# 面向分类任务的隐私保护协作学习技术*

黎兰兰，张信明
（中国科学技术大学　计算机学院，安徽　合肥　230026）

**摘　要**：随着相关法律法规的发布和人们隐私意识的觉醒，隐私保护要求不断提高。针对分类任务场景，提出了一种隐私性与效用性并重的面向分类任务的隐私保护协作技术（PCTC）。在隐私安全方面，将同态加密和差分隐私相结合，并设计了一种安全聚合机制，实现更加健壮的隐私保护；在效用性方面，引入信息熵的计算因子对标签可信度进行度量，降低标注噪声对模型性能的影响。最后对所提出的方案进行了安全性分析，并在公开数据集上进行了实验验证。结果表明 PCTC 在保证数据隐私安全的同时大幅度提升了模型性能，具有较为广阔的应用前景。

**关键词**：隐私保护；数据标注；分类任务；同态加密；差分隐私
**中图分类号**：TP393　　**文献标识码**：A　　**DOI**：10. 19358/j. issn. 2097 - 1788. 2023. 05. 007
**引用格式**：黎兰兰，张信明. 面向分类任务的隐私保护协作学习技术［J］. 网络安全与数据治理，2023，42(5)：36 - 43.

# Privacy-preserving collaborative learning technology for classification

Li Lanlan，Zhang Xinming
（School of Computer Science and Technology，University of Science and Technology of China，Hefei 230026，China）

**Abstract**：With the release of relevant laws and regulations and the awakening of people's privacy awareness，the requirements for privacy protection are constantly increasing. Aiming at the scenario of classification，this paper proposes a Privacy-preserving Collaborative Learning Technology for Classification（PCTC）that emphasizes both privacy and utility. In terms of privacy，homomorphic encryption and differential privacy are combined and a secure aggregation mechanism is designed to achieve more robust privacy protection. In terms of utility，the calculation factor of information entropy is introduced to measure the credibility of labels，which can reduce the impact of labeling noise on model performance. Finally，the security analysis of the proposed scheme is carried out，and the experiments are implemented on public datasets. The results show that PCTC significantly improves model performance while ensuring privacy and security of the data，and has broad application prospects.

**Key words**：privacy preservation；data labeling；classification task；homomorphic encryption；differential privacy



## 0　引言

近年来，随着数据产生速度和计算机算力的持续提升，机器学习在目标识别、语音识别、自然语言处理和对象检测等许多领域都取得了巨大突破[1]。新兴的机器学习尤其是深度学习更是为产业的升级和变革提供了推动力量，其中包括智慧农业、智慧医疗等行业[2]。良好的机器学习框架特别是有监督的人工神经网络往往需要大量的标注数据，然而现实中任何单一实体都不可能总是拥有全部标注数据[3]，多方协作学习是解决这一问题的有效方案。

传统的多方协作学习方法是将多方的私有原始数据都共享到一个中央服务器进行集中处理，共同训练指定模型。不幸的是，这种直接共享原始数据的方法存在极大的隐私泄露隐患[4]。一方面，收集的数据可能包含用户的隐私信息，例如身份信息等；另一方面，收集的数据可能是敏感的，如金融数据、军事数据等。在数据共享的过程中，数据拥有者无法限制数据的使用，而且在数据传输、存储和处理过程中很容易受到第三方攻击[5]。

随着相关法律法规的发布和人们隐私意识的提升，

*基金项目：国家重点研发计划（2020YFB2103803）

如何在保护各方数据隐私安全的同时充分发挥数据的要素价值是各类主体面临的重要课题[6]。在现实生活的许多场景下，参与方可能不希望披露私人信息，导致无法受益于大规模的机器学习。例如，当事各方可以是旨在利用其持有的敏感患者信息进行合作研究的医疗机构。由于数据不能共享，各方只能通过本地私有数据进行训练，在这种数据规模较小的情况下训练的模型易出现过拟合，导致泛化能力差、鲁棒性差等问题。为了解决上述问题，本文分析了多方协作学习中隐私保护的重难点，提出了一种面向分类任务的隐私保护技术 PCTC。该技术基于众包机制，不再直接共享原始数据，而是通过使用参与方本地模型对待标记样本进行标注的方法来传递知识，以此来达到各方之协作学习的目的。并且综合考虑隐私性和效用性，在保护数据和模型隐私的同时，提高模型的再训练精度。

本文主要贡献如下：

（1）将同态加密和差分隐私相结合，实现即使攻击者通过某种手段不断查询也无法推断各参与方的隐私数据的技术。

（2）针对分类任务中预测向量结构单一易被攻击和浮点数加密开销大的问题，引入随机种子设计了一种安全高效的聚合机制。

（3）对于众包机制下因标注者水平导致的标注结果准确性未知的问题，提出了基于信息熵的模型学习算法（Information-Entropy Loss，IEL），使模型能够聚焦于可信度更高的样本，从而提高模型再训练的有效性和鲁棒性。

（4）设计了基于众包机制的协作系统，并通过实验和安全性分析证明了 PCTC 算法在分类任务中能够在保护数据隐私的同时提高模型的训练效果，具有广泛的应用前景。

## 1 相关工作

传统的多方协作学习方法采用集中式学习，然而受到相关法律和道德的约束，集中式的方法逐渐变得不可行，隐私保护的多方协作学习已成为新的发展趋势。目前，许多研究致力于实现多方协作学习下的隐私保护，主要包括联邦学习[7]和基于众包标注的协作学习[8]，这两种方法都存在各自的优点和不足。联邦学习一直被用作协作学习的方法被广泛使用，但存在不适用于异构的模型集合并且通信开销较高等问题[9]，众包标注的方法通过传递标注来传递知识可以避免数据出本地，是现在各类市场主体常用的方式[10]，但是标注数据容易被攻击导致用户的隐私泄露[3]。攻击者可以通过差分攻击获取本地分类器的标注，进而推断出原始数据的隐私信息，如成员推断和属性推断[11]。在众包分类任务下，标注向量结构单一，一方面标注向量容易受到选择明文攻击，另一方面半可信第三方也会对敏感数据产生好奇心[12]。

针对分布式协作学习面临的安全威胁，学者们研究了各种防护方法来提高协作学习的安全性。研究者们利用同态加密算法对参与方的数据进行加密，服务器对密文进行聚合以防止其提取参与方隐私。其中，文献［13］、文献［14］中提出的方法是所有参与方共享公钥和私钥，参与方用公钥对需上传的数据进行加密，聚合密文后再用私钥各自解密得到计算结果。这种方法虽然可行性高，但是由于参与方之间共享私钥，所以不适用于存在恶意参与方的场景，容易受到勾结攻击和差分攻击。为解决上述问题，文献［15］采用二次加密的方式进行改进，先用各自公钥加密再用共享公钥加密来防止恶意勾结。同样地，文献［16］为进一步保护用户隐私不被窃取，在整个过程中采用全同态加密，不存在明文信息。这两种方式都可以提升系统安全性，但是会造成计算开销显著增加。CaPC（Confidential and Private Collaborative）[3]是目前广泛使用的协作学习方法，它将混淆电路和差分隐私相结合提升系统在面对第三方攻击时的鲁棒性，但未考虑差分噪声对标签的影响。

## 2 预备知识

### 2.1 同态加密

与一般加密算法相比，同态加密可以将在密文进行的运算直接映射到明文，利用同态加密技术可以将密文数据直接交给无密钥第三方处理而不泄露信息，在多方协作学习中这种方式既可以减少通信代价，又可以转移计算任务平衡计算代价。目前同态加密算法可分为：部分同态机密、类同态加密和全同态加密。一般而言，全同态加密算法功能更加强大，支持各种计算任务，但其计算复杂度很高，目前仍处于研究阶段。相较于全同态加密，部分同态加密计算效率更高，在实际中使用最多。

本文根据分类任务需求选择开销较小的分布式部分同态加密 Paillier 算法来保护数据安全，该方法是 ISO 同态加密国际标准中唯一指定的加法同态加密算法。Paillier 具体加解密流程如下：

（1）生成密钥 KeyGen（）→（pk，sk）；

随机选择两个大素数 $p$、$q$，计算 $n=p\times q$ 和 $\lambda=\mathrm{lcm}(p-1,q-1)$，lcm 计算最小公倍数。选择随机数 $g$，$g\in\mathbb{Z}_{n^2}^{*}$ 且满足 $\mu=(L(g^{\lambda}\bmod n^2))^{-1}$ 存在，其中函数 $L(x)=(x-1)/n$。此时公钥为（$n$，$g$），私钥为（$\lambda$，$\mu$）。

（2）数据加密 Encrypt（pk，$m$）→$c$；

明文 $m\in\mathbb{Z}_n$，选择随机数 $r<n$，加密算法如下：

$$c = g^m r^n \ (\mathrm{mod}n^2) \tag{2}$$

（3）解密密文 Decrypt（sk，$c$）$\rightarrow m$；

$$m = \mathrm{L}\ (c^\lambda \mathrm{mod}n^2) \times \mu(\mathrm{mod}n) = \frac{\mathrm{L}\ (c^\lambda \mathrm{mod}n^2)}{\mathrm{L}\ (g^\lambda \mathrm{mod}n^2)} \mathrm{mod}n \tag{3}$$

### 2.2 差分隐私

**定义 1** 差分隐私（Differential Privacy，DP）

域$\mathbb{N}^{|x|}$上的随机算法 $\mathcal{M}$ 如果对于所有的 $\mathcal{S} \in \mathrm{Range}(\mathcal{M})$，并且任意相邻数据库 $d$，$d' \in \mathbb{N}^{|x|}$即$\|d - d'\| \leqslant 1$，若满足以下不等式：

$$\Pr[\mathcal{M}(D) \in \mathcal{S}] \leqslant e^{\varepsilon} \Pr[\mathcal{M}(d' \in \mathcal{S})] + \delta \tag{4}$$

则称它是满足（$\varepsilon$，$\delta$）－DP 的。其中 $\varepsilon$ 代表隐私预算，$\delta$ 是松弛项表示可以容忍的误差（一般设置为 $1 \times 10^{-5}$）。

为了度量隐私泄露程度，Dowrk 等人[17]提出了差分隐私，该技术最初是为数据库查询隐私安全设计的，后续随着机器学习的发展被广泛使用在模型的隐私保护上，目前主要分为本地差分隐私和中心化差分隐私。本文在参与方本地标注中添加满足（$\varepsilon$，$\delta$）－DP 的高斯噪声来保护参与者的隐私，防止攻击者进行差分攻击从聚合标签中推断参与方的隐私信息。进一步地，考虑到差分隐私引入的浮点数噪声，本文设计了一种安全高效的聚合机制，可以有效减少浮点数加解密开销以及精度误差。

### 2.3 信息熵

熵的概念（T. Clausius，1854）由热力学推导出来，是表征系统无序程度的参数。在热力学熵的基础上，香农提出了一种信息值的数学量化即信息熵，它可以衡量信息不确定性，在信息论中起着核心作用。其定义为：给定随机变量 $X$，$P(X)$ 描述了变量 $P(X) = [p(x_1), p(x_2), \cdots, p(x_n)]$ 的分布（即 $p(x_i)$，其中 $p(x_i)$ 表示变量 $X = x_i$ 的概率），则变量 $X$ 的信息熵为：

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log p(x_i) \tag{5}$$

可以观察到如果随机变量 $X$ 有 $k$ 个可能的取值，则 $H(X) \in [0, \log k]$。当 $X$ 分布越均匀时，$H(X)$ 越接近 $\log k$，分布越集中时越接近 0。

本文基于以上信息熵的性质改进了使用标记样本训练的模型的目标函数，利用信息熵的计算因子来确定标记样本在训练中的权重，使模型更加专注于高置信度样本，提高模型再训练的有效性和鲁棒性。

## 3 形式化描述

本节对 PCTC 算法系统进行形式化描述，并给出相关安全定义。

### 3.1 系统形式化

在基于众包的协作学习框架中，主要有三类实体：需求方（受益方）、半可信第三方（诚实且好奇）和合作参与方（拥有本地模型可完成标注任务），如图 1 所示。需求方 Pr 拥有本地已分类私有数据集 $\mathbb{D}^{\mathrm{label}}$ 和存疑待标注数据集$\mathbb{X}$，需求方希望通过协作学习获取存疑数据$\mathbb{X}$的标签$\mathbb{Y}$来提高模型的准确性，是系统中的受益方。首先，num 个合作参与方 $P_i$ 利用本地数据集 $\mathbb{D}^{\mathrm{local}}$ 训练得到本地模型 $\mathcal{M}^{\mathrm{local}}$ 并对接收到的样本 $x$（$x \in \mathbb{X}$）进行分类，生成标注向量 $f_i(x)$ 并发送给半可信第三方 PG。然后，PG 对 $\mathrm{votes}(x) = \sum f_i(x) = [\mathrm{vote}_1(x), \mathrm{vote}_2(x), \cdots, \mathrm{vote}_k(x)]$ 进行聚合得到分类结果，其中 $k$ 和 $\mathrm{vote}_i$ 分别代表类别数和第 $i$ 类的投票数，根据分类结果可以进一步得到聚合标签 $y = \mathrm{OneHot}(\mathrm{argmax}(\mathrm{votes}(x)))$。最后需求方将标注好的数据集$\mathbb{X}$加入训练集 $D$，使用机器学习算法对模型进行再训练，提高模型精度。

### 3.2 安全定义形式化

在 PCTC 中，用户的私有数据集不会离开本地，仅传递标注的分类信息。因此，在本节中主要考虑在选择明文攻击下的分类结果隐私安全即算法的 IND-CPA 安全。假设所有实体都按照约定协议工作（半诚实状态），可采用不可区分性游戏（IND 游戏）实验来刻画 IND-CPA 安全，该游戏中有两个参与者，一个挑战者 Challenger 和一

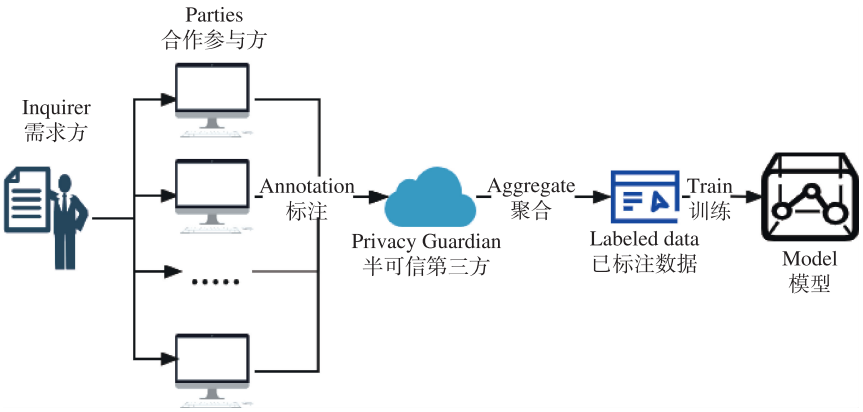


图 1 基于众包的协作学习框架

个敌手 Foe，敌手对系统发出挑战，挑战者接受敌手的挑战。IND 实验的形式化描述如下：

初始化：

Init：Challenger generates $\psi$，（pk，sk）←KeyGen（$k$），Foe←pk；

挑战：

挑战 1：

①Foe select $r_1$ and $r_2$ for $P_i$；

②$P_i$ output［Enc（$r_{b'}$）］，$b' \in \{1, 2\}$；

③Foe←［Enc（$r_{b'}$）］。

挑战 2：

①Foe select $r_{i1}$ and $r_{i2}$ for all $P_i$；

②$b' \in \{1, 2\}$，all $P_i$ output［Enc（$r_{ib'}$）］；

③Foe←$\sum_{i=1}^{num}$［Enc($r_{ib'}$)］。

猜测：

①Foe query，Foe output $b$，$b \in \{1, 2\}$；

②if $b = b'$，Foe success，else fail。

实验分为三个部分，初始化、挑战和猜测，其中 $\psi$ 代表系统，$k$ 是安全系数用来确定加密密钥长度。在该系统中，敌手的身份可以是能接触到分类结果的半可信第三方和合作参与方，分别对应挑战 1 和挑战 2。根据定理，如果对任何多项式时间的敌手 Foe，存在一个可忽略 negl（$k$），使得 $\mathrm{Adv}_{\psi,\mathrm{Foe}}(k) \leqslant \mathrm{negl}(k)$，就称该算法具有不可区分性。其中 $\mathrm{Adv}_{\psi,\mathrm{Foe}}$ 为敌手优势，是敌手通过努力得到的，可表述为 $|\Pr[b = b'] - 1/2|$。因此只要敌手在多项式时间内不能以不可忽视的概率区分分类结果，则 PCTC 在选择明文攻击下满足 IND-CPA 安全。

## 4 PCTC 概述

本节对 PCTC 进行概述，算法主要包括安全聚合机制和基于信息熵的模型学习算法。安全聚合机制用于对标注数据进行聚合，实现数据的可用不可见。基于信息熵的模型学习算法将使得标注样本被更有效地应用于模型训练，提高模型训练的鲁棒性。本节分别对以上两个算法进行介绍，最后给出 PCTC 的完整实现方案，并对方案的安全性进行分析。

### 4.1 安全聚合机制

问题描述：num 个协作参与方 $P_i$ 拥有向量 $\boldsymbol{x}_i$，希望在不泄露原始数据的情况下得到一个满足差分隐私的计算结果 $f(x) = \sum \boldsymbol{x}_i + \mathrm{Gaussian\ noise}(0, \sigma^2)$。其中向量元素 $x_{ij} \in \{0, 1\}$ 且分布集中，采用加密算法极易被对手攻击。进一步地，高斯噪声会引入浮点数，现有加密算法对浮点数运算会造成计算开销增加。针对上述问题，本文在可控的计算开销内设计了一个简单高效的聚合机制 Aggregation（$X$，$\sigma^2$），可以有效规避浮点数加解密运算，算法描述如下：

算法 1：聚合机制 Aggregation（$X$，$\sigma^2$）

输入：各参与方 $P_i$ 拥有的向量 $\boldsymbol{x}_i$，噪声大小 $\sigma^2$

输出：$\sum_{i=1}^{num} \boldsymbol{x}_i + \mathrm{Guassian\ noise}(0, \sigma^2)$

参与方：

①生成与 $\boldsymbol{x}_i$ 维数相同的随机向量 $\boldsymbol{v}_i$，向量元素 $v_{ij} \in N$；

②计算 $m_i = \boldsymbol{x}_i + \boldsymbol{v}_i$，加密 Enc（$m_i$）；

③将 Enc（$m_i$）和 $\boldsymbol{v}_i$ + Gaussian（0，$\sigma^2/k$）分别发送给半可信第三方。

半可信第三方：

①分别计算 $c = \prod_{i=1}^{num} \mathrm{Enc}(m_i) (\mathrm{mod} n^2)$，$v = \prod_{i=1}^{num} (\boldsymbol{v}_i + \mathrm{Gaussian}(0, \sigma^2/k))$；

②将聚合密文 $c$ 发送给各参与方进行解密，并进一步聚合解密结果得到 $m = \sum_{i=1}^{num} m_i$，计算 $f(x) = m - v = \sum_{i=1}^{num} \boldsymbol{x}_i + \mathrm{Guassian}(0, \sigma^2)$。

### 4.2 基于信息熵的模型学习算法

为解决低可信度的标注样本误导模型学习的问题，本文提出了基于信息熵的模型学习算法 IEL。基于信息熵的模型学习算法将能够表达变量不确定性程度的信息熵引入到损失函数中动态调整标记样本的权重，提高模型的训练的有效性。具体实现步骤如算法 2 所示。

算法 2：基于信息熵的模型学习算法 IEL（$\mathbb{D}^{\mathrm{label}}$，votes（$\mathbb{X}$），$\mathrm{Model}^{\mathrm{label}}$，$\mathbb{X}$）

输入：$\mathbb{D}^{\mathrm{label}}$，votes（$\mathbb{X}$），$\mathrm{Model}^{\mathrm{label}}$，$\mathbb{X}$

输出：$\mathrm{Model}^{\mathrm{retrain}}$

①将标记数据加入 $\mathbb{D}^{\mathrm{label}}$ 形成新的训练数据集 $\mathbb{D} = \mathbb{D}^{\mathrm{label}} + (\mathbb{X}, \mathrm{votes}(\mathbb{X}))$；

②训练模型 $\mathrm{Model}^{\mathrm{label}}$，输出 $X$ 的预测结果 $\hat{Y}$，$X \in D$；

③计算 $P = \mathrm{votes}(X) / \sum_{i=1}^{k} \mathrm{vote}_i(X)$ 获取标签结果的分布；

④$Y = \mathrm{OneHot}(\mathrm{argmax}(\mathrm{votes}(X)))$ 得到标签的独热编码；

⑤最小化目标函数，并计算 $\nabla J(\theta) = \partial J(\theta) / \partial \theta$

$$J(\theta) = -\frac{1}{N}\sum_{i=1}^{N}\left(\left(1 + \frac{\sum_{j=1}^{k} P_j \log P_j}{\log k}\right)\sum_{j=1}^{k} Y_j \log \hat{Y}\right) + \lambda J(f)$$

⑥梯度下降 $\theta' = \theta - \alpha \nabla J(\theta)$；

⑦循环②～⑥。

### 4.3 PCTC 具体实现

为了实现通信过程中不泄露局部数据集中的个体敏感信息，以及不引入过高计算的开销，本文设计了基于 Paillier 半同态加密的隐私保护协作算法，在保障数据的安全聚合的同时降低计算开销，改进模型的学习算法。

算法 3：PCTC（$\mathcal{D}^{\text{label}}$，$\mathbb{X}$，$\mathcal{M}^{\text{local}}$，$\varepsilon$，$\delta$）

输入：$\mathcal{D}^{\text{label}}$，$\mathbb{X}$，$\mathcal{M}^{\text{local}}$，$\varepsilon$，$\delta$

输出：$\mathcal{M}^{\text{retrain}}$

初始化：权威机构初始化并生成公私钥对 KeyGen（）→（pk，sk），其中公钥公开，私钥分发给各参与方。

①需求方利用本地已标注数据集 $\mathcal{D}^{\text{label}}$ 和训练模型 $\mathcal{D}^{\text{label}}$，将存疑数据$\mathbb{X}$发送给合作参与方标注；

②参与方利用本地模型 $\mathcal{M}^{\text{local}}$ 对接收到的数据$\mathbb{X}$进行预测，输出预测结果 $\text{result}_i$（$\mathbb{X}$）；

③选定噪声强度 $\sigma = \sqrt{2\ln(1.25/\delta)}/\varepsilon$，调用算法 Aggregation（$\text{result}_i$（$\mathbb{X}$），$\sigma^2$）输出最终聚合结果 $\text{votes}(\mathbb{X}) = \sum_{i=1}^{\text{num}} \text{result}_i(\mathbb{X}) + \text{Gaussian}(0, \text{num} * \sigma^2)$ 并发送给需求方；

④需求方使用基于信息熵的模型训练算法 IEL（$\mathcal{D}^{\text{label}}$，votes（$\mathbb{X}$），$\text{Model}^{\text{label}}$，$\mathbb{X}$）对模型进行再训练输出 $\mathcal{M}^{\text{retrain}}$。

### 4.4 安全分析

**定理 1** PCTC 提供了保密性和可衡量的隐私保证。

证明：首先在 PCTC 中各合作参与方的私有数据均不需要离开本地，这保证了数据的机密性即本地数据只能在本地读取。其次，参与方需要发送标注结果给半可信第三方，通过引入同态加密和随机种子设计的安全聚合方案确保了参与者（半可信第三方和合作参与方）不能破译单个参与方的预测结果，并且通过差分隐私噪声机制保证了即使在参与方很少的情况下也能提供不同的隐私保证。

**定理 2** 在半诚实的设置下，PCTC 是 IND-CPA 安全的。

PCTC 的安全性是以采用的 Paillier 半同态加密的安全性为基础的，Paillier 半同态加密是语义安全的，进一步地本文用 IND 游戏证明 PCTC 是 IND-CPA 安全的。根据 3.2 节的安全定义描述，如果系统中任何一个能接触到分类结果密文的腐败方（半可信第三方或者合作参与方）都不能在多项式时间内以不可忽略的优势区分密文，则 PCTC 在选择明文攻击下具有不可区分性即 IND-CPA 安全。

证明：首先 PCTC 采用的是概率性的加密方案，在每次加密时不仅会生成随机数，而且还会加入随机种子扰乱真实明文的分布，因此同一明文在不同的加密过程中得到的密文是不同的，即无法通过加密选定明文并逐一对比的方式来区分密文。因此，要想区分密文只能通过多项式时间内的数学方法。在 PCTC 中，半可信第三方可收到各方发送的密文 $c = \text{Enc}(m_i) = (g^{m_i} r_{ij}^n) \bmod n^2$，参与方收到来自半可信第三方的聚合密文 $c = \prod_{i=1}^{\text{num}} \text{Enc}(m_i) (\bmod n^2)$，其中 $m_i = \boldsymbol{v}_i + \text{result}_{ij}$。如果腐败方想要以不可忽略的优势成功攻击，则需同时满足以下两点：

①能在多项式时间内判断是否存在 $\boldsymbol{v}_i$，使得 $\boldsymbol{v}_i + \text{result}_{ij}$ 在模 $n^2$ 是 $n$ 次剩余，$\boldsymbol{v}_i \in [0, n]$。

②能在多项式时间内判断满足条件①的 $\boldsymbol{v}_i$ 是否与 $P_i$ 生成的随机种子相同。

条件①中，遍历 $\boldsymbol{v}_i$ 的时间复杂度为 $O(n)$，判定模 $n^2$ 是 $n$ 次剩余问题需要在 $O(p(n))$ 内完成才能满足要求。而判定性合数剩余阶假设（DCRT）问题即给定合数 $n$ 和正数 $z$ 判断模 $n^2$ 的 $n$ 阶剩余 $z$ 是否存在，是没有办法在多项式时间内解决的。因此条件①不成立。对于条件②，基于保密性原则该条件不成立。综上，腐败方无法在多项式时间内以不可忽略的优势区分密文，PCTC 是 IND－CPA 安全的得证。

**定理 3** 在半诚实的设置下，PCTC 可以容忍部分实体勾结。

证明：在 PCTC 中密钥 $\lambda$ 是对所有参与者保密的，每个合作参与方都只能获取部分密钥，而要想破解密文必须获取全部密钥，因此当勾结的参与方小于参与方总数时就无法在多项式时间内破解密文。

**定理 4** PCTC 在部分合作参与方退出时，分类结果仍然是隐私安全的。

证明：在部分合作参与方退出时，攻击者可通过对退出前和退出后的查询结果进行比对推断出退出参与方的标注结果，这种攻击在只有一个参与方退出时尤为有效。PCTC 通过添加随机噪声的方式使查询结果的分布在一定概率上偏离原始数据，这使得攻击者得出的推断信息也会在概率上偏离真实数据导致无法判断其真伪。

## 5 实验性能及分析

为了验证提出的算法的有效性，本节构建基于众包的分类系统对 PCTC 进行模拟和评估。系统实验中，使用三个指标对算法性能进行评估：加密时间、解密时间和模型准确率。实验的运行环境为 Ubuntu 18.04.6 LTS 操作系统，处理器为 3.5-GHz Intel Core i9-9900x，并依赖于

PHE 和 PyTorch 等库。

### 5.1 实验设置

本文用经典计算机视觉分类任务 MNIST 来评估模型精度。MNIST 是一个手写数字的图像数据集，它包含 10 个类别。实验过程中，首先选用不相交的数据集子集训练一个局部模型的集合用于模拟实际场景下的参与方，然后采用 PCTC 算法对所有局部模型预测标签进行安全聚合，最后使用聚合标签对模型进行再训练以提高精度。实验参数上，设置合作参与方数量 $N=15$；除非另有说明，训练中每一 batch 的训练样本数为 32，momentum SGD 的参数为 0～9，初始学习率设置为 0～1。为了避免实验结果的偶然性，将每个实验重复 10 次并计算平均值作为最终的实验结果。

### 5.2 加密时间开销

在加密算法中，密钥长度是影响安全等级的一个重要因素。一般而言，密钥长度越长，加解密时间开销越高，破解难度越大，安全级别越高。为实现安全聚合，PCTC 中使用了半同态加密算法，并设计了聚合机制避免了浮点数加解密开销。为验证 PCTC 的算法性能，测试了当查询数据集 $|\lambda|=500$ 时，系统在不同密钥长度下的加密时间开销（此处的时间开销是所有参与方总的时间开销）。实验结果如图 2 所示。

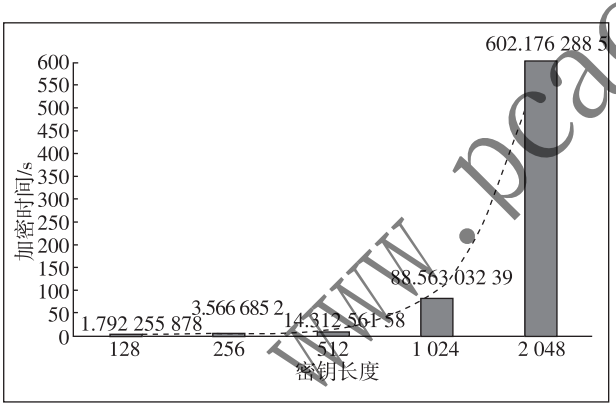


图 2 不同密钥长度的加密时间开销

从图 2 中趋势线可以看到，随着密钥长度的增长，加密时间开销呈现指数级的上升。当密钥长度 $\lambda=2\ 048$ 时，每张图片的时间开销最高为 1.20 s，对并行的参与方来说是一个可接受的时间开销范围（每个参与方的时间消耗约为 0.1 s），并且由于并没有直接传递加密的真实数据，因此采用较低等级的密钥长度也可达到相对高等级的安全保护，使时间开销相比同等级的安全算法更小。

### 5.3 解密时间开销

基于之前加密的查询数据集标注结果，对其进行聚合解密。因为本算法没有浮点数误差问题，所以只对聚合解密的时间进行报告。实验结果如图 3 所示（注：加解密采用同样长度的密钥）。

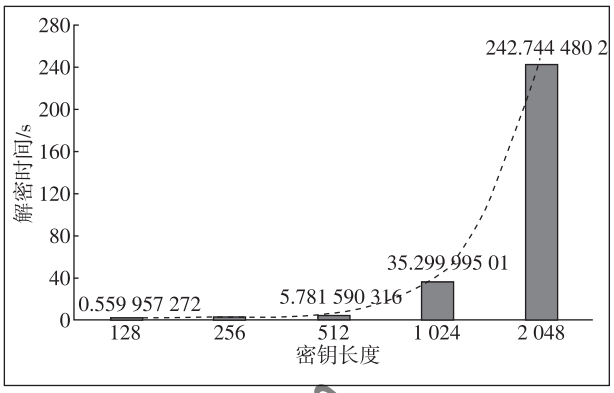


图 3 不同密钥长度的的解密时间开销

每个图片标注结果的聚合解密时间也只需要 0.49 s。因为在系统中设计了针对浮点数计算开销的聚合机制，避免了浮点数的运算，可以有效减少系统加解密的计算复杂度，从而降低时间开销。因此本文提出的隐私保护算法在时间开销上是乐观的，能够满足大部分应用场景的性能需求。

### 5.4 模型准确性提升比较

为了研究 PCTC 中基于信息熵的模型学习算法 IEL 对效用性的提升效果，本文根据本地模型在验证集上的准确性来表征专业知识水平，并设计 3 组不同专业水平的参与者集合进行测试。首先，将数据集分别划分为 15 个数量为 200、400、600 的不相交子集，并分别训练用于模拟 15 个本地参与方模型，参与方本地模型的精度如图 4 所示（其中①②③分别对应 3 组不同大小的数据子集）。基于异质性的思想，参与者的局部模型为 VGG11、VGG16 和 Resnet18 等不同的经典模型。然后随机选取 200 个样本作为未标记数据集并设定 $P_1$ 为需求方，使用 PCTC 算法对所有参与方模型对样本预测标签进行安全聚合，最后使用加入标注数据后的数据集对 $P_1$ 进行训练以提高再训练精度。训练采用 IEL 和 CaPC 两种方式，用于对比两种算法的模型效果。同时为了保证一系列查询组合一起后整体仍然满足差分隐私，考虑到查询函数之间的关系性，对于 $k$ 次查询的隐私预算消耗采用强组合原理[18]计算，单次查询的隐私预算和误差分别为 1.2 和 $1\times10^{-5}$。实验结果如表 1 所示，其中 Expertise 表示参与方模型的平均精度（专业水平），Init 表示 $P_1$ 模型的初始精度，Retrain_ CaPC 和 Retrain_ IEL 分别表示用 CaPC 和 IEL 再训练后的 $P_1$ 模型精度。

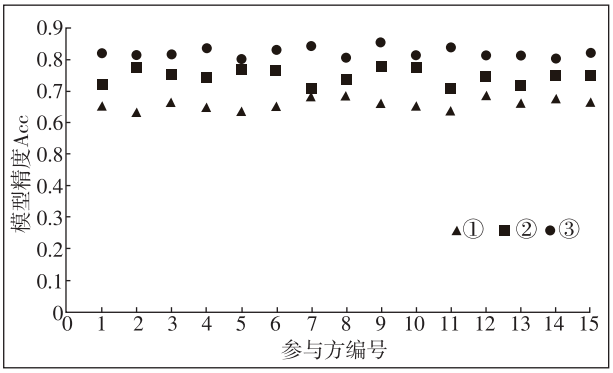


图 4　15 个参与方本地模型精度

表 1　使用 IEL 和 CAPC 再训练模型精度比较

| | Expertise | Init | Retrain_ CaPC | Retrain_ IEL |
|---|---|---|---|---|
| ① | 0. 659 5 | 0. 654 2 | 0. 640 1 | 0. 714 5 |
| ② | 0. 745 8 | 0. 717 | 0. 756 7 | 0. 774 3 |
| ③ | 0. 822 2 | 0. 820 4 | 0. 848 3 | 0. 859 7 |

根据表 1 的结果，可以观察到在不同的参与者专业水平下，使用 IEL 进行再训练的模型精度均得到了提升。具体地，当参与者的平均专业知识较低时（65. 95%），IEL 相较于 CaPC 的提升效果最为显著。随着参与者平均专业水平的不断提高，IEL 与 CaPC 的再训练性能也不断接近，实验结果与预期一致。当参与者的专业水平较低时，引入信息熵的计算因子来衡量样本标签的置信度作为损失的比例因子，从而有效地利用可信度高的标注结果，更好地引导模型训练；当各个参与者的专业水平较高时，样本标签准确率较高，此时使用 IEL 可能会导致准确但有争议的聚合标签的样本权重被降低，但是由于相对地提高了本地私有（准确）样本的权重和在一定程度上降低了差分噪声的影响，因此仍旧取得了相对较高准确率。综上所述，本文方法在协作标注的场景下优于 CaPC 算法，对模型的性能提升具有显著效果，并且在加强模型精度的同时能够保护数据隐私。

## 6　结论

本文提出了一种隐私性与效用性并重的面向分类任务的隐私保护协作技术。在隐私性方面，将同态加密和差分隐私相结合，创新性地设计了一种高效的聚合方案，在避免浮点数运算开销和误差的同时实现数据的安全聚合；在效用性方面，提出了基于信息熵的模型学习算法，动态调整标记样本的权重，使模型聚焦于质量更稳定的样本，优化模型对标注样本的学习策略；最后对安全性和性能进行了分析。结果表明，该方案能够在保障数据隐私的情况下提高模型的训练精度。未来的研究工作将集中于参与方退出时的隐私保护。

## 参考文献

[1] YANG M, CUI S, Zhang Y, et al. Data and image classification of haematococcus pluvialis based on svm algorithm [C] //2021 China Automation Congress (CAC). IEEE, 2021: 522 - 525.

[2] SHAIKH, TAWSEEF AYOUB, et al. Machine learning for smart agriculture and precision farming: towards making the fields talk [J]. Archives of Computational Methods in Engineering, 2022, 29 (7): 4557 - 4597.

[3] CHOQUETTE-CHOO, CHRISTOPHER A, et al. CaPC learning: confidential and private collaborative learning [J]. arXiv preprint arXiv: 2102. 05188, 2021.

[4] SAMARATI P, SWEENEY L . Generalizing data to provide anonymity when disclosing information (abstract) [C] // Seventeenth Acm Sigact-sigmod-sigart Symposium on Principles of Database Systems. ACM, 1998: 188.

[5] SHOKRY M, AWAD A I, ABD-ELLAH M K, et al. Systematic survey of advanced metering infrastructure security: vulnerabilities, attacks, countermeasures, and future vision [J]. Future Generation Computer Systems, 2022 (136): 358 - 377.

[6] 周吴尧，梁淼. 隐私计算让数据“可用不可见”[N]. 中国城乡金报，2021 - 09 - 03 (A04). DOI: 10. 28148/n. cnki. ncxjr. 2021. 001475.

[7] KONEN J, MCMAHAN H B, RAMAGE D, et al. Federated optimization: distributed machine learning for on-device intelligence [J]. arXiv preprint arXiv: 1610. 02527, 2016.

[8] HAMM J, CAO P, BELKIN M . Learning privately from multiparty data [J]. arXiv prepoint arXiv: 1602. 03552, 2016.

[9] ALI M, KARIMIPOUR H, TARIQ M . Integration of blockchain technology and federated learning in vehicular (IoT) networks: a comprehensive survey [J]. Computers & Security, 2021 (5): 102355

[10] 项威，刘文卓，王邦. 基于 Web 的众包文本标注平台构建与应用 [J]. 计算机应用，2022，42 (S1): 1 - 6.

[11] ZHANG J, ZHANG J, CHEN J, et al. GAN enhanced membership inference: a passive local attack in federated learning [C]// ICC 2020 - 2020 IEEE International Conference on Communications (ICC). IEEE, 2020.

[12] YE F, DONG X, SHEN J, et al. A verifiable dynamic multi-user searchable encryption scheme without trusted third parties [C]// 2019 IEEE 25th International Conference on Parallel and Distributed Systems (ICPADS). IEEE, 2019.

[13] CHAI D, WANG L, CHEN K, et al. Secure federated matrix factorization [J]. arXiv preprint arXiv: 1906. 05108v1, 2019.

[14] HAO M, LI H, XU G, et al. Towards efficient and privacy-preserving federated deep learning [C]// ICC 2019 - 2019 IEEE International Conference on Communications (ICC). IEEE,

2019.

[15] FANG C, GUO Y, WANG N, et al. Highly efficient federated learning with strong privacy preservation in cloud computing [J]. Computers & Security, 2020 (96): 3 - 4.

[16] MIAO Y, LIU Z, Li H, et al. Privacy-preserving Byzantine-robust federated learning via blockchain systems [J]. IEEE Transactions on Information Forensics and Security, 2022 (17): 2848 - 2861.

[17] DWORK C. Differential privacy [J]. Lecture Notes in Computer Science, 2006 (4052): 1 - 12.

[18] DWORK C, ROTHBLUM G N, VADHAN S. Boosting and differential privacy [C] //2010 IEEE 51st Annual Symposium on Foundations of Computer Science. IEEE, 2010.

(收稿日期：2023 - 03 - 21)

**作者简介：**

黎兰兰 (1999 - )，女，硕士，主要研究方向：隐私保护、多方协作学习。

张信明 (1964 - )，男，博士，教授，博士生导师，主要研究方向：无线网络、大数据、深度学习。

## 中国（大湾区）工业互联网发展与安全峰会在深圳成功召开

4 月 8 日，中国（大湾区）工业互联网发展与安全峰会在深圳成功召开。本次峰会与第十一届中国电子信息博览会同期举办，由中国电子信息产业集团有限公司主办，中国电子信息产业集团有限公司第六研究所、工业控制系统信息安全技术国家工程研究中心、中电会展与信息传播有限公司承办，来自线上线下 2000 余位行业专家和技术人员收看本次峰会。

中国电子信息产业集团有限公司总经理助理、中电工业互联网有限公司董事长朱立锋代表主办方在致辞中表示，工业互联网作为新一代信息技术与工业经济深度融合的关键基础设施，正以体系化发展方式推动中国制造业迈向高质量发展。中国电子深入贯彻党的二十大精神，主动服务国家战略，持续优化产业结构，南迁深圳以来，中国电子把握大变局、融入大湾区，全面加强网信领域战略科技创新，着力攻克关键核心技术，大力推动信息技术应用创新产业发展，为大湾区建设发展作出了积极贡献。中国电子将围绕以数字技术支撑国家治理体系和治理能力现代化、服务数字经济高质量发展、保障国家网络安全三大核心任务，着力发展计算产业、集成电路、网络安全、数据应用、高新电子、工业互联网等重点业务，打造国家网信事业核心战略科技力量。

工业和信息化部网络安全产业发展中心副处长郭威在题为《工业互联网数据安全发展机遇与挑战》的视频致辞中指出，工业互联网面临安全事件频发、安全建设同步不够、安全意识亟待加强、应急响应体系有待完善等主要问题，在工业经济向数字经济转型过程中，工业互联网呈现关键技术加速融合、行业应用持续深化、能力建设逐渐成熟等特点。围绕工业互联网安全管理需求，应以“原生安全、密码应用、纵深防御、测评加固、功防演练、统一管理”为重点方向，构建工业企业安全综合防护体系。

中国科学院院士、通信网络领域专家尹浩在题为《把握工业信息安全发展态势，合力应对工业信息安全挑战》主旨演讲中着重讲解了工业信息安全面临的新形势和安全风险应对举措。尹院士认为，工业信息安全是制造强国、网络强国的基石，全球工业经济正加速数字化转型，工业数据流动范围不断延展，工业信息安全已经成为国际竞合、国际规则制定的核心议题和重要领域。为此，他提出应从工业设备安全、传输网络安全、控制系统安全、应用服务安全、工业数据安全五个维度来构建工业信息安全保障体系。

来自北京航空航天大学、中电鹏程智能装备有限公司、北京惠而特科技有限公司等产学研用各领域的专家学者围绕工业信息安全和工业互联网产业发展展开交流。

本次中国（大湾区）工业互联网发展与安全峰会聚焦工业制造业数字化转型，搭建高端合作交流平台，汇聚政产学研用各领域资源，通过政策解读、案例分享、行业前瞻等研讨环节，探讨大湾区制造业数字化转型的机遇与挑战，为加速推进工业互联网赋能制造业数字化转型升级，实现制造业高质量发展贡献力量。

# 版权声明

凡《网络安全与数据治理》录用的文章，如作者没有关于汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权等版权的特殊声明，即视作该文章署名作者同意将该文章的汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权授予本刊，本刊有权授权本刊合作数据库、合作媒体等合作伙伴使用。同时，本刊支付的稿酬已包含上述使用的费用，特此声明。

《网络安全与数据治理》编辑部