机制,DSFL和FD方法都没有获得较高的检测性能。从表4可以得出结论,由于同时具有较高的检测准确率和通信效率,笔者提出的方法具有最佳的整体性能。

3.5 消融实验

为了进一步检测笔者提出的方法中,每个组成部分给性能带来的增益,笔者进行了消融实验。

首先在笔者方法中分别去除鉴别器模块、投票机制以及2者同时去除。在没有鉴别器时,客户端对所有未标记的流量样本进行预测,即同时包括熟悉和不熟悉的样本。如图4所示,实验结果表明,在所有情况下取消鉴别器模块会导致性能显著下降,显然鉴别器的使用对检测性能的影响最大,这说明加入鉴别器模块对提高每个客户端预测标签质量的必要性。对投票机制的消融实验研究表明,当客户端作出过多错误的预测时,投票机制会使模型的性能进一步降低,因此投票机制需要与提高客户端预测标签质量的鉴别器结合使用。综上所述,可以得出结论,鉴别器和投票机制的结合能够大幅提升笔者方法的检测性能。

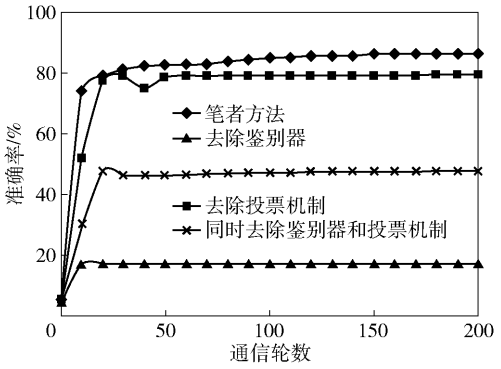

图4 笔者方法鉴别模块和投票机制消融实验的准确率曲线

笔者继续研究了硬标签策略对检测性能和通信开销的影响,采用硬标签策略不是为了提高检测性能,而是为了减少笔者方法在联邦训练过程中的通信开销。对于每个流量样本,样本的软标签是1个概率向量,其中每个数字都是双精度浮点数。在客户端上传该向量之前,可以将该向量中的每个双精度浮点数保留特定的小数位数,从而减少软标签占用的内存大小。实验结果如图5和图6所示,保留

8,6,4,2 位小数达到的检测精度基本相同,但通信开销截然不同。实验结果表明,保留的小数位数越少,通信开销越低。因此,在笔者方法中上传硬标签在保证较高的检测性能的前提下,可以大大降低通信开销。

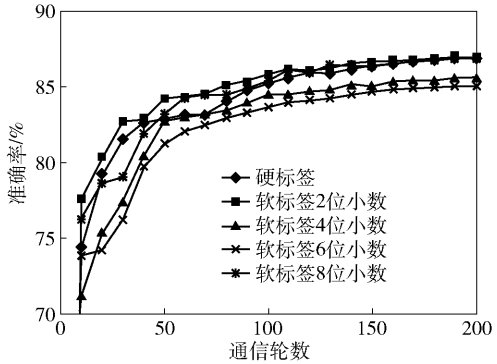

图 5　测试不同标签策略的准确率曲线

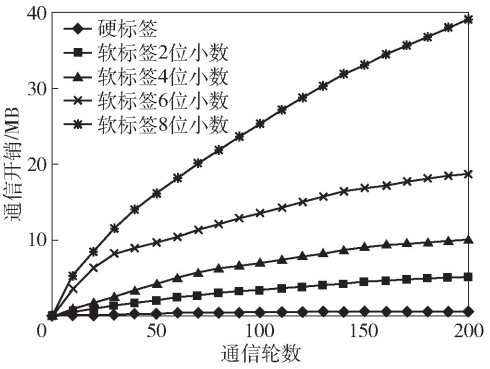

图 6　测试不同标签策略的通信开销曲线

4　结束语

笔者提出了一种用于物联网入侵检测的高效半监督联邦学习方法,核心是通过利用未标记的公开数据,增强模型对数据的理解能力,进而提高客户端分类器的性能。通过引入鉴别器,提高每个客户端预测标签的质量,有效避免了在 non-IID 数据分布下客户端作出大量错误预测,进而影响联邦训练的效果。此外,笔者还通过客户端上传硬标签的策略和中央服务器的投票机制相结合,进一步降低联邦训练过程中的通信开销。同时,笔者在实验过程中模拟了真实物联网环境下的联邦训练,设置了较多的客户端设备、不均匀的流量数据分布和不同的客户端训练模型。实验结果表明,笔者提出的方法可以在联邦训练中获得更好的检测性能和更低的通信开销,满足物联网环境中联邦训练和入侵检测所需要的安全性、准确性和高效性。

参考文献:

[1] THAKKAR A, LOHIYA R. Fusion of statistical importance for feature selection in deep neural network-based intrusion detection system [J]. Information Fusion, 2023, 90: 353-363.

[2] NGUYEN T, THAI M T. Preserving privacy and security in federated learning[J]. IEEE/ACM Transactions on Networking, 2023, 32: 833-843.

[3] KARIYAPPA S, GUO C, MAENG K, et al. Cocktail party attack: Breaking aggregation-based privacy in federated learning using independent component analysis [C]// International Conference on Machine Learning, PMLR, 2023: 15884-15899.

[4] CHEN Z, TIAN P, LIAO W, et al. Resource-aware knowledge distillation for federated learning[J]. IEEE Transactions on Emerging Topics in Computing, 2023, 11: 706-719.

[5] TRAN H Y, HU J, YIN X, et al. An efficient privacy-enhancing cross-silo federated learning and applications for false data injection attack detection in smart grids[J]. IEEE Transactions on Information Forensics and Security, 2023, 18: 2538-2552.

[6] ZHANG Z, MA S, YANG Z, et al. Robust semisupervised federated learning for images automatic recognition in Internet of drones[J]. IEEE Internet of Things Journal, 2022, 10(7): 5733-5746.

[7] LAZZARINI R, TIANFIELD H, CHARISSIS V. Federated learning for IoT intrusion detection[J]. AI, 2023, 4(3): 509-530.

[8] CHAI Z, YANG C, LI Y. Communication efficiency optimization in federated learning based on multi-objective evolutionary algorithm [J]. Evolutionary Intelligence, 2023, 16(3): 1033-1044.

[9] JAHANI-NEZHAD T, MADDAH-ALI M A, LI S, et al. SwiftAgg + +: Achieving asymptotically optimal communication loads in secure aggregation for federated learning [J]. IEEE Journal on Selected Areas in Communications, 2023, 41(4): 977-989.

[10] JIANG Z, XU Y, XU H, et al. Computation and communication efficient federated learning with adaptive model pruning[J]. IEEE Transactions on Mobile Computing, 2024, 23(3): 2003-2021.

[11] ITAHARA S, NISHIO T, KODA Y, et al. Distillation-based semi-supervised federated learning for communication-efficient collaborative training with non-IID private data[J]. IEEE Transactions on Mobile Computing, 2021, 22(1): 191-205.