计算机研究与发展
Journal of Computer Research and Development
DOI:10.7544/issn1000-1239.20200717
59(2):430-439, 2022

满足本地差分隐私的分类变换扰动机制

朱素霞　王　蕾　孙广路
(哈尔滨理工大学计算机科学与技术学院　哈尔滨　150080)
(哈尔滨理工大学信息安全与智能技术研究中心　哈尔滨　150080)
(zhusuxia@hrbust.edu.cn)

A Perturbation Mechanism for Classified Transformation Satisfying Local Differential Privacy

Zhu Suxia, Wang Lei, and Sun Guanglu
(*School of Computer Science and Technology, Harbin University of Science and Technology, Harbin* 150080)
(*Research Center of Information Security and Intelligent Technology, Harbin University of Science and Technology, Harbin* 150080)

Abstract　As the state-of-the-art privacy protection technology, local differential privacy is widely used to compute the mean value of continuous numerical data. The perturbation mechanism will directly affect the accuracy of the mean value. In order to further improve the accuracy of mean value estimation, a perturbation mechanism for classified transformation satisfying differential privacy is proposed. In this mechanism, continuous numerical data is divided into transformation range, which is then segmented. What's more, it transforms the segmentation into one-dimensional binary category data. After transformation, the mechanism of random response is used to perturb the data. More importantly, it extracts the value randomly as well as uniformly from the numerical segment identified by the perturbation data as the perturbed value. The experimental results of mean value estimation in both real data and synthetic data show that the mechanism proposed in the paper greatly improves the accuracy. In addition, this perturbation mechanism is used to build a mini-batch gradient descent algorithm satisfying local differential privacy and the linear regression learning task is completed successfully. The experimental results show that this method not only is superior to other existing mechanisms but also can obtain a smaller mean square error at the same time.

Key words　local differential privacy; data transformation; mean value estimation; mini-batch gradient descent; random response

摘　要　本地差分隐私作为一种隐私保护技术,被广泛用于连续数值型数据的均值估计,使用的扰动机制将直接影响均值的准确度.为进一步提高均值估计的准确性,提出了一种满足差分隐私的分类变换扰动机制.该机制对连续数值型数据划分变换范围并进行分段,根据分段将其变换为1维二元分类数据.

收稿日期:2020-09-15;**修回日期**:2021-03-04
基金项目:国家自然科学基金项目(61502123);黑龙江省留学归国人员科学基金项目(LC2018030);黑龙江普通高校基本科研业务专项资金(JMRH2018XM04);黑龙江省自然科学基金项目(LH2021F032)
This work was supported by the National Natural Science Foundation of China (61502123), the Science Foundation for Returned Overseas Students of Heilongjiang Province (LC2018030), the Heilongjiang University Special Foundation for Basic Scientific Research (JMRH2018XM04), and the Natural Science Foundation of Heilongjiang Province (LH2021F032).
通信作者:孙广路(sunguanglu@hrbust.edu.cn)

转换后使用随机响应机制进行扰动,再根据扰动后的数据标识的数值段从中随机均匀抽取数值作为扰动值.在真实数据和合成数据中的均值估计实验结果表明该机制极大地提高了准确性.除此之外,将分类变换扰动机制用于构建满足本地差分隐私的小批量梯度下降算法,并完成线性回归学习任务,实验结果证明该方法同样优于其他已有机制,可得到更小的均方误差.

关键词 本地差分隐私;数据转换;均值估计;小批量梯度下降;随机响应

中图法分类号 TP309

随着云计算和大数据技术的发展,用户端产生的海量数据被服务器收集起来进行各种数据分析任务.虽然对这些数据进行分析可以为人们带来可观的效益,但是却造成了用户隐私暴露的问题.差分隐私由于其强大的隐私保障已经成为了一种标准的隐私保护模型.随着差分隐私的广泛使用,服务器变得越来越重要.然而,在真实世界中保证所有服务器都是可信的是不实际的,而不可信的服务器可能会因为某些原因泄露用户的隐私.为了解决这一问题,本地差分隐私[1-2]作为一种新的隐私保护技术被提出用来保护用户的隐私,其最典型的扰动机制是随机响应机制.在本地差分隐私中,服务器假设是不可信的,每个用户端对本地数据进行扰动使其满足本地差分隐私,然后再将扰动后的数据发送给服务器.服务器对收集的噪声数据进行计算,得到所需的统计信息.本地差分隐私方法可以在获得较为准确的统计信息的同时有效地对用户的数据进行保护,从而避免了用户隐私泄露的问题.

本地差分隐私由于其强大的隐私保证,已经被运用到很多实际的工作任务中.例如谷歌的 Chrome 浏览器使用的 RAPPOR(randomized aggregatable privacy-preserving ordinal response)方法[3]以及微软的遥测数据采集[4].这些方法使得在保护用户隐私的同时,可以利用用户的数据进行分析得到有效的统计结果.人们针对不同的数据类型提出了适用于不同计算任务的本地差分隐私框架,目前主要研究的统计任务有均值估计和频率估计.例如,谷歌 Chrome 使用的 RAPPOR[3]方法是针对分类型数据的频率估计,Nguyên 等人[5]提出了针对离散型数据的均值估计的扰动方法 Harmony. Ye 等人[6]针对键值数据类型提出了 PrivKVM 方法,可以在满足本地差分隐私的同时估计键的频率以及键对应的所有值的均值.针对连续型数值数据的均值估计,Duchi 等人[7]的方法对数据进行扰动之后一共有 2 种可能得到的扰动值.由于这 2 种扰动值的绝对值都大于 1,即不管隐私预算如何变化,其方差始终大于 1.所以当隐私预算比较大时,该方法得到的均值估计的准确性相比于拉普拉斯方法要更差.随后,Wang 等人[8]针对 Duchi 方法的缺点,提出了分段机制(piecewise mechanism, PM).该机制不同的是,其扰动输出为一段连续值,且这段连续值的中间部分有更高的概率输出.虽然分段机制改善了 Duchi 方法中存在的问题,但是当隐私预算较小时,该方法并没有很好地提高均值估计的准确性,其最坏情况下噪声方差仍与 Duchi 方法接近.

除此之外,机器学习作为当前比较热门的学习领域,其中也涉及了大量用户的隐私保护问题.为了更好地保护用户的隐私,可以将其与本地差分隐私的相关扰动机制结合使用.目前,机器学习中较常使用的隐私保护方式是在模型训练时对用户梯度进行扰动,服务器收集扰动后的梯度进行更新.例如,Nguyên 等人[5]将 Harmony 运用到了随机梯度下降中,对每次迭代的梯度进行扰动,并且证明了本地差分隐私下的小批量梯度下降要优于随机梯度下降.Wang 等人[8]则利用多维数据扰动的方式,将分段机制用于迭代中的梯度扰动.这些方法虽然在机器学习训练过程中保护了用户的隐私,但由于机制本身的缺点,其训练结果的准确性仍然具有提升空间.

为了改善已有扰动方法引入的准确性问题,论文针对连续型数值数据,提出了一种满足本地差分隐私的分类变换扰动机制(differential classified transformation, DCT).跟已有的方法直接对所属数据类型使用对应的扰动方法进行扰动不同,本文提出的方法首先对数据类型进行了转换,将数值型数据转换为了 1 维二元分类数据,再对分类数据进行扰动.在真实数据以及合成数据中使用该方法进行均值估计,与已有的方法进行对比,可以得到一个更为准确的估计结果.在机器学习的隐私保护中,考虑到本地差分隐私中隐私预算的分配问题,为了在训练中得到更为准确的结果,论文将提出的分类变换机制用于构建满足本地差分隐私的小批量梯度下降,并在该框架下进行线性回归的学习任务.

总的来说,本文主要贡献如下:

1) 提出了一种数据变换扰动方法,并且得到了较好的结果,这给本地差分隐私的扰动提供了一个新方向,可以通过变换数据,使其在提高数据的可用性的同时又保障了用户的隐私;

2) 提出的分类变换机制具有良好的性能,在满足本地差分隐私保证的同时,在均值估计方面可以得到更为准确的结果;

3) 将提出的方法用于构建小批量梯度下降算法,并用该算法完成线性回归的学习任务,使得参与用户的数据受到良好保护的同时,可以得到一个较为准确的模型结果;

4) 在真实的数据集以及合成的数据集上进行实验,以对提出的机制进行评估.实验结果表明,不管是在均值估计还是在经验风险最小化任务中,使用分类变换扰动机制得到的结果误差要小于已有的方法.

1 相关定义

1.1 本地差分隐私

在本地差分隐私中,服务器收集各个用户的数据,并利用数据计算得到所需的统计信息.用户在将数据发送给服务器前,先对本地的数据进行扰动,再将扰动后的数据发送给服务器.服务器无法根据收集的噪声数据来获得用户的隐私信息.隐私预算的大小代表了用户隐私保护程度的强弱,其控制了隐私和效用之间的平衡,一个更小的隐私预算代表了更强的隐私保护程度.本地差分隐私的定义如下:

定义 1. ε-本地差分隐私[1].随机函数 M 满足 ε-本地差分隐私当且仅当域 M 中的任意 2 个输入 t, t'以及对于 M 中的任意可能的输出 t^*,有:

$$Pr(M(t)=t^*)\leqslant e^{\varepsilon}\times Pr(M(t')=t^*), \quad (1)$$

其中 $Pr(\cdot)$代表概率.

本地差分隐私作为差分隐私[9]的分支,提供了比差分隐私还要强大的隐私保障.根据上面的隐私定义,服务器无论具备怎样的背景知识,都无法以高概率从接收到的用户的扰动元组 t^* 来判断用户的真实值是 t 还是 t'.

本地差分隐私中最经典的方法是随机响应机制,该方法主要用来收集用户的敏感数据以获得准确的统计信息,下面举例来介绍这个机制.假设服务器想知道有多少个用户是抽烟的,它会向每个用户发送问题"你抽烟吗?"用户接受到问题后采用抛硬币的方法来决定它的答案.假如硬币的正面朝上,那么用户将真实答案告诉服务器,否则的话它将告诉服务器一个相反的答案.使用该方法,服务器可以根据所有用户的回答得到一个无偏估计.假设每个用户抛硬币正面朝上的概率为 p,即用户正确回答服务器的概率为 p,则其提供错误答案的概率为 $1-p$.为了使该方法满足 ε-本地差分隐私,概率 p 应满足式(2):

$$p=\frac{e^{\varepsilon}}{e^{\varepsilon}+1}. \quad (2)$$

1.2 问题定义

论文主要研究的问题是连续型数值的均值估计,为了后续研究方便,这里简单假设每个用户 u_i 都有 1 个数值型数据 v_i,本文所有使用到的符号如表 1 所示:

Table 1 Symbol Definition

表 1 符号定义

符号	描述
$U=\{u_1,u_2,\cdots,u_n\}$	用户集,用户数量 $n=\|U\|$
$V=\{v_1,v_2,\cdots,v_n\}$	用户值集,v_i 为用户 u_i 的值
ε	扰动机制的隐私预算
$\tilde{V}=\{\tilde{v}_1,\tilde{v}_2,\cdots,\tilde{v}_n\}$	值的扰动输出
$tra(\cdot)$	分类变换
$retra(\cdot)$	分类逆变换
d	数据 v_i 的变换范围

在本地差分隐私中,不同用户可以根据需求使用不同的隐私预算 ε 来保护自己的隐私.在本文中,为了便于分析,假设了一个统一的隐私预算参数值 ε,目标是在满足 ε-本地差分隐私的条件下完成下列 2 种类型的分析任务:

1) 均值估计.对于连续型数值数据,假设包含 n 个用户,需要计算其均值,计算方式为 $\frac{1}{n}\sum_{i=1}^{n}v_i$.

2) 经验风险最小化.论文中主要将线性回归及小批量梯度下降结合使用,并计算模型训练结果的均方误差来评估方法性能.

2 数值数据的均值估计

均值估计是目前本地差分隐私中主要进行研究的统计任务之一,它在统计分析中具有重要作用.本节主要讨论在满足 ε-本地差分隐私的条件下收集用户的数据进行数值属性均值估计的问题,对 2 种已有的方法进行介绍.

1) Duchi 方法

Duchi 等人[7]提出了在本地差分隐私下用来扰动 1 维数值型数据的方法,如算法 1 所示:

算法 1. Duchi 等人[7]的 1 维数值数据扰动方法.

输入:$v_i\in[-1,1]$、隐私预算 ε;

输出:$\bar{v}_i\in\left\{-\frac{(e^\varepsilon+1)}{(e^\varepsilon-1)},\frac{(e^\varepsilon+1)}{(e^\varepsilon-1)}\right\}$.

① 抽取一个伯努利变量 u 满足 $Pr(u=1)=\frac{(e^\varepsilon-1)}{(2e^\varepsilon+2)}\times v_i+\frac{1}{2}$;

② if $u=1$

③ $\quad\bar{v}_i=\frac{e^\varepsilon+1}{e^\varepsilon-1}$;

④ else

⑤ $\quad\bar{v}_i=-\frac{e^\varepsilon+1}{e^\varepsilon-1}$;

⑥ endif

⑦ 返回 $\bar{v}_i$.

根据算法 1,给定一个值 $v_i\in[-1,1]$,算法返回的扰动值只有 2 种:$\frac{e^\varepsilon+1}{e^\varepsilon-1}$和$-\frac{e^\varepsilon+1}{e^\varepsilon-1}$.其中,返回$\frac{e^\varepsilon+1}{e^\varepsilon-1}$的概率为$\frac{(e^\varepsilon-1)}{(2e^\varepsilon+2)}\times v_i+\frac{1}{2}$,返回$-\frac{e^\varepsilon+1}{e^\varepsilon-1}$的概率为$-\frac{(e^\varepsilon-1)}{(2e^\varepsilon+2)}\times v_i+\frac{1}{2}$.除此之外,Duchi 等人[7]证明了该方法得到的扰动输出为输入的无偏估计.服务器在接收了所有用户的扰动输出后,可以计算出所有用户的均值.Duchi 方法虽然一定程度上提高了均值估计的准确性,但是由于其扰动输出的 2 种值的绝对值都大于 1,使得不管隐私预算如何变化,扰动值的方差始终大于 1.所以该方法只有在隐私预算较小时可以得到比较好的效果,当隐私预算较大时其准确性较差.

2) PM

Wang 等人[8]提出的分段机制 PM 是另一种在本地差分隐私下进行均值估计的扰动方法.与 Duchi 方法不同,PM 分段抽取输入数值的扰动值.算法 2 描述了 PM 方法:

算法 2. PM 数值扰动方法[8].

输入:$v_i\in[-1,1]$、隐私预算 ε;

输出:$\bar{v}_i\in[-C,C]$.

① 从$[0,1]$中均匀随机抽取 x;

② if $x<\frac{e^{\varepsilon/2}}{(e^{\varepsilon/2}+1)}$

③ $\quad$从$[l(v_i),r(v_i)]$中随机均匀抽取 $\bar{v}_i$;

④ else

⑤ $\quad$从$[-C,l(v_i))\cup(r(v_i),C]$中随机均匀抽取 $\bar{v}_i$;

⑥ endif

⑦ 返回 $\bar{v}_i$.

从算法 2 可以看出,PM 方法的输出域是一段连续的值$[-C,C]$,其中 $C=\frac{e^{\varepsilon/2}+1}{e^{\varepsilon/2}-1}$.扰动值有更高的概率为输出值域中的中间段的值,较低的概率为两端的值,式(3)给出了其概率密度函数:

$$pdf(\bar{v}_i=x\mid v_i)=\begin{cases}p, & x\in[l(v_i),r(v_i)],\\ \frac{p}{e^\varepsilon}, & x\in[-C,l(v_i))\cup(r(v_i),C],\end{cases}\tag{3}$$

其中

$$p=\frac{e^\varepsilon-e^{(\varepsilon/2)}}{2e^\varepsilon+2},$$

$$l(v_i)=\frac{C+1}{2}\times v_i-\frac{C-1}{2},$$

$$r(v_i)=l(v_i)+C-1.$$

Wang 等人[8]证明了分段方法得到的扰动输出值为输入的无偏估计,并且将该方法用于均值估计能够得到比其他现有方法更为准确的结果.分段机制虽然改善了 Duchi 方法的缺点,当隐私预算较大时可以得到更为准确的结果.但是当隐私预算较小时,其最坏情况下的方差与 Duchi 方法的相近,准确性没有得到很好的提高.

虽然这 2 种方法对已有的方法进行了改善,一定程度上提高了连续数值型数据均值估计的准确性,但是仍然存在缺点,准确性仍具有较大的改善空间.如 Duchi 方法由于输出的扰动值的绝对值都大于 1,所以在隐私预算较大时性能较差;而分段机制虽然对 Duchi 方法进行了改良,但是由于提出的分段机制在隐私预算较小时的最坏情况下方差与 Duchi 方法的接近,所以准确性在隐私预算较小时没有得到提升.

3 分类变换扰动机制

为进一步提高均值估计的准确性,本文提出了满足本地差分隐私的分类变换扰动机制,即 DCT.与 Duchi 和 PM 直接根据无偏估计得到较为准确的结果不同,论文提出使用数据变换的方法使得在满足本地差分隐私的条件下可以得到更为准确的估计

值.该机制不对原数据进行扰动,而是将数据先进行分类变换,对其转换后得到的1维二元分类数据进行扰动.对于分类变换扰动机制,其输入值 $v_i \in [-1,1]$,扰动值的输出范围为$[-A,A]$,其中

$$A=1+d.$$

该机制主要分成3个阶段,分别是分类变换、分类扰动以及分类逆变换.

3.1 分类变换

变换前需要对用户数据进行预处理,这里假设用户拥有的数据为浮点数.为了减少实验时的计算开销,方便后续的数据分析,将用户数据 v_i 标准化到$[-1,1]$.一般情况下,假设该属性的值域为$[L,U]$,式(4)给出了用户计算方式:

$$v'=\frac{2}{U-L}\times v+\frac{L+U}{L-U}. \tag{4}$$

将数据预处理之后,进入到分类变换阶段.分类数据特指一类反映事务类别的数据,分类变换的过程主要是将用户的连续数值型数据转换为1维二元分类数据的过程.考虑到如果转换成高维分类数据,会导致插入的噪声增多从而降低统计分析的准确性,所以将其转换为1维分类数据.该分类数据主要包含2种值:1和0,用来标识不同范围数据段的数据.变换时以用户原始数据为中心,与其距离绝对值小于等于 d 的数值为其变换范围,$2d$ 为该数据变换的范围.取 $m=d/2$,用 m 将该范围划分为4段.也就是说,被分成的4段以该值为中心,满足左右2段对称.其中中间2段标识为同一类数据,用1表示,两端的2段为同一类数据,用0表示.为了使该方法更加贴合实际,距离 d 应随着隐私预算的减小(即隐私强度的增大)而增大,为使得到的数据更具效用性,这里将其设置为 $d=\frac{1}{\alpha\times\varepsilon}$,其中 α 为距离参数,该值越大,数据效用性越高,但同时隐私性越差,其具体取值在后面实验部分进行了分析.该设置使得变换范围可以根据用户隐私需求动态调整,同时引入的 α 参数可以使数据具备更好地效用性.确定扰动范围后,将该数值转换为二值分类数据,式(5)给出了二值化的计算方式:

$$tra(v_i')=\begin{cases}1, & x\in[l_2,R_1],\\ 0, & x\in[l_1,l_2)\cup(R_1,R_2],\end{cases} \tag{5}$$

其中

$$l_1=v'-d,$$
$$l_2=v'-m,$$
$$R_1=v'+m,$$
$$R_2=v'+d.$$

由于该数值位于扰动范围中心,所以在对数值型数据进行二值化时采用随机抽取的方式取值,即该数值对应的分类数值可能为1也可能为0.

3.2 分类扰动

用户 u_i 的数据 v_i 已经转换为了1维二元分类数据,可以在本地直接使用随机扰动机制[10]来对数据进行扰动.这里采用Xia等人[11]提出的对单个位进行扰动的方法,式(6)给出了具体的扰动规则:

$$\tilde{v}_i=\begin{cases}1, & \text{w.p. } 1/2f,\\ 0, & \text{w.p. } 1/2f,\\ v_i, & \text{w.p. } 1-f,\end{cases} \tag{6}$$

其中 f 代表的是扰动时数据改变的概率.也就是说,用户 u_i 的数据 v_i 有 f 的概率会改变,$1-f$ 的概率保持不变.当 $f=\frac{2}{1+e^{\varepsilon}}$ 时,该方法满 ε-本地差分隐私.

性质1. 当 $f=\frac{2}{1+e^{\varepsilon}}$ 时,该扰动方法满足

$$\frac{Pr(\tilde{v}=\hat{v}\mid v=v')}{Pr(\tilde{v}=\hat{v}\mid v=1-v')}\leqslant e^{\varepsilon}. \tag{7}$$

证明. 根据式(6),可以推得

$$Pr(\tilde{v}=1\mid v=1)=1-\frac{1}{2}f, \tag{8}$$

$$Pr(\tilde{v}=1\mid v=0)=\frac{1}{2}f, \tag{9}$$

$$Pr(\tilde{v}=0\mid v=1)=\frac{1}{2}f, \tag{10}$$

$$Pr(\tilde{v}=0\mid v=0)=1-\frac{1}{2}f. \tag{11}$$

取 $f=\frac{2}{1+e^{\varepsilon}}$,则有

$$\frac{Pr(\tilde{v}=\hat{v}\mid v=v')}{Pr(\tilde{v}=\hat{v}\mid v=1-v')}=\frac{1-1/2f}{1/2f}=e^{\varepsilon}. \tag{12}$$

所以当 f 设置为 $\frac{2}{1+e^{\varepsilon}}$ 时,式(12)计算得该扰动满足 ε-本地差分隐私. 证毕.

3.3 分类逆变换

对二元分类数据进行扰动之后,对其扰动输出值进行分类逆变换操作,输出数值型数据.转换规则为

$$retra(\tilde{v})\in\begin{cases}[l_2,R_1], & \tilde{v}=1,\\ [l_1,l_2)\cup(R_1,R_2], & \tilde{v}=0.\end{cases} \tag{13}$$

如果扰动后分类数据为1,则将其转换回数值数据时从中间的2段距离中随机均匀抽取1个值作为其转换后的值,如果分类数据为0则从两端的2段数据中进行均匀抽取.

算法 3. 分类变换扰动机制.

输入:$v_i\in[-1,1]$、隐私预算 ε;

输出:$\bar{v}_i\in[-A,A]$.

① 从[0,1]随机均匀抽取 u;

② if $u<0.5$

③ $\quad v_i'=1$;

④ else

⑤ $\quad v_i'=0$;

⑥ endif

⑦ 从[0,1]中随机均匀抽取 x;

⑧ if $x<\dfrac{e^{\varepsilon}-1}{e^{\varepsilon}+1}$

⑨ $\quad \bar{v}_i'=v_i'$;

⑩ else

⑪ $\quad$ 从{0,1}中随机均匀抽取一个数作为 $\bar{v}_i'$;

⑫ endif

⑬ if $\bar{v}_i'=1$

⑭ $\quad$ 从$[l_2,R_1]$中随机抽取 $\bar{v}_i$;

⑮ else

⑯ $\quad$ 从$[l_1,l_2)\cup(R_1,R_2]$中随机抽取 $\bar{v}_i$;

⑰ endif

⑱ 返回 $\bar{v}_i$.

算法 3 描述了分类变换扰动机制的伪代码.算法中假设输入域是[−1,1],一般情况下使用式(4)中的计算方法对数据进行标准化.算法中采用随机抽取方式将数值数据转换成分类数据后,使用 3.2 节描述的随机响应机制对变换得到的分类数据进行扰动,其具体扰动规则如式(6)所示,最后根据式(13)中提出的规则将扰动后数据逆变换为数值型数据.用户端使用算法 3 对数据进行扰动,将最后得出的数据发送给服务器.服务器在接收了所有用户的数据值后进行均值估计,均值计算方式为 $\dfrac{1}{n}\sum_{i=1}^{n}\bar{v}_i$.对于多维数据的处理,可采用 Wang 等人[8]提出的多维属性抽取的方法,然后再对抽取出来需要提交的属性使用本文提出的机制进行扰动.

引理 1. 算法 3 满足 ε-本地差分隐私.

证明. 根据分类变换和扰动规则,$\bar{v}$ 概率密度函数如式(14)所示:

$$pdf(\bar{v}_i=x\mid v_i)=\begin{cases}\dfrac{1}{2}(1-f)+\dfrac{f}{2}, & x\in[l_2,R_1],\\ \dfrac{f}{2}+\dfrac{1}{2}(1-f), & x\in[l_1,l_2)\cup(R_1,R].\end{cases}\tag{14}$$

对任意 $\bar{v}\in[-A,A]$和任意 2 个输入值 $v,v'\in[-1,1]$,$\varepsilon>0$,有 $\dfrac{pdf(\bar{v}\mid v)}{pdf(\bar{v}\mid v')}=\dfrac{1/2(1-f)+f/2}{1/2f+(1-f)/2}<e^{\varepsilon}$.所以,算法 3 满足 ε-本地差分隐私. 证毕.

4 本地差分隐私下的小批量梯度下降

本节主要研究构建满足 ε-本地差分隐私下的经验风险最小化的机器学习模型,使用梯度下降法实现.参照 Nguyên 等人[5]的对比实验结果,使用小批量梯度下降可以得到比随机梯度下降法更为准确的结果,所以论文使用小批量梯度下降法实现经验风险最小化.构建了满足本地差分隐私的小批量梯度下降法之后,使用其完成线性回归任务来验证该框架性能.

假设每一个用户 u_i 都有一对$\langle \boldsymbol{x}_i,y_i\rangle$,其中 $\boldsymbol{x}_i\in[-1,1]^k$,$y_i\in[-1,1]$.使用 $L(\cdot)$表示损失函数,其表示的是将由参数 $\boldsymbol{x}_i$ 和 y_i 组成的 k 维参数向量 $\boldsymbol{\beta}$ 映射成 1 个实数产生的损失.最终的目标是得到 1 组参数向量 $\boldsymbol{\beta}^*$,其满足条件:

$$\boldsymbol{\beta}^*=\arg\min_{\boldsymbol{\beta}}\left[\frac{1}{n}\sum_{i=1}^{n}L(\boldsymbol{\beta};\boldsymbol{x}_i,y_i)+\frac{\lambda}{2}\|\boldsymbol{\beta}\|_2^2\right],\tag{15}$$

其中 λ 为正则化因子.在本文中,主要考虑线性回归的损失函数,损失函数如式(16)所示:

$$L(\boldsymbol{\beta};\boldsymbol{x}_i,y_i)=(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}-y_i)^2,\tag{16}$$

在机器学习中,获得 $\boldsymbol{\beta}^*$ 的最普遍的计算方式是使用随机梯度下降法.使用该方法时,首先初始化一组向量 $\boldsymbol{\beta}_0$,然后进行迭代更新,得到新的向量 $\boldsymbol{\beta}_1$,$\boldsymbol{\beta}_2$,…,使用式(17)实现迭代更新:

$$\boldsymbol{\beta}_{j+1}=\boldsymbol{\beta}_j-\gamma\times\nabla\boldsymbol{l}(\boldsymbol{\beta}_j;\boldsymbol{x},\boldsymbol{y}),\tag{17}$$

其中$\langle \boldsymbol{x},\boldsymbol{y}\rangle$是随机抽取的用户的数据,$\nabla\boldsymbol{l}(\boldsymbol{\beta}_j;\boldsymbol{x},\boldsymbol{y})$是在 $\boldsymbol{\beta}_j$ 中 L 的梯度,γ 则表示第 j 次迭代中使用的学习率.在实验中为了方便,将其设置为 0.01.

与非隐私状态下不同的是,在本地差分隐私的条件下,$\nabla\boldsymbol{l}$ 不会被用户端直接发送给聚合器,而是以隐私的方式进行收集.基于这个原因,已有的工作[12-13]提出聚合器可以在每次迭代中收集用户加噪之后的$\nabla\boldsymbol{l}$.因为每次迭代中的梯度都是数值型数据,所以可以使用针对数值型数据的本地差分隐私扰动方法来对梯度进行扰动,本文使用算法 3 对梯度进行扰动.

考虑到本地差分隐私中的隐私分配问题,如果使用随机梯度下降进行计算的话会导致加入的噪声过多,从而导致结果偏差较大,准确性较低.所以,

这里使用的是小批量梯度下降法.也就是说,在每一次迭代中,随机选取一组用户 G,G 中每一个用户都提交扰动后的梯度给服务器,服务器再将梯度更新为这组用户提交的梯度的均值,式(18)给出了梯度更新的公式:

$$\boldsymbol{\beta}_{j+1}=\boldsymbol{\beta}_j-\gamma\times\frac{1}{|G|}\sum_{i=1}^{|G|}\nabla\boldsymbol{l}_i^*, \tag{18}$$

其中 $\nabla\boldsymbol{l}_i^*$ 代表的是 G 中第 i 个用户扰动后的梯度,$|G|=\Omega\left(\frac{k(\lg(k))}{\varepsilon^2}\right)$,$k$ 代表用户数据的维度.使用小批量梯度下降进行计算,加入梯度 $\nabla\boldsymbol{l}$ 的噪声为 $O\left(\frac{\sqrt{k\lg(k)}}{\varepsilon\sqrt{|G|}}\right)$,如果使用随机梯度下降的话,加入梯度的噪声数量为 $O\left(\frac{\sqrt{k\lg(k)}}{\varepsilon}\right)$.由此可见,使用小批量梯度下降进行计算,可以得到更为准确的训练结果.在对梯度进行扰动时,需要对梯度进行预处理.如果梯度大于 1,则用户需要将其设置为 1,如果小于 -1 则需要将其设置为 -1,然后再对其进行扰动.

5 实 验

为了更好地评估论文中提出的方法的性能,论文使用了多种真实数据以及合成数据对该方法进行实验.

对于真实数据,使用了:1)从 Integrated Public Use Microdata Series[14] 抽取的 2 个公共数据集,BR 和 MX,它们分别是巴西和墨西哥的人口普查记录.BR 包含了 16 种属性,其中 6 种为数值型属性,10 种为分类型属性.MX 则包含 19 种属性,分别为 5 种数值型属性以及 14 种分类型属性.2)人类活动识别数据集 WISDM[15],这是来自 35 名参与者在安卓手机上的加速度计数数据,将其中的时间戳一列数据删除,剩下包含 3 种数值型数据以及 2 种分类型属性在内的 5 种属性.3)抽取了 ADULT 数据集[16]中属性 Age 一列.将这 4 种真实数据集的数值型属性域都规范到[-1,1].

除了真实数据集之外,论文还使用了合成数据集,分别是:1)服从高斯分布的 GAUSS 数据集,其中设置数据均值为 0,标准差为 0.25.2)服从指数分布的 EXP 数据集,将标准差设置为 0.5.3)服从均匀分布的 UNIFORM 数据集.在均值估计实验中,为了消除误差影响,每种方法重复运行了 100 次取其平均值.

5.1 参数 α 的影响

d 的取值影响了扰动后得到的均值的准确性,在前面的分类变换机制中已经介绍了 d 值的计算方法为 $d=\frac{1}{\alpha\times\varepsilon}$.$d$ 值的变化通过改变距离参数 α 实现,参数 α 增大则 d 值减小,即变换范围减小.所以,实验分析了 α 带来的影响,主要是通过改变 α 的取值来找到一个最合适的 d 值.该实验采用的数据集是 ADULT 数据集,在使用不同 d 值的条件下使用分类变换扰动对该数据集中的 Age 属性进行扰动,并使用绝对误差对扰动后的均值进行评估.式(19)给出了其计算方式:

$$AE(m)=\frac{1}{T}\sum_{i=1}^{T}|m_\circ-m_i^*|, \tag{19}$$

其中,T 代表运行的次数,$m_\circ$ 代表真实的均值,m_i^* 代表均值的估计值.

从图 1 的实验结果可以看出,在使用相同隐私预算进行扰动时,参数 α 的值越大即 d 值越小,均值估计的绝对误差越小,结果越准确.当 $\alpha=5$ 时,在不同隐私预算下均值估计的绝对误差都较小,可以得到一个较为准确的结果.所以在后续的所有实验中,将 d 设置为 $d=\frac{1}{\alpha\times\varepsilon}$.考虑到隐私预算应设置为大于零,同时应该具备一定的隐私保护效果,即该隐私保护应是有意义的,参考 Sun 等人[17]对用户隐私偏好进行的实验,论文中认为隐私保护环境中隐私预算的上界应为 10,所以论文中采用的 DCT 机制中变换范围 d 的下界为 $\frac{1}{50}$,即 A 边界值的下界为 $\frac{51}{50}$,A 的上界值随着隐私预算变化而变化.当 α 设置

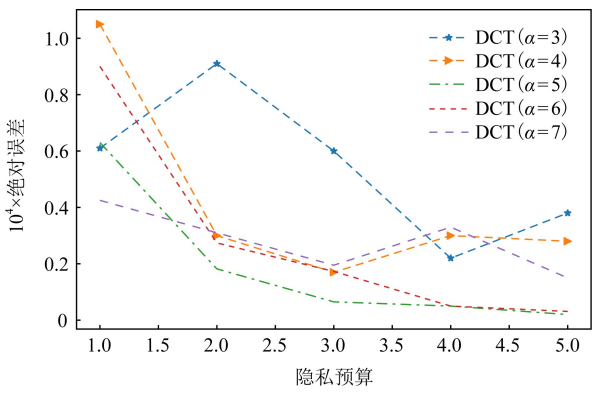


Fig. 1 The influence of different α on the mean estimation

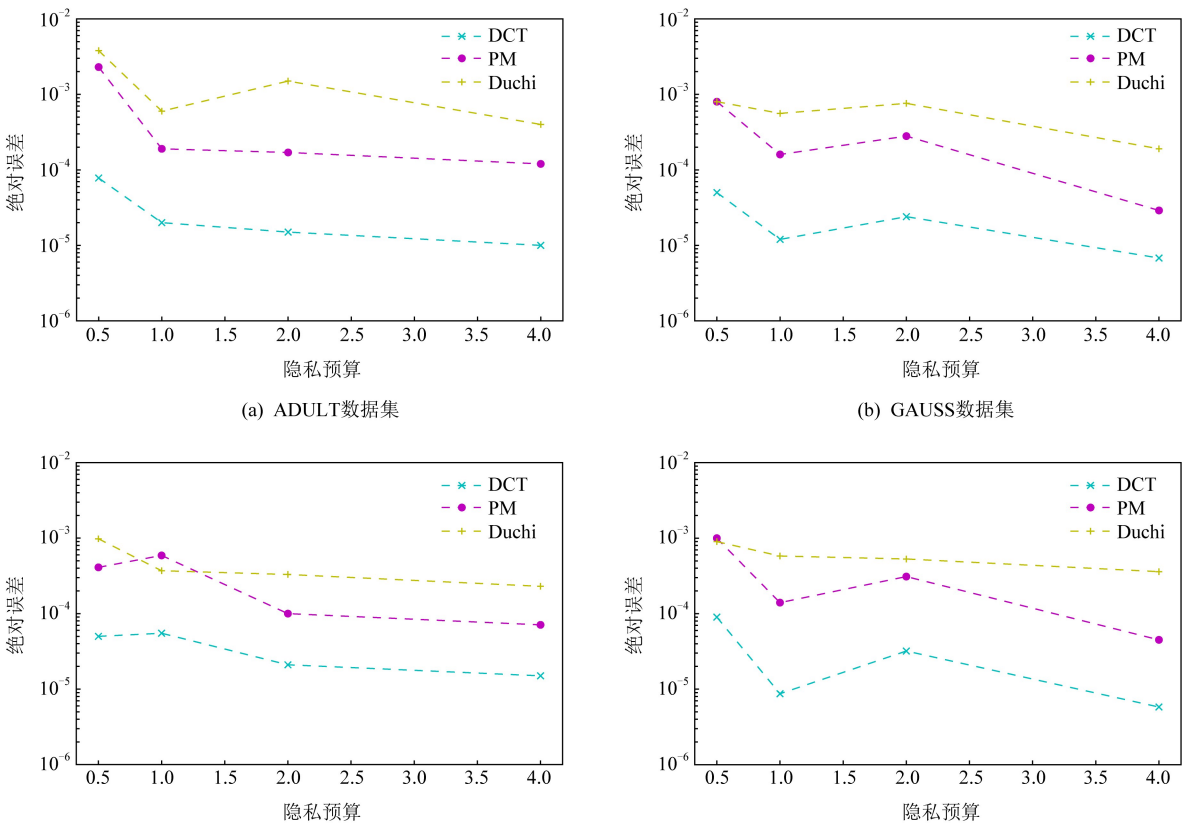
图 1 不同 α 值对均值估计的影响

为 5 时,该机制的噪声方差为$\frac{15\varepsilon v^2(5\varepsilon-1)+2}{3(5\varepsilon)^3}$,其中 v 表示用户数据.随着隐私预算变化,该方差要小于 PM 方法最坏情况下噪声方差,进一步验证了 DCT 机制的性能要优于 PM 方法.

5.2 不同数据集的 *AE* 值对比

为评估分类变换扰动机制的性能,论文计算不同机制扰动后均值估计的绝对误差进行对比.每个用户对本地数据进行扰动,服务器收集用户扰动后的数据之后计算数值属性的均值.除了使用论文中提出的机制,还使用了已有的较新的扰动方法来进行比较,包括 Wang 等人[8]提出的方法 PM(如算法 2 所示)和 Duchi 等人[7]的方法(如算法 1 所示),这也是目前连续数值型数据扰动比较有代表性的方法.为了使结果更加的真实可靠,论文在 1 个真实数据 ADULT 和 3 个合成数据上进行了实验.

由图 2 中不同类型数据集中的实验结果可看出,绝对误差随着隐私预算的增大而减少.由于 Duchi 方法的最坏情况下误差方差在隐私预算较小时与 PM 接近,所以当隐私预算小于 1 时,Duchi 和 PM 方法的结果较为接近.而论文中提出的分类变换扰动机制的绝对误差则要比这 2 种方法的误差小的多,不管隐私预算如何变换,该机制的绝对误差比其他 2 种方法要小几乎 1 个数量级.也就是说,论文中提出的方法在均值估计中的准确性得到了明显的改善.



(a) ADULT数据集 (b) GAUSS数据集

(c) EXP数据集 (d) UNIFORM数据集

Fig. 2 The mean estimate of different datasets

图 2 不同数据集的均值估计

5.3 数据量的影响

在统计任务分析中,数据量的大小通常会影响最终结果的准确性,这里将论文中提出的机制与现有机制的性能受数据集大小影响进行对比.为更好地对比数据量变化对算法性能影响,使用不同大小的高斯数据集进行均值估计,最后比较其绝对误差的值.从图 3 的实验结果可以看出,绝对误差随着数据集的增大呈下降趋势,也就是说数据量越大结果往往越准确.PM 方法的绝对误差和 Duchi 方法的比较接近,而分类变换扰动机制的误差始终要比 PM 方法以及 Duchi 方法的要更小.在不同的数据集大小中,分类变换扰动机制均体现出更好的性能,这主要是因为该机制使用了数据变换的方法,使得扰动满足本地差分隐私的同时数据能获得更高的准确性.

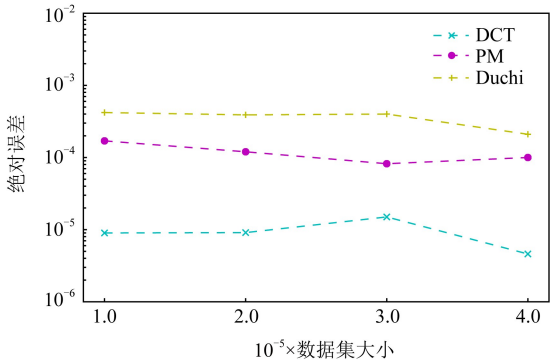


Fig. 3 The impact of dataset size

图 3 数据集大小的影响

5.4 经验风险最小化

在经验风险最小化实验中,采用小批量梯度下降算法完成线性回归的学习任务,将用户无放回的进行分组,同时为降低迭代过程中的噪声,将训练轮数设置为训练集长度除以每组人数的向下取整的值,使得用户最多参与 1 次训练.对于数据集 BR 和 MX,将"totalincome"数值属性作为因变量,其他所有属性作为自变量.对于 WISDM 数据集,将最后 1 个数值属性作为因变量,其他所有属性作为自变量.对于数据集中的分类属性,其处理方式与文献[8]中的方法一样.将每个具有 k 种值的分类型属性 A_j 转换成 $k-1$ 元属性,每一个属性的值域为 $\{0,1\}$,使得:1)A_j 中的值如为第 i 个值($i<k$)则第 i 元属性会被设置为 1,其余的 $k-2$ 个属性会被设置为 0. 2)A_j 中的值如为第 k 个值则其转换的属性所有的值都为 0.转换之后,BR 的维度为 90,MX 的维度为 94,WISDM 的维度为 43.

论文中使用的是小批量梯度下降算法,每一次迭代中抽取 1 组用户,该组中的用户对梯度进行扰动.用户将扰动后的梯度发送给服务器,服务器根据接收到的用户的梯度进行梯度更新后返回给用户.该实验包含了 3 种方法:DCT,PM,Duchi.对于所有的方法,都将正则化因子设置为 $\lambda=10^{-4}$.对于每一个数据集,使用 5 折交叉验证 5 次来评估每种方法的性能.使用均方误差(mean square error, MSE)比较使用不同扰动机制构建的小批量梯度下降算法的优劣.

图 4 描述了在不同的隐私预算下,每一种机制在不同数据下的线性回归模型的均方误差.可从实验结果看出,PM 方法和 Duchi 方法构建的满足本地差分隐私的小批量梯度下降模型的训练效果更为接近,论文中提出的分类变换扰动机制计算出的均方误差要小于这 2 种机制,获得的模型准确度更高,

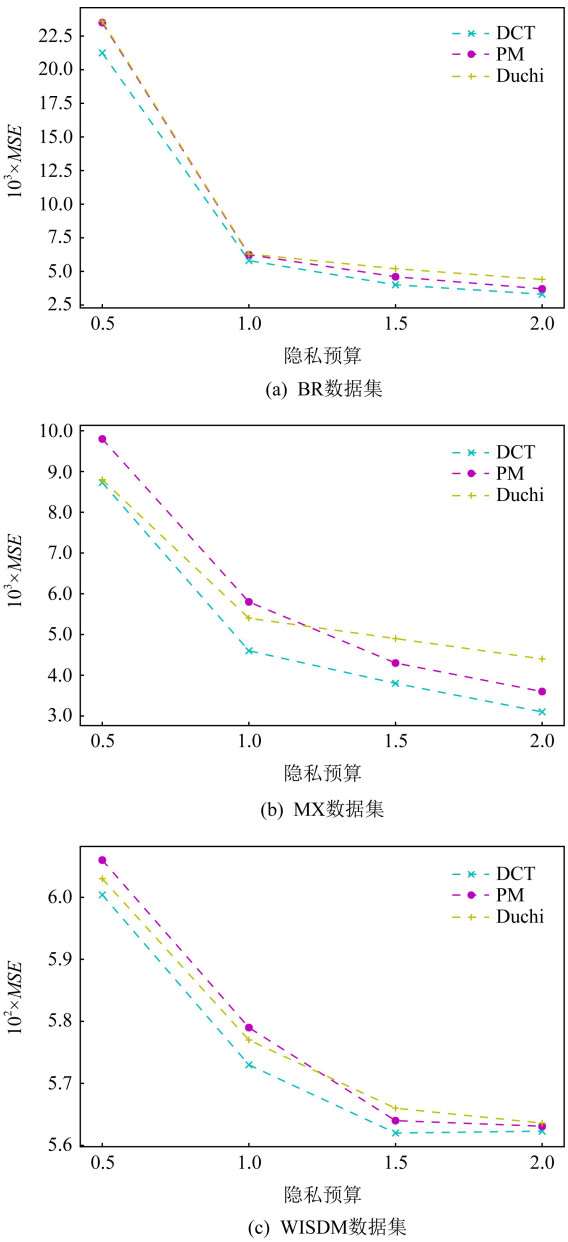


Fig. 4 Linear regression using different perturbation mechanisms

图 4 使用不同扰动机制的线性回归

性能要更优.总的来说,所有实验结果表明,不管是在均值估计中还是在经验风险最小化的任务中,分类扰动机制的性能都要优于现有的本地差分隐私的解决方法,其在简单和复杂的数据分析任务中均能获得较高的准确性.

6 结 论

为了防止用户隐私泄露,论文提出了满足本地差分隐私的分类变换扰动机制.该机制将数值型数据的扰动与分类型数据的扰动进行结合,提高了均值

估计的准确性.同时,将该机制用于梯度下降中的每次迭代的梯度扰动,保护了训练过程中用户隐私的同时得到了一个较为准确的模型.而且,本文也从本地差分隐私定义的角度,理论证明了提出的方法满足 ε-本地差分隐私.最后通过多组真实数据集以及合成数据集验证了分类变换扰动机制的性能,证明了其在相同条件下要优于现有的同类方法.下一步工作将研究如何在更为复杂的数据分析中实现隐私保护并提高准确性.

作者贡献声明:朱素霞对研究思路提供指导意见,协助设计论文框架,并对论文初稿、修改稿等提供审阅意见;王蕾提出研究思路,设计研究方案,进行实验和数据分析,并撰写论文;孙广路对研究思路提供指导意见,并对论文初稿、修改稿等提供审阅意见.

参 考 文 献

[1] Duchi J C, Jordan M I, Wainwright M J. Local privacy and statistical minimax rates [C] //Proc of the 54th IEEE Annual Symp on Foundations of Computer Science. Piscataway, NJ: IEEE, 2013: 429-438

[2] Zhu Tianqing, He Muqing, Zou Deqing. Big data privacy protection based on differential privacy [J]. Journal of Information Security Research, 2015, 1(3): 224-229 (in Chinese)
(朱天清, 何木青, 邹德清. 基于差分隐私的大数据隐私保护 [J]. 信息安全研究, 2015,1(3): 224-229)

[3] Erlingsson U, Pihur V, Korolova A. RAPPOR: Randomized aggregatable privacy-preserving ordinal response [C] //Proc of 2014 ACM SIGSAC Conf on Computer and Communications Security. New York: ACM, 2014: 1054-1067

[4] Ding Bolin, Kulkarni J, Yekhanin S. Collecting telemetry data privately [C] //Proc of the 31st Int Conf on Neural Information Processing Systems. Cambridge: MIT Press, 2017: 3571-3580

[5] Nguyên T T, Xiao Xiaokui, Yang Yin, et al. Collecting and analyzing data from smart device users with local differential privacy [OL]. (2016-06-05)[2020-03-24]. https://arxiv.org/abs/1606.05053

[6] Ye Qingqing, Hu Haibo, Meng Xiaofeng, et al. PrivKV: Key-value data collection with local differential privacy [C] //Proc of IEEE Symp on Security and Privacy (S&P). Piscataway, NJ: IEEE, 2019: 317-331

[7] Duchi J C, Jordan M I, Wainwright M J. Minimax optimal procedures for locally private estimation [J]. Journal of the American Statistical Association, 2018, 113(521): 182-201

[8] Wang Ning, Xiao Xiaokui, Yang Yin, et al. Collecting and analyzing multidimensional data with local differential privacy [C] //Proc of IEEE Int Conf on Data Engineering (ICDE). Piscataway, NJ: IEEE, 2019: 638-649

[9] Dwork C, Roth A. The algorithmic foundations of differential privacy [J]. Foundations and Trends in Theoretical Computer Science, 2014, 9(3/4): 211-407

[10] Stanley L W. Randomized response: A survey technique for eliminating evasive answer bias [J]. Journal of the American Statistical Association, 1965, 60(309): 63-69

[11] Xia Chang, Hua Jingyu, Tong Wei, et al. Distributed *K*-means clustering guaranteeing local differential privacy [J]. Journal of Computers & Security, 2020, 90(3): No.101699

[12] Duchi J C, Jordan M I, Wainwright M J. Privacy aware learning [J]. Journal of the Association for Computing Machinery, 2014, 61(6): 1-57

[13] HammJ, Champion A C, Chen Guoxing, et al. Crowd-ML: A privacy-preserving learning framework for a crowd of smart devices [C] //Proc of IEEE ICDCS. Piscataway, NJ: IEEE, 2015: 11-20

[14] IPUMS. Integrated public use microdata series [OL]. [2020-04-28]. https://www.ipums.org

[15] Kwapisz J R, Weiss G M, Moore S A. Activity recognition using cell phone accelerometers [J]. ACM SIGKDD Explorations Newsletter, 2011, 12(2): 74-82

[16] Kohavi R, Becker B. Uci repository of machine learning databases: Adult data set [OL]. [2020-03-26]. https://archive.ics.uci.edu/ml/datasets/Adult

[17] Sun Lin, Ye Xiaojun, Zhao Jun, et al. BiSample: Bidirectional sampling for handling missing data with local differential privacy [C] //Proc of the 25th Int Conf on Database Systems for Advanced Applications. Berlin: Springer, 2020: 88-104

Zhu Suxia, born in 1978. PhD, associate professor and PhD supervisor. Member of CCF. Her main research interests include privacy and security, IoT and parallel computing.
朱素霞,1978 年生.博士,副教授,博士生导师,CCF 会员.主要研究方向为隐私与安全、物联网和并行计算.

Wang Lei, born in 1997. Master candidate. Her main research interests include differential privacy and IoT.
王 蕾,1997 年生.硕士研究生.主要研究方向为差分隐私和物联网.

Sun Guanglu, born in 1979. PhD, professor and PhD supervisor. Distinguished member of CCF. His main research interests include artificial intelligence, network and information security, intelligent information processing.
孙广路,1979 年生.博士,教授,博士生导师,CCF 杰出会员.主要研究方向为人工智能、网络与信息安全、智能信息处理.