基于目标扰动的 AdaBoost 算法

张淑芬[1,2,3]，董燕灵[1,2,4]，徐精诚[1,2,4]，王豪石[1,2,4]

（1. 华北理工大学理学院，河北 唐山 063210；2. 河北省数据科学与应用重点实验室，河北 唐山 063210；
3. 唐山市大数据安全与智能计算重点实验室，河北 唐山 063210；4. 唐山市数据科学重点实验室，河北 唐山 063210）

摘 要：针对 AdaBoost 算法的多轮迭过程会放大为实现差分隐私保护而添加的噪声，从而导致模型收敛缓慢、数据可用性大幅降低的问题，提出了一种基于目标扰动的 AdaBoost 算法——DPAda，采用目标扰动的方式对样本权值进行加噪，精确计算其敏感度，并赋予其动态的隐私预算。为了解决噪声叠加过多的问题，提出基于摆动数列、随机响应和改进的随机响应 3 种噪声注入算法。实验结果表明，与 DPAda_Random 算法和 DPAda_Swing 算法相比，DPAda_Improved 算法能实现数据的隐私保护，拥有更高的分类准确率，优于其他差分隐私 AdaBoost 算法，并能解决连续加噪带来的噪声过大的问题。

关键词：差分隐私；摆动数列；随机响应；隐私预算分配；AdaBoost 算法

中图分类号：TP309.2

文献标志码：A

DOI: 10.11959/j.issn.1000−436x.2023028

AdaBoost algorithm based on target perturbation

ZHANG Shufen[1,2,3], DONG Yanling[1,2,4], XU Jingcheng[1,2,4], WANG Haoshi[1,2,4]

1. College of Science, North China University of Science and Technology, Tangshan 063210, China
2. Hebei Key Laboratory of Data Science and Application, Tangshan 063210, China
3. Tangshan Key Laboratory of Big Data Security and Intelligent Computing, Tangshan 063210, China
4. Tangshan Key Laboratory of Data Science, Tangshan 063210, China

Abstract: Aiming at the problem that the multi-round iteration process in the AdaBoost algorithm will amplify the noise added to achieve differential privacy protection, which leads to slow model convergence and greatly reduced data availability, an AdaBoost algorithm based on target perturbation—DPAda was proposed. Target perturbation was used to add noise to sample weights, accurately calculated their sensitivity, and a dynamic privacy budget was given. In order to solve the problem of excessive noise superposition, three noise injection algorithms based on swing sequence, random response and improved random response were proposed. The experimental results show that compared with DPAda_Random and DPAda_Swing, DPAda_Improved achieves the privacy protection of data, has higher classification accuracy, as well as better than other differential privacy AdaBoost algorithm, and can also solve the problem of excessive noise caused by continuous noise addition.

Keywords: differential privacy, swing sequence, random response, privacy budget allocation, AdaBoost algorithm

0 引言

大数据时代下，海量数据以及强大的数据分析与数据挖掘能力无不反映了数据的高价值性，然而由于一些数据发布者缺乏管理和保护数据的意识，使攻击者有机可乘。虽然现有的隐私保护技术如数据脱敏[1]、k-匿名[2]、同态加密[3]等方法能在一定范围内保护数据中的隐私信息，但当攻

收稿日期：2022−10−13；**修回日期：**2022−12−25

通信作者：董燕灵，dongyl@stu.ncst.edu.cn

基金项目：国家自然科学基金资助项目（No.U20A20179）

Foundation Item: The National Natural Science Foundation of China (No.U20A20179)

击者拥有强大的背景知识时，同样无法抵抗攻击者的攻击。于是，差分隐私保护技术[4-6]应运而生。差分隐私保护技术不需要假设攻击者有背景知识，且成功地避免了攻击者会因为新样本的出现而得到新的知识。该技术允许数据收集者采用拉普拉斯机制[7]、指数机制[8]或高斯机制[9]对数据添加噪声进行扰动，从而使攻击者无法辨别某一样本是否在数据集中。

差分隐私通过添加噪声来扰动数据，使查询结果具有一定的随机性。主要的扰动方式有输入扰动、输出扰动和目标扰动。输入扰动对原始数据直接添加噪声，扰动后的数据成为模型学习的输入，直接参与之后的训练。输出扰动是指对模型的最终输出添加噪声，得到扰动后的结果。目标扰动通过对目标函数添加噪声，使最终模型的输出具有随机性。

当前，随机森林[10-18]是将差分隐私应用于集成学习中的主要研究方向，然而将差分隐私与 AdaBoost 算法结合的研究却甚少。这是因为差分隐私 AdaBoost 算法面对的挑战[19]主要是算法迭代引起的噪声叠加造成最后模型无法收敛和分类准确率较低。Gambs 等[20]提出了 2 种隐私保护 AdaBoost 算法，适用于 2 个或 2 个以上参与者在不直接交流数据集的情况下保护参与者的数据隐私。Li 等[21]选取 AdaBoost 算法作为每个参与者的本地学习器，提出一种基于集成策略的自适应分布式隐私保护数据挖掘技术。沈思倩[22]分别研究了基于完全数据集和不完全数据集的差分隐私 Adabooot 算法。贾俊杰等[23]进一步将分类与回归树（CART）取代树桩作为 AdaBoost 算法的基分类树。但上述研究都只是对弱学习器的类标签或类计数进行了加噪，而且在敏感度的计算上也不够严谨。Li 等[24]首次将梯度提升决策树与差分隐私相结合，在敏感度的计算上更加精确。在这些已有的将差分隐私与 AdaBoost 算法结合的研究中，主要有 3 个缺陷：一是隐私预算的分配不具有针对性，较泛化；二是在敏感度的计算上不够严谨；三是在添加噪声上，尚未解决 AdaBoost 算法由于多轮迭代而放大噪声的问题。为此，本文提出对每轮迭代过程中的最大最小样本权值进行加噪，针对该加噪对象提出一种动态的隐私预算分配策略，严谨地根据敏感度定义计算了查询函数最大变化范围，并提出了 3 种能有效控制噪声叠加过大并提高模型可用性的方案。实现在保护数据的同时让模型拥有更高的分类准确率。具体来说，本文的主要工作如下。

1) 提出了一种基于目标扰动的 AdaBoost 算法。根据 AdaBoost 的目标函数，提出本文实现 AdaBoost 差分隐私保护的加噪时机。

2) 针对 AdaBoost 算法每轮迭代的样本权值分布，选取最具代表性的最大权值和最小权值，并通过动态的隐私预算分配策略分配不同的隐私预算。

3) 根据敏感度定义，准确计算最大权值和最小权值的动态敏感度。

4) 为有效解决因每轮迭代加噪而引起的分类准确率失真、模型不可用问题，提出基于摆动数列、随机响应和改进的随机响应 3 种噪声注入方案，并在真实数据集上进行实验，以评估不同算法的有效性。

1　差分隐私保护技术

差分隐私保护技术最早被应用于数据库安全领域，查询 2 个仅相差一条记录的数据集，发现得到相同结果的概率非常接近，使攻击者无法进行差分攻击。

1.1　差分隐私的定义及相关概念

设数据集 D 与具有相同属性结构的数据集 D' 的对称差为 $D\Delta D'$，$|D\Delta D'|$ 表示 $D\Delta D'$ 中记录的数量。若 $|D\Delta D'|=1$，则称 D 和 D' 为邻近数据集。

定义 1　差分隐私[4-6]。设有随机算法 M，PM 为 M 所有可能的输出构成的集合。对于任意 2 个邻近数据集 D 和 D' 以及 PM 的任何子集 SM，若算法 M 满足

$$\Pr[M(D)\in \mathrm{SM}]\leqslant \mathrm{e}^{\varepsilon}\Pr[M(D')\in \mathrm{SM}] \tag{1}$$

则称算法 M 提供 ε – 差分隐私保护，其中，参数 ε 为隐私预算，表示数据集中增加或减少一条记录时，算法 M 的输出结果一致的概率；$\Pr[\cdot]$ 表示发生某一事件的概率。

定义 2　全局敏感度[7]。对于 2 个邻近数据集 D 和 D'，查询函数 f 最大的变化范围为全局敏感度 Δf，则 Δf 的计算式为

$$\Delta f=\max_{D,D'}\|f(D)-f(D')\|_1 \tag{2}$$

定义 3　串行组合[7]。差分隐私具有灵活的组合特性。假设有 n 个随机算法 K，其中 K_i 满足 ε_i –

差分隐私，那么对于同一数据集，由 $\{K_i\}(1 \leqslant i \leqslant n)$ 组合后的算法满足 $\text{sum}(\varepsilon_i)$ – 差分隐私。

1.2 实现机制

拉普拉斯机制通过向查询结果中添加随机噪声来实现差分隐私保护，并且添加的随机噪声是服从拉普拉斯分布的。设尺度参数为 λ，位置参数 $\mu = 0$，则该拉普拉斯分布为 $\text{Lap}(\lambda)$，它的概率密度函数为

$$P(x) = \frac{1}{2\lambda} \mathrm{e}^{-\frac{|x|}{\lambda}} \tag{3}$$

不同尺度参数λ的拉普拉斯分布如图 1 所示。

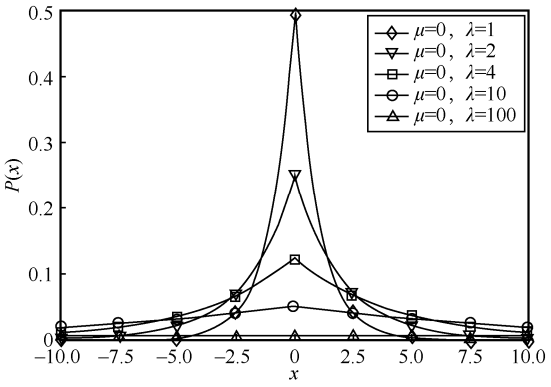


图 1 不同尺度参数 λ 的拉普拉斯分布

定义 4 拉普拉斯机制[7]。给定数据集 D 和隐私预算 ε，设有查询函数 $f(D)$，其敏感度为 Δf，那么随机算法 $M(D) = f(D) + Y$ 提供 ε – 差分隐私保护。其中，$Y \sim \text{Lap}\left(\frac{\Delta f}{\varepsilon}\right)$ 为随机噪声，并服从尺度参数为 $\frac{\Delta f}{\varepsilon}$ 的拉普拉斯分布。拉普拉斯机制的整体思想是以一定的概率输出异常值。为了使加噪后的数据脱离真实值，需要提高输出异常值的概率，即稳定地得到一个异常值，因此当拉普拉斯函数曲线越平缓时，异常值输出的概率就越高，数据相对于正确的中心值更加分散。

2 AdaBoost 算法

AdaBoost[25-27]是前向分步加法算法的特例，通过迭代，将前一轮强学习器的学习结果与当前一轮弱学习器的学习结果加权，来更新当前的强学习器。在训练弱学习器时，减小上一轮样本被正确分类的样本权值，提高样本被错误分类的样本权值。在最终加权组合成强学习器时，加大分类误差率小的弱学习器的权重，减小分类误差率大的弱学习器的权重。

本文使用的主要符号如表 1 所示。

表 1 主要符号

符号	含义
D_k	第 k 轮样本权值分布
$w_{k,i}$	第 k 轮迭代中第 i 个样本权值
Δf	全局敏感度
$G_k(x)$	第 k 轮弱学习器
$G(x)$	最终分类器
$\tilde{w}_{k,i}$	加噪后的 $w_{k,i}$
$\bar{w}_k$	第 k 轮样本权值均值
e_k	第 k 轮分类误差率
α_k	第 k 轮弱学习器权重
ε	隐私预算

2.1 AdaBoost 目标函数

AdaBoost 算法的损失函数为指数函数[28]，即定义损失函数为

$$L\big(y, F(x)\big) = \exp\big(-yF(x)\big) \tag{4}$$

将损失函数代入目标函数中，则 AdaBoost 的目标函数为

$$\sum_{i=1}^{N} \exp(-y_i F(x)) \tag{5}$$

由于 AdaBoost 算法是前向分步算法，假设第 k 轮的强学习器为 $f_k(x)$，第 $k-1$ 轮的强学习器为 $f_{k-1}(x)$，第 k 轮的弱学习器为 $G_k(x)$，则有以下关系

$$f_k(x) = f_{k-1}(x) + \alpha_k G_k(x) \tag{6}$$

AdaBoost 的目标是使前向分步算法得到的 α_k 和 $G_k(x)$ 能让 $f_k(x)$ 在训练集上的指数损失最小，即 AdaBoost 的目标是最小化目标函数，表示为

$$(\alpha_k, G_k(x)) = \underbrace{\arg\min}_{\alpha, G} \sum_{i=1}^{N} \exp[(-y_i)(f_{k-1}(x_i) + \alpha_k G_k(x_i))] \tag{7}$$

令 $w'_{ki} = \exp(-y_i f_{k-1}(x_i))$，不难发现，它的值不随着 α 和 G 的改变而改变，因此 w'_{ki} 与最小化无关，只依赖于 $f_{k-1}(x_i)$，并随着每一轮迭代而改变。

将 w'_{ki} 代入式(7)，此时最小化目标函数为

$$(\alpha_k, G_k(x)) = \underbrace{\arg\min}_{\alpha, G} \sum_{i=1}^{N} w'_{k,i} \exp[-y_i \alpha_k G_k(x_i)] \tag{8}$$

将 $\sum_{i=1}^{N} w'_{k,i}\exp[-y_i\alpha_k G_k(x_i)]$ 展开，可得

$$\sum_{i=1}^{N} w'_{k,i}\exp[-y_i\alpha_k G_k(x_i)]=$$
$$\sum_{i=1}^{N} w'_{k,i}\mathrm{e}^{-\alpha_k} I\{y_i = G_k(x_i)\}+\sum_{i=1}^{N} w'_{k,i}\mathrm{e}^{\alpha_k} I\{y_i \neq G_k(x_i)\}=$$
$$\mathrm{e}^{-\alpha_k}\sum_{i=1}^{N} w'_{k,i} I\{y_i = G_k(x_i)\}+\mathrm{e}^{\alpha_k}\sum_{i=1}^{N} w'_{k,i} I\{y_i \neq G_k(x_i)\}+$$
$$\mathrm{e}^{-\alpha_k}\sum_{i=1}^{N} w'_{k,i} I\{y_i \neq G_k(x_i)\}-\mathrm{e}^{-\alpha_k}\sum_{i=1}^{N} w'_{k,i} I\{y_i \neq G_k(x_i)\}=$$
$$\mathrm{e}^{-\alpha_k}\sum_{i=1}^{N} w'_{k,i}+(\mathrm{e}^{\alpha_k}-\mathrm{e}^{-\alpha_k})\sum_{i=1}^{N} w'_{k,i} I\{y_i \neq G_k(x_i)\} \tag{9}$$

其中，分类误差率 $e_k=\sum_{i=1}^{N} w'_{k,i} I\{y_i \neq G_k(x_i)\}$。式(9)右边对 α 求导，并使其等于 0，可得

$$\alpha_k=\frac{1}{2}\ln\frac{1-e_k}{e_k} \tag{10}$$

由式(6)和 $w'_{k,i}=\exp(-y_i f_{k-1}(x_i))$ 可得

$$w'_{k+1,i}=w'_{k,i}\exp(-y_i\alpha_k G_k(x_i)) \tag{11}$$

2.2　AdaBoost 算法流程

算法 1　AdaBoost 算法

输入　训练集 T，迭代次数 m

输出　最终分类器 $G(x)$

1) 初始化样本权值分布 $D_1=(w_{1,1},w_{1,2},w_{1,3},\cdots,w_{1,n})$，$w_{k,i}=\frac{1}{n}$，$i=1,2,3,\cdots,n$

2) for $k=1$ to m

3)　训练弱学习器 $G_k(x)//G_k(x):X\to\{1,-1\}$

4) 根据 $e_k==\sum_{i=1}^{n} w'_{k,i} I\{y_i \neq G_k(x_i)\}$，计算 $G_k(x)$ 的分类误差率 e_k // $I\{y_i \neq G_k(x_i)\}=\{1,\text{others}^{0,y_i=G_k(x_i)}\}$

5) if $e_k>0.5$ then break

6) 计算 $G_k(x)$ 的权重 $\alpha_k=\frac{1}{2}\ln\frac{1-e_k}{e_k}$

7) 计算更新后的样本权值，并进行归一化 $w_{k+1,i}=\frac{w_{k,i}\exp[-y_i\alpha_k G_k(x_i)]}{Z_k}$ // Z_k 为规范化因子，$i=1,2,3,\cdots,n$

8) end for

9) 输出最终分类器 $G(x)=\mathrm{sign}\left(\sum_{k=1}^{m}\alpha_k G_k(x)\right)$

3　AdaBoost 算法的隐私保护

在 AdaBoost 算法的实际应用过程中，算法模型会在生命周期的各个阶段面临不同的安全威胁，导致模型的隐私信息被泄露，模型的可用性、完整性被破坏。为了减少此类现象的发生，需要对 AdaBoost 算法进行隐私保护。

从隐私保护的角度来看，根据注入噪声的时机，实现差分隐私有输入扰动、输出扰动和目标扰动这 3 种方法。本文选取目标扰动作为 AdaBoost 算法实现差分隐私保护的加噪时机，并根据 2.1 节 AdaBoost 目标函数，取样本权值作为加噪对象。

3.1　基于样本权值的动态隐私预算分配策略

在 k 轮迭代后，更新的样本权值 $w_{k+1,i}$ 成为下一次迭代中计算错分样本率的输入。由于每一轮迭代过程中，权值小的样本是前一轮被正确分类的样本，前一轮被错分的样本在下一轮中会被赋予更高的权值来加以重视。同时，为了能够实现 AdaBoost 算法的快速收敛，在添加差分隐私保护时，需对权值较大的样本分配多的隐私预算，对权值较小的样本分配较少的隐私预算。基于此，提出一种基于样本权值的动态隐私预算分配策略。

假设总的隐私预算为 B，迭代次数为 m，样本数为 N，对于第 k 轮更新后的权值分布 D_{k+1}，寻找出最大权值 $\max_{w_{k+1,i}}$ 和最小权值 $\min_{w_{k+1,i}}$，分别为最大权值和最小权值添加噪声。其中样本的初始权值不需要添加噪声，即 $w_{1,i}=\frac{1}{N}$。

为 $\max_{w_{k+1,i}}$ 和 $\min_{w_{k+1,i}}$ 分别添加动态的大小为 $\frac{B}{m}\max_{w_{k+1,i}}$ 和 $\frac{B}{m}\min_{w_{k+1,i}}$ 的隐私预算后，$\widetilde{\max}_{w_{k+1,i}}$ 的值不能小于权值均值 $\overline{w}_{k+1}$，$\widetilde{\min}_{w_{k+1,i}}$ 的值不能大于权值均值 $\overline{w}_{k+1}$，即

$$\widetilde{\max}_{w_{k+1,i}}\geqslant\overline{w}_{k+1} \tag{12}$$

$$\widetilde{\min}_{w_{k+1,i}}\leqslant\overline{w}_{k+1} \tag{13}$$

此时，全局敏感度 Δf 是动态变化的，根据定义 2，计算最大权值 $\max_{w_{k+1,i}}$ 的全局敏感度 Δf 为

$$\Delta f_{\max}=\max_{w_{k+1,i}}-\overline{w}_{k+1} \tag{14}$$

计算最小权值 $\min_{w_{k+1,i}}$ 的全局敏感度 Δf 为

$$\Delta f_{\min}=\overline{w}_{k+1}-\min_{w_{k+1,i}} \tag{15}$$

其中，

$$\overline{w}_{k+1}=\frac{1}{N} \tag{16}$$

最后，将加完噪的样本权值 $\tilde{w}_{k+1,i}$ 归一化，得到 $w'_{k+1,i}$，且 $\sum_{i}^{N}w'_{k+1,i}=1$

算法 2 隐私预算分配算法 DPweight_dynamic

输入 更新后的权值 $w_{k+1,i}$，迭代次数 m，隐私预算 B

输出 样本权值 $w'_{k+1,i}$

1) 计算全局敏感度 $\Delta f_{\max}$, $\Delta f_{\min}$

2) 计算最大权值和最小权值的隐私预算，

$$\varepsilon_{\max}=\frac{B}{m}\max_{w_{k+1,i}},\varepsilon_{\min}=\frac{B}{m}\min_{w_{k+1,i}}$$

3) 为 $\max_{w_{k+1,i}}$ 和 $\min_{w_{k+1,i}}$ 分别添加拉普拉斯噪声 $\mathrm{Lap}\left(\frac{\Delta f_{\max}}{\varepsilon_{\max}}\right)$ 和 $\mathrm{Lap}\left(\frac{\Delta f_{\min}}{\varepsilon_{\min}}\right)$

4) if $\widetilde{\max}_{w_{k+1,i}}<\frac{1}{n}$

5) $\widetilde{\max}_{w_{k+1,i}}=\frac{1}{n}$

6) end if

7) if $\widetilde{\min}_{w_{k+1,i}}>\frac{1}{n}$

8) $\widetilde{\min}_{w_{k+1,i}}=\frac{1}{n}$

9) end if

10) 对加噪完的样本权值进行归一化处理，得到 $w'_{k+1,i}$

将 DPweight_dynamic 算法与 AdaBoost 算法结合，得到 DPAda 算法。

算法 3 DPAda 算法

输入 训练集 T，迭代次数 m，隐私预算 B

输出 最终分类器 $G(x)$

1) 初始化样本权值分布 D_1

2) for $k=1$ to m

3) 训练弱学习器 $G_k(x)$

4) 计算 $G_k(x)$ 的分类误差率 e_k

5) if $e_k>0.5$ then break

6) 计算 $G_k(x)$ 的权重 α_k

7) 计算更新后的样本权值 $w_{k+1,i}$，并进行归一化

8) 执行算法 2，获得样本权值 $w'_{k+1,i}$

9) end for

10) 输出最终分类器 $G(x)=\mathrm{sign}\left(\sum_{k=1}^{m}\alpha_k G_k(x)\right)$

3.2 迭代过程中的噪声控制

若对每轮最大权值或最小权值 $w_{k+1,i}$ 都添加噪声，则在多轮迭代后，当前的噪声会叠加到后续的分类器构造中，累加后的噪声会极大程度地影响最终分类器结果的质量。

此外，AdaBoost 算法在第一轮迭代后，权值分布两极化，即只有最大权值和最小权值。在对最大权值和最小权值添加 2 种不同的噪声时，由于拉普拉斯噪声是随机的，且有正有负，就可能造成最大权值添加了负噪声，最小权值添加了正噪声。但由于最小权值在添加了正噪声后不可能超过权值均值，最大权值在添加了负噪声后不可能小于权值均值，因此就会出现最小权值或最大权值变为均值的情况。为避免这种极端情况的发生，在后文的算法中均从第二轮开始加噪。

基于此，本文提出 3 种可以有效控制噪声叠加过多的注入方案，以减小每轮加噪带来的影响，提高模型的可用性和准确性。

3.2.1 基于摆动数列的噪声注入

定义 5 摆动数列[29]。一个数列从第 2 项起，有些项大于它的前一项，有些项小于它的前一项，这样的数列叫做摆动数列。通过寻找摆动的平衡位置与摆动的振幅，可以得到摆动数列的通项公式。

假设有数列 $a,b,a,b,\cdots$，则该数列的平衡位置为 $\frac{a+b}{2}$，振幅为 $\frac{b-a}{2}$，通过 $(-1)^n$ 或 $(-1)^{n+1}$ 调节，则该数列的通项为

$$a_n=\frac{a+b}{2}+(-1)^n\frac{b-a}{2} \tag{17}$$

本文选取 $a=0,b=1$ 的摆动数列，根据摆动数列第 n 项的值，判断是否为 AdaBoost 算法的第 n 轮添加噪声。此时该数列的通项为

$$a_n=\frac{1}{2}+(-1)^n\frac{1}{2} \tag{18}$$

将基于摆动数列的噪声注入与 DPAda 算法结合，得到 DPAda_Swing 算法。

算法 4　DPAda_Swing 生成算法

输入　训练集 T，迭代次数 m，隐私预算 B

输出　最终分类器 $G(x)$

1) 初始化样本权值分布 D_1

2) for　$k=1$　to　m

3)　　训练弱学习器 $G_k(x)$

4)　　计算 $G_k(x)$ 的分类误差率 e_k

5)　　if　$e_k>0.5$　then break

6)　　计算 $G_k(x)$ 的权重 α_k

7)　　计算更新后的样本权值 $w_{k+1,i}$，并进行归一化

8)　　计算摆动数列第 k 项值 a_k

9)　　if　$a_k==1$

10)　　执行算法 2，获得样本权值 $w'_{k+1,i}$

11)　end if

12) end for

13) 输出最终分类器 $G(x)=\mathrm{sign}\left(\sum_{k=1}^{m}\alpha_k G_k(x)\right)$

3.2.2　基于随机响应的噪声注入

由于摆动数列稳定地、间隔性地输出相同值，不具有随机性，因此本文提出基于随机响应的噪声注入。

定义 6　随机响应。一随机实验在同样的条件下重复并相互独立地进行，该随机实验只有 2 种可能结果：结果 a 和结果 b。随机数 x 在 (a,b) 上服从均匀分布，即 $x\sim\mathrm{U}(a,b)$，$\mathrm{E}(x)=\dfrac{a+b}{2}$，给定以下规则

$$h(x)=\begin{cases}a, a<x<\dfrac{a+b}{2}\\ b, \dfrac{a+b}{2}\leqslant x<b\end{cases}\tag{19}$$

则称 $h(x)$ 满足随机响应，即随机实验满足随机响应。

本文在 AdaBoost 算法的每轮迭代过程中，生成一随机数 x，其中 x 服从 $a=0$，$b=1$ 标准均匀分布，即 $x\sim\mathrm{U}(0,1)$，$\mathrm{E}(x)=0.5$，则每轮迭代 $h_k(x)$ 满足随机响应

$$h_k(x)=\begin{cases}0, 0<x<0.5\\ 1, 0.5\leqslant x<1\end{cases}\tag{20}$$

根据随机响应第 k 轮迭代输出的值，判断是否为 AdaBoost 算法的第 k 轮添加噪声。

将基于随机响应的噪声注入与 DPAda 算法结合，得到 DPAda_Random 算法。

算法 5　DPAda_Random 生成算法

输入　训练集 T，迭代次数 m，隐私预算 B

输出　最终分类器 $G(x)$

1) 初始化样本权值分布 D_1

2) for　$k=1$　to　m

3)　　训练弱学习器 $G_k(x)$

4)　　计算 $G_k(x)$ 的分类误差率 e_k

5)　　if　$e_k>0.5$　then break

6)　　计算 $G_k(x)$ 的权重 α_k

7)　　计算更新后的样本权值 $w_{k+1,i}$，并进行归一化

8)　　根据随机响应，获得随机数 $x(0<x<1)$

9)　　计算 $h_k(x)$ 的值

10)　if　$k>1$　　//从第二轮开始加噪

11)　　if　$h_k(x)==1$

12)　　执行算法 2，获得样本权值 $w'_{k+1,i}$

13)　　end if

14)　end if

15) end for

16) 输出最终分类器 $G(x)=\mathrm{sign}\left(\sum_{k=1}^{m}\alpha_k G_k(x)\right)$

3.2.3　基于改进的随机响应的噪声注入

由于随机响应在每轮迭代中较随机地生成 a 和 b，即可能出现连续几轮的输出值都为 a 或 b，因此，本文提出了一种改进的随机响应。根据第 k 轮迭代输出的值，判断是否为 AdaBoost 算法的第 k 轮添加噪声。

定义 7　改进的随机响应。若第 i 次迭代的输出值 $h_i(x)=a$，第 $i+1$ 次迭代的输出值 $h_{i+1}(x)=b$，且第 $i+2$ 次～第 $j-1$ 次迭代中 $(j>i+2)$，每次输出值 $h_k(x)$ 均为 b，直至第 j 次迭代的输出值 $h_j(x)=a$，则将第 $i+2$ 次～第 $j-1$ 次迭代的输出值 $h_k(x)$ 全置 a。

取 $a=0, b=1$，将基于改进的随机响应的噪声注入与 DPAda 算法结合，得到 DPAda_Improved 算法。

算法 6　DPAda_Improved 生成算法

输入　训练集 T，迭代次数 m，隐私预算 B

输出　最终分类器 $G(x)$

1) 初始化样本权值分布 D_1

2) for　$k=1$　to　m

3)　　训练弱学习器 $G_k(x)$

4) 计算 $G_k(x)$ 的分类误差率 e_k
5) if $e_k > 0.5$ then break
6) 计算 $G_k(x)$ 的权重 α_k
7) 计算更新后的样本权值 $w_{k+1,i}$，并进行归一化
8) 根据随机响应，获得随机数 $x(0 < x < 1)$
9) 计算 $h_k(x)$ 的值
10) if $k > 1$
11) 根据改进的随机响应，更新 $h_k(x)$ 的值
12) if $h_k(x) == 1$
13) 执行算法 2，获得样本权值 $w'_{k+1,i}$
14) end if
15) end if
16) end for
17) 输出最终分类器 $G(x) = \text{sign}\left(\sum_{k=1}^{m} \alpha_k G_k(x)\right)$

3.3 算法分析

3.3.1 加噪轮数分析

定理 1 令 R 为算法累计添加噪声的轮数，则 $\begin{cases} R_{\text{DPAda\_Improved}} \leqslant R_{\text{DPAda\_Swing}} < R_{\text{DPAda}} \\ R_{\text{DPAda\_Random}} < R_{\text{DPAda}} \end{cases}$。

证明 DPAda 算法具有从第二轮起每轮都添加拉普拉斯噪声的特点，则

$$R_{\text{DPAda}} = m - 1 \tag{21}$$

摆动数列可以稳定地输出 0101…，则 DPAda_Swing 算法具有间隔性添加拉普拉斯噪声的特点，因此

$$R_{\text{DPAda\_Swing}} = \left\{ \begin{matrix} \frac{m}{2}, m = 2n \\ \frac{m-1}{2}, m = 2n+1 \end{matrix} \right\}, n \in N \tag{22}$$

随机响应在输出 0 与 1 时太过随机，因此，DPAda_Random 具有连续加噪或连续不加噪的特点，假设 DPAda_Random 算法添加拉普拉斯噪声共 s 次，则

$$R_{\text{DPAda\_Random}} = s \leqslant m - 1 \tag{23}$$

改进的随机响应有效规避了随机响应连续几轮加噪或不加噪的缺点，且具有间隔随机轮次加噪的优点，因此假设 DPAda_Improved 算法添加拉普拉斯噪声共 t 次，则

$$R_{\text{DPAda\_Improved}} = t < m - 1 \tag{24}$$

因此可得

$$\begin{cases} R_{\text{DPAda\_Improved}} \leqslant R_{\text{DPAda\_Swing}} < R_{\text{DPAda}} \\ R_{\text{DPAda\_Random}} \leqslant R_{\text{DPAda}} \end{cases} \tag{25}$$

证毕。

3.3.2 算法时间复杂度分析

DPAda、DPAda_Swing、DPAda_Random 和 DPAda_Improved 算法的时间复杂度估计为

$$T_{\text{DPAda}} = T_{\text{DPAda\_Swing}} = T_{\text{DPAda\_Random}} = T_{\text{DPAda\_Improved}} = O(m(dn)) \tag{26}$$

其中，m 为迭代次数，d 为训练集中属性数量，n 为训练集中样本个数。

3.3.3 算法隐私性分析

假设共有 m 轮迭代，对于第 k 轮迭代后更新的样本权值 $w_{k+1,i}$，$|\max_{w_{k+1,i}}|$ 为最大权值的个数，$|\min_{w_{k+1,i}}|$ 为最小权值的个数。

定理 2 DPAda、DPAda_Swing、DPAda_Random 和 DPAda_Improved 算法满足 $B-$ 差分隐私。

证明 由于 $|\max_{w_{k+1,i}}| + |\min_{w_{k+1,i}}| \leqslant N$（$N$ 为训练集中样本个数），因此令 $|\max_{w_{k+1,i}}| + |\min_{w_{k+1,i}}| = t$，将 $\min_{w_{k+1,i}}$ 和 $\max_{w_{k+1,i}}$ 升序排序，并重新编号，得到 $w_{k+1,j}, 1 \leqslant j \leqslant t$，又因为

$$\sum_{i=1}^{N} w_{k+1,i} = 1 \tag{27}$$

所以，有

$$\sum_{j=1}^{t} w_{k+1,j} \leqslant 1 \tag{28}$$

由定义 3 差分隐私的串行组合性质可知

$$\sum_{k=1}^{m} \sum_{j=1}^{t} \frac{B}{m} w_{k,j} \leqslant B \tag{29}$$

因此，DPAda、DPAda_Swing、DPAda_Random 和 DPAda_Improved 算法满足 $B-$ 差分隐私。证毕。

4 实验及分析

4.1 实验数据

实验所选数据集来源于公开的 UCI 机器学习数据库中的 Adult 数据集和 Bank Marketing 数据集，如表 2 所示。

表 2　　实验数据集信息

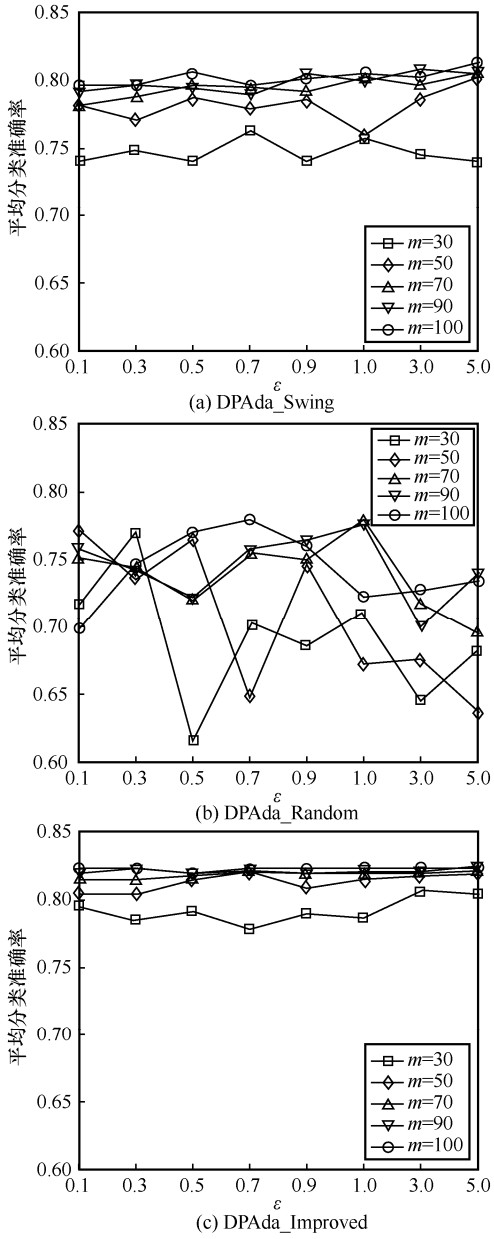
数据集名称	样本数	连续型特征数	离散型特征数	类数
Adult_train	325 61	6	8	2
Adult_test	162 81	6	8	2
Bank Marketing	411 88	10	10	2

通过对 Adult 数据集的属性集进行分析发现，属性 native-country 为非美国籍的样本数仅占总样本数的 16%，因此本文实验仅选取美国籍的样本记录进行训练、测试。此外，属性 fnlwgt 的值是通过对和当前样本具有相似人口特征的人进行加权统计而得出的人口总数，该属性对本文的实验结果无帮助，因此将它删除。本文选取 Adult_train 作为训练集，Adult_test 作为测试集，选取 Bank Marketing 数据集的 70%作为训练集，30%作为测试集。

4.2　实验结果与分析

在相同参数下，每组实验重复运行 50 次取平均值来计算算法的平均分类准确率。x 轴为总的隐私预算 $\varepsilon = B$，y 轴为算法平均分类准确率，其中 $B = 0.1, 0.3, 0.5, 0.7, 0.9, 1.0, 3.0, 5.0$，迭代次数 $m = 30, 50, 70, 90, 100$。

图 2 和图 3 分别是不同数据集下 DPAda_Swing 算法、DPAda_Random 算法与 DPAda_Improved 算法在不同迭代次数下的平均分类准确率对比。观察图 2 和图 3 不难发现，DPAda_Swing 算法与 DPAda_Improved 算法的分类准确率虽偶有波动，但整体趋势基本随着 ε 的增加而增长，而 DPAda_Random 算法的平均分类准确率波动很大，且当迭代次数 m 较小时，波动特别明显。这是因为 DPAda_Random 算法在给样本权值 $w_{k+1,i}$ 添加噪声时，根据随机响应的输出特点，在输出 0 与 1 时太过随机，即可能连续几次迭代的输出值都为 1，从而连续加噪，因此无法有效控制是否对第 k 轮迭代进行加噪。m 越小，越难克服随机性带来的影响，当 m 较大时，可以通过多轮迭代来减少随机性带来的影响。DPAda_Swing 算法能够根据摆动数列稳定的输出值 0101…，来间隔性地添加噪声。对于 DPAda_Improved 算法，虽然在输出 0 与 1 时仍具有随机性，但根据定义 7，相较于 DPAda_Random 算法中的随机响应，DPAda_Improved 算法使用改进的随机响应，有效减少了添加噪声的次数，达到了小于或等于 DPAda_Swing 算法中基于摆动数列添加噪声的轮数。观察图 2(a)和图 3(c)、图 3(a)和图 3(c)可以看出，在不同迭代次数 m 下，DPAda_Improved 算法的分类准确率均高于 DPAda_Swing 算法。



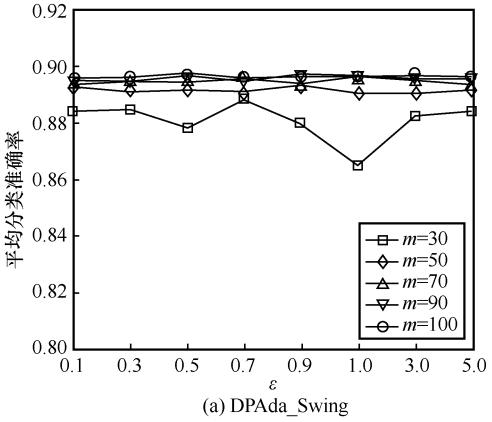
(a) DPAda_Swing

(b) DPAda_Random

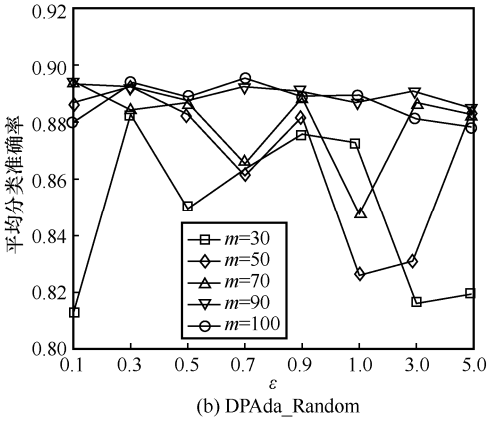
(c) DPAda_Improved

图 2　Adult 数据集下 3 种算法的平均分类准确率

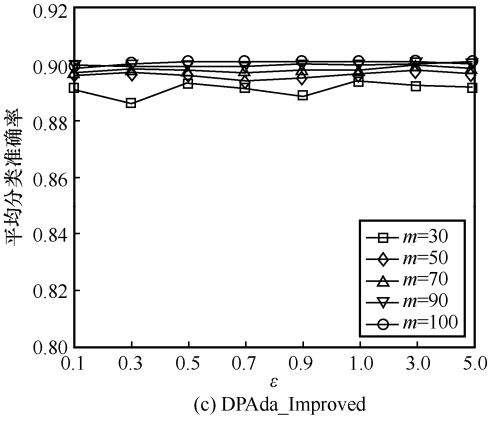
本文还设置了在 $m = 50, 70, 90$ 和 $\varepsilon = 0.1, 0.3, 0.5, 0.7, 0.9, 1.0, 3.0, 5.0$ 条件下，将不结合差分隐私的 AdaBoost 算法、给类标签添加差分隐私保护的 DP-AdaBoost 算法[22]、给 CART 树的内部节点和叶子节点添加差分隐私保护的 CART-DPsAdaBoost 算法[30]、从第二轮起每轮添加差分隐私保护的 DPAda 算法与本文提出的 DPAda_Swing、DPAda_Random、DPAda_Improved 算法进行比较，实验结果如图 4～图 6 所示。



(a) DPAda_Swing

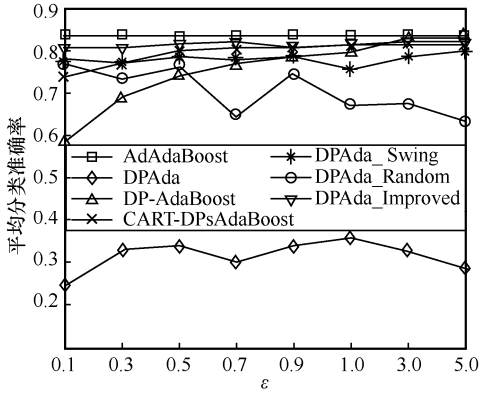


(b) DPAda_Random

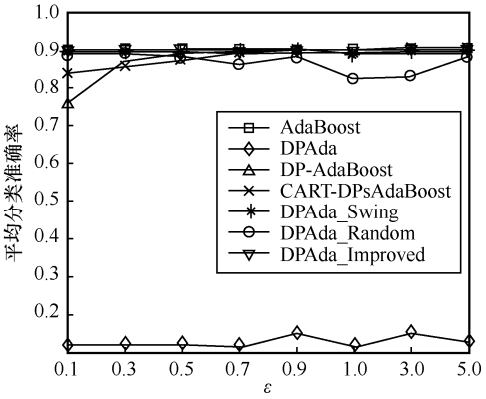


(c) DPAda_Improved

图 3　Bank Marketing 数据集下 3 种算法的平均分类准确率



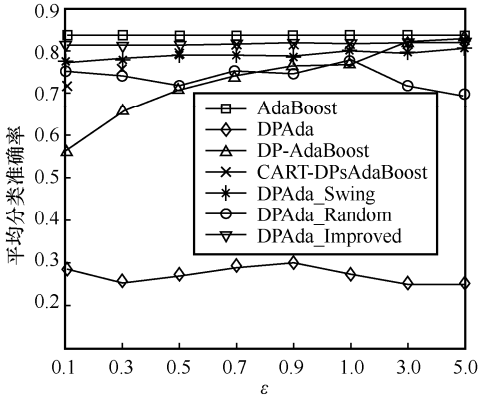
(a) Adult



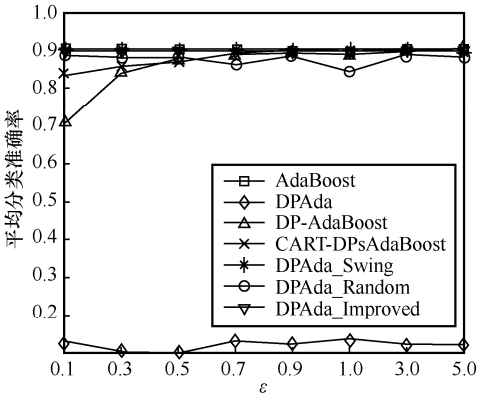
(b) Bank Marketing

图 4　m=50 时不同算法平均分类准确率

从图 4～图 6 可以看出，DPAda 算法的分类准确率很低，并随着迭代次数 m 的增加大体上呈下降趋势。这是由于 DPAda 算法除第一轮迭代外，对之后每轮迭代都连续加噪。因此，在迭代次数 m 增加时，噪声便会随着迭代数的增加而叠加，最后使 DPAda 算法的分类准确率不增反降。当 m 一定时，DP-AdaBoost、CART-DPsAdaBoost、DPAda_Swing 和 DPAda_Improved 算法的分类准确率都随着 ε 的增加而增加；但当 ε 一定时，只有 DP-AdaBoost 算法的分类准确率随着 m 的增加而



(a) Adult



(b) Bank Marketing

图 5　m=70 时不同算法平均分类准确率

下降。这是因为 DP-AdaBoost 算法[22]添加噪声的对象为弱学习器的类标签，它并没有参与迭代过程，因此当 ε 一定时，随着迭代次数 m 的增加，弱学习器越多，分配到每一轮迭代的隐私预算减小，加噪就越明显，分类准确率就越低。CART-DPsAdaBoost 算法[30]是在形成 CART 基分类器的过程中，对 CART 的叶子节点和内部节点进行加噪，当基分类器个数增加时，噪声对构造最终分类器的影响逐渐减小。因此，随着 m 的增加，分类准确率逐渐提升；而 DPAda_Swing 和 DPAda_Improved 算法中添加噪声的对象为样本权值，且结合的是动态隐私预算分配策略 DPweight_dynamic，该策略只对最大样本权值和最小样本权值加噪，保证了在添加噪声的过程中，最大样本权值在加噪后不会小于样本权值均值，最小样本权值在加噪后不会大于样本权值均值，因此仍能保证样本权值的主要分布。此外，DPAda_Swing 和 DPAda_Improved 算法根据摆动数列和改进的随机响应，有效控制了加噪过程为非连续加噪，给予了弱学习器矫正收敛的机会，因此 DPAda_Swing 和 DPAda_Improved 算法的分类准确率随着迭代次数 m 的增加而增加。

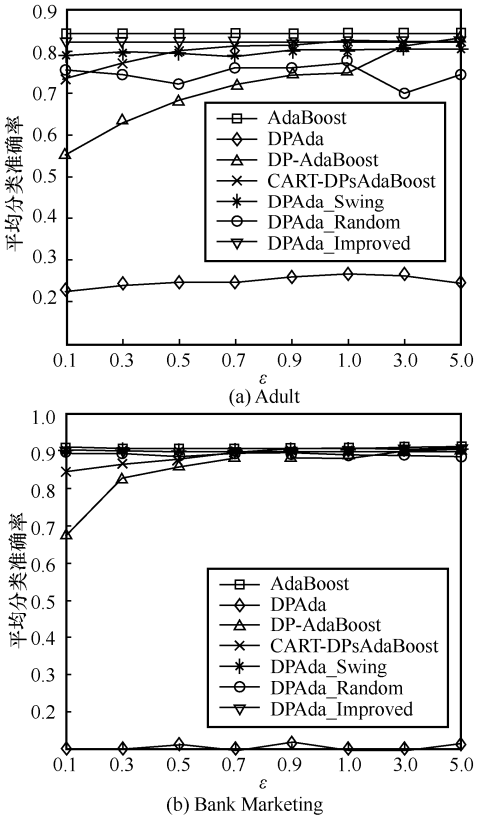


图 6　m=90 时不同算法平均分类准确率

通过观察图 4～图 6 也不难看出，在不同迭代次数 m 和隐私预算 ε 下，在所有对比算法中，DPAda_Improved 算法在分类结果上表现得最稳定，分类准确率最接近不添加噪声的 AdaBoost 算法，实现了在保护隐私的同时，仍拥有更高的分类准确率。

综上，DPAda_Improved 算法与对弱学习器类标签添加噪声的 DP-AdaBoost 算法[22]相比，不会有随着迭代次数的增加准确率反而下降的问题；在分类准确率的整体表现上，优于给 CART 树添加噪声的 CART-DPsAdaBoost 算法[30]；有效解决了 DPAda 算法由于连续加噪带来的噪声不断叠加、精确度极低、模型无法收敛导致模型不可用的问题；不仅在分类准确率上表现得比其他算法要好，还保留了 DPAda_Random 算法的随机轮次加噪的特点，并有效地减少了随机性带来的影响，达到了小于或等于基于摆动数列的 DPAda_Swing 算法中添加噪声的轮数。最后，DPAda_Improved 算法很好地做到了隐私保护与数据效用的均衡。

5　结束语

本文提出了一种基于目标扰动的差分隐私 AdaBoost 算法，与现有的差分隐私 AdaBoost 算法相比，本文算法的贡献如下。1) 选择了迭代过程中每个样本的样本权值作为添加差分隐私保护的对象，解决了在 AdaBoost 算法多轮迭代过程中加噪的主要问题，即算法迭代带来的噪声叠加导致最终无法收敛，模型不可用，准确率极低的问题。2) 设计了一种动态的隐私预算分配策略。3) 给出了严谨的敏感度计算。4) 提出了 3 种有效控制噪声叠加过多的算法，并通过实验证明了相较于 DPAda_Random 算法和 DPAda_Swing 算法，DPAda_Improved 算法在拥有更高的分类准确率的同时，仍能实现数据的隐私保护。下一步，笔者将考虑把差分隐私应用于其他集成学习算法。

参考文献：

[1] 陈天莹, 陈剑锋. 大数据环境下的智能数据脱敏系统[J]. 通信技术, 2016, 49(7): 915-922.
CHEN T Y, CHEN J F. Intelligent data masking system for big data productive environment[J]. Communications Technology, 2016, 49(7): 915-922.

[2] SWEENEY L. k-anonymity: a model for protecting privacy[J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2002, 10(5): 557-570.

[3] GENTRY C. Fully homomorphic encryption using ideal lattices[C]//Proceedings of the 41st Annual ACM Symposium on Theory of Computing. New York: ACM Press, 2009: 169-178.

[4] DWORK C. Differential privacy[C]//Proceedings of the 33rd International Conference on Automata, Languages and Programming (ICALP'06). Berlin: Springer, 2006: 1-12.

[5] DWORK C, AGRAWAL M, DU D, et al. Differential Privacy: a survey of results[C]// International Conference on Theory and Applications of Models of Computation. Berlin: Springer, 2008: 1-19.

[6] DWORK C, LEI J. Differential privacy and robust statistics[C]//Proceedings of the 41st Annual ACM Symposium on Theory of Computing. New York: ACM Press, 2009: 371-380.

[7] DWORK C, MCSHERRY F, NISSIM K, et al. Calibrating noise to sensitivity in private data analysis[C]//Theory of Cryptography. Berlin: Springer, 2006: 265-284.

[8] MCSHERRY F, TALWAR K. Mechanism design via differential privacy[C]//Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science. Piscataway: IEEE Press, 2007: 94-103.

[9] DWORK C, ROTH A. The algorithmic foundations of differential privacy[J]. Foundations and Trends in Theoretical Computer Science, 2013, 9(3/4): 211-407.

[10] JAGANNATHAN G, PILLAIPAKKAMNATT K, WRIGHT R N. A practical differentially private random decision tree classifier[C]//Proceedings of 2009 IEEE International Conference on Data Mining Workshops. Piscataway: IEEE Press, 2009: 114-121.

[11] PATIL A, SINGH S. Differential private random forest[C]//Proceedings of 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI). Piscataway: IEEE Press, 2014: 2623-2630.

[12] FLETCHER S, ISLAM M Z. A differentially private decision forest[J]. Expert Systems with Applications, 2017, 78: 16-31.

[13] 穆海蓉, 丁丽萍, 宋宇宁, 等. DiffPRFs: 一种面向随机森林的差分隐私保护算法[J]. 通信学报, 2016, 37(9): 175-182.
MU H R, DING L P, SONG Y N, et al. DiffPRFs: random forest under differential privacy[J]. Journal on Communications, 2016, 37(9): 175-182.

[14] 李远航, 陈先来, 刘莉, 等. 面向差分隐私保护的随机森林算法[J]. 计算机工程, 2020, 46(1): 93-101.
LI Y H, CHEN X L, LIU L, et al. Random forest algorithm for differential privacy protection[J]. Computer Engineering, 2020, 46(1): 93-101.

[15] 邓蔚, 陈秀婷, 张清华, 等. 基于树模型的差分隐私保护算法[J]. 重庆邮电大学学报(自然科学版), 2020, 32(5): 848-856.
DENG W, CHEN X T, ZHANG Q H, et al. Differential privacy protection algorithms based on tree model[J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2020, 32(5): 848-856.

[16] FLETCHER S, ISLAM M Z. Differentially private random decision forests using smooth sensitivity[J]. Expert Systems With Applications, 2017, 78: 16-31.

[17] GUAN Z T, SUN X W, SHI L Y, et al. A differentially private greedy decision forest classification algorithm with high utility[J]. Computers & Security, 2020, 96: 101930.

[18] FLETCHER S, ISLAM M Z. A differentially private random decision forest using reliable signal-to-noise ratios[C]//Australasian Joint Conference on Artificial Intelligence. Berlin: Springer, 2015: 192-203.

[19] LI X X, LIU J, LIU S F, et al. Differentially private ensemble learning for classification[J]. Neurocomputing, 2021, 430: 34-46.

[20] GAMBS S, KÉGL B, AÏMEUR E. Privacy-preserving boosting[J]. Data Mining and Knowledge Discovery, 2007, 14(1): 131-170.

[21] LI Y, BAI C X, REDDY C K. A distributed ensemble approach for mining healthcare data under privacy constraints[J]. Information Sciences, 2016, 330: 245-259.

[22] 沈思倩. 关于差分隐私保护分类算法的研究[D]. 南京: 南京航空航天大学, 2017.
SHEN S Q. Research on the differential privacy classification algorithms[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2017.

[23] 贾俊杰, 邱万勇, 马慧芳. 差分隐私保护约束下集成分类算法的研究[J]. 信息安全学报, 2021, 6(4): 106-118.
JIA J J, QIU W Y, MA H F. Research on an ensemble classification algorithm under differential privacy[J]. Journal of Cyber Security, 2021, 6(4): 106-118.

[24] LI Q B, WU Z M, WEN Z Y, et al. Privacy-preserving gradient boosting decision trees[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(1): 784-791.

[25] FREUND Y, SCHAPIRE R E. A decision-theoretic generalization of on-line learning and an application to boosting[J]. Journal of Computer and System Sciences, 1997, 55(1): 119-139.

[26] FREUND Y, SCHAPIRE R E. Experiments with a new Boosting algorithm[C]//Proceedings of the 13th Conference on Machine Learning. San Francisco: Morgan Kaufmann Publishers Inc., 1996: 148-156.

[27] 曹莹, 苗启广, 刘家辰, 等. AdaBoost 算法研究进展与展望[J]. 自动化学报, 2013, 39(6): 745-758.
CAO Y, MIAO Q G, LIU J C, et al. Advance and prospects of Ada-

Boost algorithm[J]. Acta Automatica Sinica, 2013, 39(6): 745-758.

[28] FRIEDMAN J, HASTIE T, TIBSHIRANI R. Additive logistic regression: a statistical view of boosting [J]. The Annals of Statistics, 2000, 28(2): 337-407.

[29] 燕会霞. 简谐振动与摆动数列的通项公式[J]. 数学教学, 2006(9): 19-20.
YAN H X. General formula of simple harmonic vibration and swing sequence[J]. Mathematics teaching, 2006(9): 19-20.

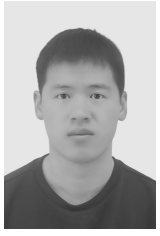
[30] 邱万勇. 基于差分隐私的集成分类算法及应用研究[D]. 兰州: 西北师范大学, 2021.
QIU W Y. A study of ensemble classification algorithms and applications based on differential privacy[D]. Lanzhou: Northwest Normal University, 2021.

[作者简介]

张淑芬（1972− ），女，河北唐山人，华北理工大学教授、硕士生导师，主要研究方向为云计算、智能信息处理、数据安全与隐私保护。

董燕灵（1998− ），女，浙江宁波人，华北理工大学硕士生，主要研究方向为数据安全与隐私保护。

徐精诚（1996− ），男，江苏常州人，华北理工大学硕士生，主要研究方向为数据安全与隐私保护。



王豪石（1997− ），男，河北邢台人，华北理工大学硕士生，主要研究方向为数据安全与隐私保护。