MCDP:基于神经网络的多集群分布式差分隐私数据发布方法

陈 思[1,2],付安民[1],柯海峰[1],苏 铓[1],孙怀江[1]
(1. 南京理工大学计算机科学与工程学院,江苏南京 210094; 2. 南京理工大学信息化建设与管理处,江苏南京 210094)

摘 要: 大数据应用能够为人们的生活和工作方式提供便捷,但包含消费记录、社交关系、地理位置等个人隐私信息的数据在发布过程中可能被服务提供商收集,用户隐私面临巨大威胁. 本文首次提出了一个基于神经网络的多集群分布式差分隐私数据发布方法,能够显著缓解单服务器的数据处理压力. 同时,利用神经网络算法进行隐私参数预测明显提高了预测精度和预测效率,并且集群之间不同的隐私参数也保证了方案的灵活性. 此外,由于中心服务器存储的是经过差分隐私处理后的统计数据,即使中心服务器由于遭受攻击导致存储的数据泄露,也能确保用户数据隐私. 实验对比分析表明,我们的方法在隐私处理效率、隐私保护强度、预测精度和预测效率等方面都有明显优势.

关键词: 数据发布; 差分隐私; 服务协同; 神经网络; 多集群; 分布式
中图分类号: TP18 **文献标识码:** A **文章编号:** 0372-2112 (2020)12-2297-07
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2020.12.002

MCDP:Multi-Cluster Differential Privacy Data Publishing Method Based on Neural Network

CHEN Si[1,2],FU An-min[1],KE Hai-feng[1],SU Mang[1],SUN Huai-jiang[1]
(1. *School of Computer Science and Engineering,Nanjing University of Science and Technology,Nanjing,Jiangsu* 210094,*China*;
2. *Division of Information Construction and Management,Nanjing University of Science and Technology,Nanjing,Jiangsu* 210094,*China*)

Abstract: Big data applications provide convenience for people's life and work style,but in the process of data publishing,personal privacy information,such as consumption records,social relations,and so on,are collected by service providers all the time,and users' privacy is threatened greatly. Aiming at the significant relief of data processing pressure on single server,we propose a multi-cluster distributed differential privacy data publishing method based on neural network (MCDP),which effectively improves the prediction accuracy and efficiency,and different privacy parameters between clusters guarantee the flexibility of the protocol. Especially,because the central server stores statistical data after differential privacy processing,it does not collect individual privacy data,even if the central server is attacked,the user data will not be leaked. Experiments show that MCDP has obvious advantages in privacy processing efficiency,privacy protection intensity,prediction accuracy and prediction efficiency.

Key words: data publishing;differential privacy;service collaboration;neural network;multi-cluster;distribution

1 引言

大数据应用改变了人们的日常生活与工作方式[1,2],有助于提升服务的准确性和及时性,并提供友好的信息推荐,使用户得到快捷的体验[3]. 但是,大数据为人们生活带来便利的同时,也引起用户隐私信息被恶意利用或者泄漏的担忧[4].

为此,国内外学者提出了许多隐私保护数据发布的方法,包括基于限制发布技术和基于数据失真技术两类. 前者主要通过有选择地发布原始数据,即不发布用户敏感数据,或只发布敏感度不高的数据,以达到保护用户隐私目的,大多集中于"匿名策略",包括 k-ano-

收稿日期:2019-09-23;修回日期:2019-12-15;责任编辑:覃怀银
基金项目:国家自然科学基金(No. 61572255,No. 61702266);江苏省"六大人才高峰"高层次人才基金资助项目(No. XYDXXJS-032);赛尔网络下一代互联网技术创新项目(No. NGII20190804,No. NGII20150117)

nymity、l-diversity、t-closeness 等[5,6]. 后者典型代表是差分隐私技术,提供可量化评估的隐私保护方式[7~9]. 然而,现有差分隐私技术大多面向的是单服务器的小型数据集,很难直接用于 TB、PB 等数量级的数据隐私保护[10,11]. 例如,在进行全国癌症统计分析时,传统的差分隐私方法将所有病人数据都集中在中心服务器,采用统一隐私参数进行处理并发布. 但这需要基于两个难以满足的前提条件:一是中心服务器可信、计算能力高且具有强抵御攻击能力,二是全国各区域(如江苏、广东、新疆等)有相同隐私保护标准和制度,能提供完整数据. 目前基于分布式场景的隐私数据保护方法侧重于数据处理前期聚类算法的差分隐私保护[12,13],并没有考虑后期各个集群里隐私预算的差异性,导致发布的数据可用性低的问题.

因此,去中心化的多集群分布式隐私保护成为大数据环境下的必然选择,但是依然面临不少难点. 首先现有隐私保护中噪声与数据之间的关联性不能被忽视,噪声无法脱离数据集的属性而独立添加,注重于数据集的整体统计属性,而分布式的处理则会切割这种统计属性. 此外,如何有效整合分布式差分隐私的计算结果也是有待解决的难题.

针对上述问题,本文首次提出了一种基于神经网络预测的多集群分布式差分隐私数据发布方法 MCDP(Multi-Cluster distributed Differential Privacy data publishing method based on neural network),能针对分布式计算的特点来进行数据发布,同时通过合作寻找标准隐私参数,能够提高数据可用性并满足隐私保护要求. 本文的主要贡献可概括如下:

(1)首次提出了一种基于神经网络预测的多集群分布式差分隐私数据发布方法,通过同步建立分布式的差分隐私处理模型,能够显著缓解单服务器的数据处理压力,并且利用神经网络算法进行隐私参数预测可以明显提高预测精度和预测效率.

(2)MCDP 允许各个集群合作寻找动态的隐私参数,从而能够保证方案的灵活性和可靠性. 并且由于中心服务器存储的是经过差分隐私处理后的统计数据,即使中心服务器由于遭受攻击导致存储的数据泄露,也能确保用户数据隐私.

(3)安全性分析表明 MCDP 不仅在各个集群的数据处理满足差分隐私,同时整个方案的查询结果也满足差分隐私. 此外,实验对比分析表明 MCDP 在隐私处理效率、隐私保护强度、预测精度和预测效率等方面都有明显优势.

2 MCDP 方案

本节提出一个包括用户协作,实时预测等功能于一体的差分隐私数据发布方法 MCDP,表 1 给出了 MCDP 的符号定义及描述.

表 1 MCDP 的符号定义及描述

符号	描述
QS	查询数据集
ε_set	隐私参数集合
t	真实数值
c	噪声数值
ln	Laplace 噪声
T	原始数据表
$[\alpha,\beta]$	真实值范围
$[\gamma,\eta]$	噪声范围
Rand(i)	$[1,i]$之间的随机数
TG	集群预测隐私参数数据表
DG	集群最优隐私参数数据表

2.1 MCDP 数据发布模型

图 1 给出了 MCDP 发布模型,该模型中存在一台可以提供对外查询服务的中心服务器和多台用于处理数据的分布式集群服务器. 本文方案实施主要包括以下三个阶段.

(1)单机处理阶段:各个集群完成查询结果隐私处理和隐私参数预测子任务,包括两个步骤:一是数据预处理,各个集群进行数据的收集和处理(包括数据存储,格式清洗等),然后针对给定的查询集以各自的隐私参数对数据进行 Laplace 机制处理;二是隐私参数预测,各个集群收到指令要进行下一次数据统计时,结合自身特性使用历史隐私参数数据集进行神经网络方法预测得到 $\varepsilon_{预测}$,通过查询次数迭代,动态调整的 $\varepsilon_{预测}$ 不断接近 $\varepsilon_{最优}$.

(2)通信阶段:集群和中心服务器进行交互,主要内容包括查询结果的传输、整体隐私参数参考值 ε 的反馈等. 具体完成两项工作:一是集群将数据集查询结果安全传输给中心服务器;二是中心服务器计算各子集群的隐私参数和整体隐私参数的差值,并反馈隐私参数参考值 ε 给集群进行调整. 本阶段可以用 SSL/TLS 协议实现数据安全通信.

(3)服务协同阶段:中心服务器整合集群上报的统计数据,不搜集个体隐私数据,并对外提供整体分析查询接口. 一方面,中心服务器收集各集群隐私参数,选取最大值作为本次整体的隐私参数参考值 ε 并反馈给各集群. 另一方面,集群利用指数机制计算 $\varepsilon_{预测}$ 和 ε 区间的 $\varepsilon_{最优}$,使用 $\varepsilon_{最优}$ 来做本集群的查询统计值处理,进行新一轮单机处理. 该阶段通过协作寻找最优隐私参数和参数动态化调整保证了数据精度统一,由中心服务器进行整体统计数据发布,保证统计值具有较高的可用性.

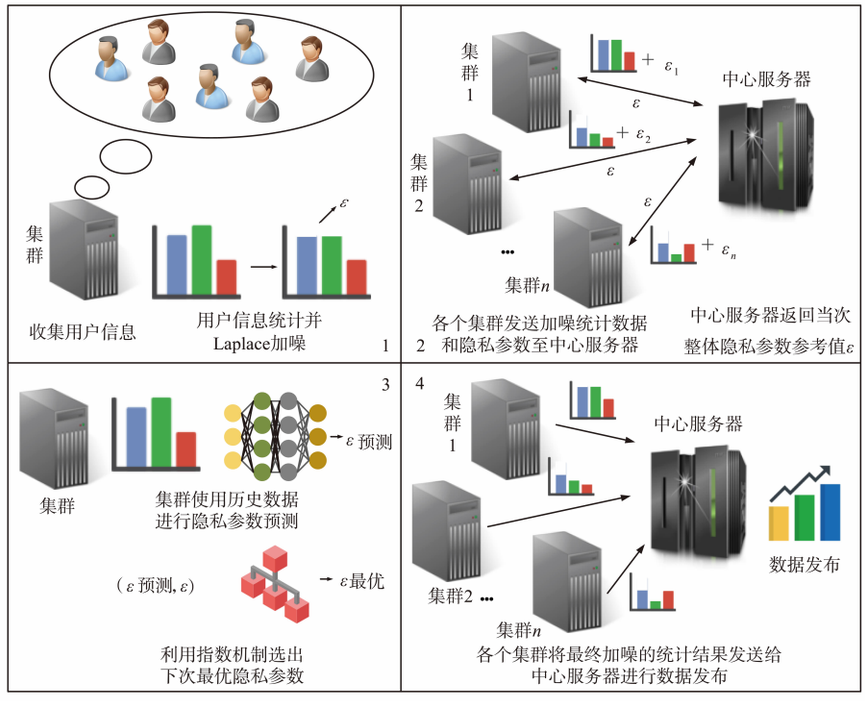

图1　MCDP数据发布模型

下面我们重点阐述单机处理和服务协同阶段两个核心环节. 由于通信阶段不涉及差分隐私,可以直接采用 SSL/TLS 协议实现,因此不再赘述.

2.2　单机处理阶段

单机处理阶段中,核心包括子集群差分隐私处理和隐私参数预测两步骤.

第一步,在各集群都存储有各自的数据,并且数据相互独立. 每个集群可以用不同的隐私参数对各自数据进行差分隐私处理. 各个集群选取不同的隐私参数是因为考虑数据集的差异性,弹性的隐私参数设置可以应对更复杂的数据场景,提供更强的保护能力. 原始各个集群的隐私参数为式(1):

$$\varepsilon_n(n=1,2,\cdots,i)=\exp\{\varepsilon_1,\varepsilon_2,\varepsilon_3,\cdots,\varepsilon_n\} \quad (1)$$

MCDP 考虑到不同集群可以取不同隐私参数的特性,结合差分隐私 Laplace 分布的特征对每一个隐私参数进行加噪,为各集群提供安全保障.

第二步,采用基于时间序列的 BP(Back Propagation)神经网络预测方法来拟合函数 f,然后预测出未来值[14]. 预测阶段可以描述为式(2):

$$X_{n+k}=f(X_n,X_{n-1},X_{n-2},\cdots,X_1) \quad (2)$$

MCDP 使用一个 5 层全连接神经网络,每层节点数为[2048,1024,512,256,100],选取 Tanh 为激活函数,Sigmoid 为传递函数,Logsig 为对数函数对模型进行训练. 在差分隐私模型的隐私参数预测中 BP 神经网络的神经元的输入能够被映射到(0,1)的空间中,可以满足差分隐私参数的预测需求. 单机处理阶段如算法 1 所示.

算法 1　单机处理算法

输入:$\alpha=0,\beta,$QS$,T,\gamma,\eta,\varepsilon_$set
输出:TG

```
1.    for all clusters:
2.      if α<0 || α>β || n<0, then
3.        return ⊥
4.      end if
5.     for all q_i in QS:
6.      get the query result tc
7.        for all r,s∈[1,n]
8.          μ = Average{t_r}
9.          Δf = max_{r,s}{|t_r − t_s|}
10.       end for
11.       α = Δf/μ, β = 2μ
12.       ln_i = pdf(x)
13.       if ln_i < η && ln_i > γ && i < n_1, then
14.       c_i = t_i, i ++
15.       BP(c_i)
16.       end if
17.       if i ≠ n_1, 转到 line 11
18.       end if
19.     end for
20.   end for
21.  return TG
```

2.3 服务协同阶段

MCDP 在服务协同阶段融入容灾机制,即使中心服务器受到攻击,用户数据也不会泄漏,中心服务器只对外提供查询接口. 服务协同分为两方面.

一方面,中心服务器收到各子集群传递来的差分隐私参数,由定理 3 可知,如果一个差分隐私保护算法序列中所有子算法处理的数据集彼此不相交,那么整体提供的隐私保护水平取决于算法序列中的保护水平最差者,即隐私预算最大者. 因此中心服务器选择最大的隐私参数值为当次的整体隐私参数参考值 ε,并反馈给子集群. 任意子集群的随机处理过程和输入数据集是相互独立的,当集群接收到 ε 后,采用指数机制对 TG 进行处理.

针对数据集 D,指数机制的关键是设计一个效用函数 $U(D,p)(p\in\tau)$,其中 p 表示从输出域 $\tau=[\varepsilon_{预测},\varepsilon]$ 所选择的输出项,U 可以计算出在 τ 区域中每一个值被选中为 $\varepsilon_{最优}$ 的可能性分值. MCDP 在服务协同阶段满足指数机制,可得:

$$\mathrm{MCDP}(D,U)=\left\{p:\left|\Pr[p\in\tau]\propto\left(\frac{\varepsilon U(D,p)}{2\Delta U}\right)\right.\right\} \tag{3}$$

由式(3)可知,效用函数分值较高的查询结果比分值较低的查询结果拥有更大的概率被选择发布,集群最优隐私参数也分布在此. 通过服务协同算法计算出集群的最优隐私参数,如算法 2.

算法 2 服务协同算法

输入:TG
输出:DG
1. for all clusters:
2. if $D=\varnothing$,then
3. return $\perp$
4. end if
5. for each TG in $D[\varepsilon,\varepsilon_{预测}]$
6. Calculate the possible region partition cases:$(p_1,p_2,\cdots,p_n)$
7. end for
8. for i 1 to n by 1 do
9. Compute $U(D,p_i)$
10. end for
11. select the max U
12. determine the partition $P_{\max}$
13. dataset ε partition
14. $\varepsilon_{最优}$ = the centre of $P_{\max}$
15. end for
16. return DG

另一方面,通过差分隐私的指数机制进行处理迭代后,得到集群数据集最优的隐私参数,基于此进行集群查询数据集的 Laplace 机制加噪,通过通信阶段传递,将经过隐私处理的数据给中心服务器,中心服务器为外界提供统计结果的查询接口.

3 安全性分析

MCDP 中首先证明每一个阶段满足 ε-差分隐私,然后根据组合特性,进一步证明经过 MCDP 算法处理的数据发布结果满足 ε-差分隐私.

定理 1 MCDP 在单机处理阶段满足 ε-差分隐私.

证明 单机处理阶段使用 Laplace 机制,对于数值型函数,我们计算输出 x 的概率:

$$\Pr\left\{\mathrm{Lap}\left(\frac{\Delta f}{\varepsilon}\right)=x\right\}=\frac{1}{2\frac{\Delta f}{\varepsilon}}\mathrm{e}^{-|x|/\frac{\Delta f}{\varepsilon}} \tag{4}$$

同样地,计算输出为 $x+d$ 的概率,噪声服从 $\mathrm{Lap}\left(\frac{\Delta f}{\varepsilon}\right)$ 的 Laplace 分布,令 $\beta=\frac{\Delta f}{\varepsilon}$,有:

$$\frac{\Pr(\mathrm{Lap}(\beta))=x}{\Pr(\mathrm{Lap}(\beta)=x+d)}\leqslant\exp\left(\frac{d}{\beta}\right)\leqslant\exp\left(\frac{\Delta f}{\varepsilon}\right)=\exp(\varepsilon) \tag{5}$$

因此,MCDP 在单机处理阶段满足 ε-差分隐私.

定理 2 MCDP 在服务协同阶段满足 ε-差分隐私.

证明 服务协同阶段使用指数机制,存在一对兄弟数据集 D 和 D' 都服从指数机制,对任意的查询函数 q,定义 r 为任意合法的输出,R 为输出值的范围,Δq 为查询函数中最大可能差异值,则有

$$\frac{\exp\left(\frac{\varepsilon q(D,r)}{2\Delta q}\right)}{\exp\left(\frac{\varepsilon q(D',r)}{2\Delta q}\right)}=\exp\left(\frac{\varepsilon(q(D,r)-q(D',r))}{2\Delta q}\right)\leqslant\exp\left(\frac{\varepsilon}{2}\right) \tag{6}$$

分别计算两个用指数机制处理的数据集的结果等于 r 的概率,并证明其结果满足差分隐私.

$$\begin{aligned}&\frac{\Pr(\mathrm{MCDP}_q^{\varepsilon}(D)=r)}{\Pr(\mathrm{MCDP}_q^{\varepsilon}(D')=r)}\\&=\frac{\exp\left(\frac{\varepsilon q(D,r)}{2\Delta q}\right)}{\int_{r'\in R}\exp\left(\frac{\varepsilon q(D,r')}{2\Delta q}\right)}\Bigg/\frac{\exp\left(\frac{\varepsilon q(D',r)}{2\Delta q}\right)}{\int_{r'\in R}\exp\left(\frac{\varepsilon q(D',r')}{2\Delta q}\right)}\\&\leqslant\exp\left(\frac{\varepsilon}{2}\right)\left(\frac{\exp\left(\frac{\varepsilon}{2}\right)\int_{r'\in R}\exp\left(\frac{\varepsilon q(D,r')}{2\Delta q}\right)}{\int_{r'\in R}\exp\left(\frac{\varepsilon q(D,r')}{2\Delta q}\right)}\right)\\&=\exp(\varepsilon)\end{aligned} \tag{7}$$

因此,MCDP 在服务协同阶段满足 ε-差分隐私.

定理 3 假设 $K(D_1),K(D_2),\cdots,K(D_n)$ 分别表示输入数据集为 $D_1,D_2,\cdots,D_n$ 的一系列满足 ε_i-差分隐私的随机算法,并且任意两个算法的随机过程和输入数

据集是相互独立的,则这些算法组合起来的算法 MCDP 满足 $\max\{\varepsilon_i\}$-差分隐私.

证明　由于算法 $K(D_n)$ 均满足 ε_i-差分隐私,假设 $\varepsilon_1=\varepsilon_2=\cdots=\varepsilon_n=\varepsilon$,算法 $K(D)$ 表示所有算法 $K(D_i)$ 组合起来的算法,输出记为 $\{s_1,s_2,\cdots,s_n\}$. 由于任意两个 $K(D_i)$ 之间的随机过程和输入的数据集相互独立,对任意 $s\in\{s_1,s_2,\cdots,s_n\}\in\mathrm{Range}(s)$,有:

$$\begin{aligned}\Pr[K(D)\in s]&=\prod_{i=1}^{n}\Pr[K(D_i)\in s_i]\\&\leqslant\prod_{i=1}^{n}\mathrm{e}^{\varepsilon\times|D_i\oplus D_i'|}\cdot\Pr[K(D_i')=s_i]\\&=\mathrm{e}^{\varepsilon_i\times\sum_{i=1}^{n}|D_i\oplus D_i'|}\cdot\Pr[K(D_i')=s_i]\end{aligned}\tag{8}$$

作为兄弟数据集的 D 和 D'应当满足$|D\oplus D'|=1$. 由于对任意的 $i\neq j$,有 $D_i\cap D_j=\varnothing\wedge D_i'\cap D_j'=\varnothing$,可知:

$$\sum_{i=1}^{n}|D_i\oplus D_i'|=|D\oplus D'|=1\tag{9}$$

则组合算法 $K(D)$ 满足 ε-差分隐私.

假设 ε_i 不完全相等,记 $Q=\langle D,D'\rangle$ 表示兄弟数据集 D 和 D'组合一起的集合,可以将 Q 按照 D 和 D'在哪个子部分出现的差异将 Q 分类成 $Q_1,Q_2,\cdots,Q_n$ 共 n 个类别. 每个类别的 Q_i 定义如下:

$$Q_i=\{\langle D,D'\rangle\mid(|D\oplus D'|=1)\wedge(D_j=D_j')\}\tag{10}$$

假设组合算法 $K(D)$ 满足 ε'-差分隐私,则有:

$$\Pr[K(D)\in S]\leqslant\mathrm{e}^{\varepsilon'}\times\Pr[K(D')\in S],\varepsilon'\geqslant\varepsilon_i\tag{11}$$

因此对 ε'推论如下:

$$\begin{aligned}\varepsilon'&=\min\{\varepsilon'\mid\wedge(\varepsilon'\geqslant\varepsilon_i)\}\\&=\min\{\varepsilon'\mid\varepsilon'\geqslant\max(\varepsilon_i)\}\\&=\max(\varepsilon_i)\end{aligned}\tag{12}$$

可得,$K(D)$ 满足 $\max\{\varepsilon_i\}$-差分隐私,进一步得证 MCDP 满足 ε-差分隐私.

4　性能分析

本文采用加州大学机器学习库里面的“Adult”数据集进行实验分析验证[15],重点从隐私处理效率、隐私保护强度、预测精度和预测效率进行分析.

4.1　隐私处理效率

在隐私处理效率实验中,将 MCDP 与单个服务器和分布式差分隐私方法 DP-MCDBSCAN[13] 进行对比. 在整体隐私参数为 0.2,集群数量分别取 5、10 和 20 的情况下,与单个中心服务器隐私计算对比效果如图 2 所示. 实验结果表明,随着数据集的增大,隐私处理的耗时也随之增加,其中单服务器的处理耗时最长. 在使用多集群处理的情况下,随着集群数量增加,隐私处理的耗时也越来越小.

在集群数量取 10 相同的情况下,我们通过实验进一步对比分析 MCDP 与 DP-MCDBSCAN 的隐私处理效率. 当整体隐私参数分别取 0.1、0.2、0.5 和 0.8 的时候,图 3 表明两种方案的整体变化趋势,随着隐私参数的增加而隐私处理的计算时间依次减少. 相比于 DP-MCDBSCAN,MCDP 运行时间较低,在运行效率方面具有明显优势,可见 MCDP 在保护用户数据隐私的同时能兼顾较好性能表现.

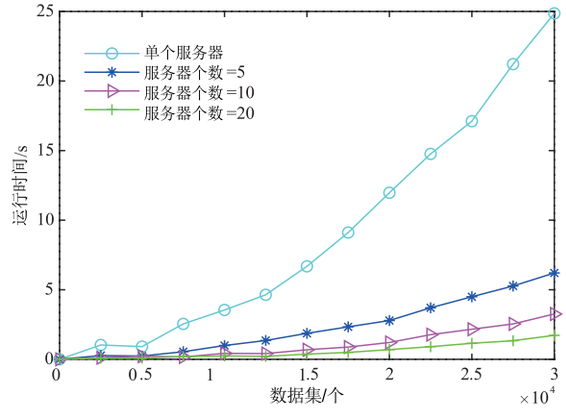

图2　隐私处理效率（与单服务器对比）

4.2　隐私保护强度

差分隐私参数 ε 用于衡量隐私保护的强度,越小的 ε 代表较高的隐私保护强度,此时数据的可用性也会降低. 实验结果如图 4 所示,随着查询次数的增加,计算五个集群的 ε,可以看出经过学习后五个集群的 ε 达到平衡并且趋近. 这是因为 MCDP 允许各个集群合作寻找动态的隐私参数,保证方案的灵活性和可靠性,在满足差分隐私同时,对整体查询结果进行求精处理,提高发布数据的可用性.

4.3　预测精度

在差分隐私预算的预测方面,我们分别将 MCDP 采用的 BP 神经网络算法与目前常用的预测方法线性回归模型和长短期记忆模型 LSTM[16] 进行对比. 经过 20 到 700 个周期内的预测实验,预测精度的实验内容为 $\varepsilon_{\text{预测}}$ 和 $\varepsilon_{\text{最优}}$ 的差异值. 从图 5 可以看出,LSTM 模型由于复杂的训练模型,预测精度要略微优于 BP 神经网络模型,而线性回归模型远远不如 BP 神经网络模型和 LSTM 模型.

4.4　预测效率

在预测效率方面,随着计算周期的增加,三个预测模型完成预测任务所需的计算时间也增加,如图 6 所示. 结果表明,线性回归模型具有最优的预测效率,但是该模型在预测精度上太低. 而对比 BP 神经网络和 LSTM 的表现来看,BP 神经网络的预测时间明显低于 LSTM 模型,预测效率较高,同时还可以达到较接近的预测精度,因此我们选择综合性能较好的 BP 神经网络进行隐私参数预测.

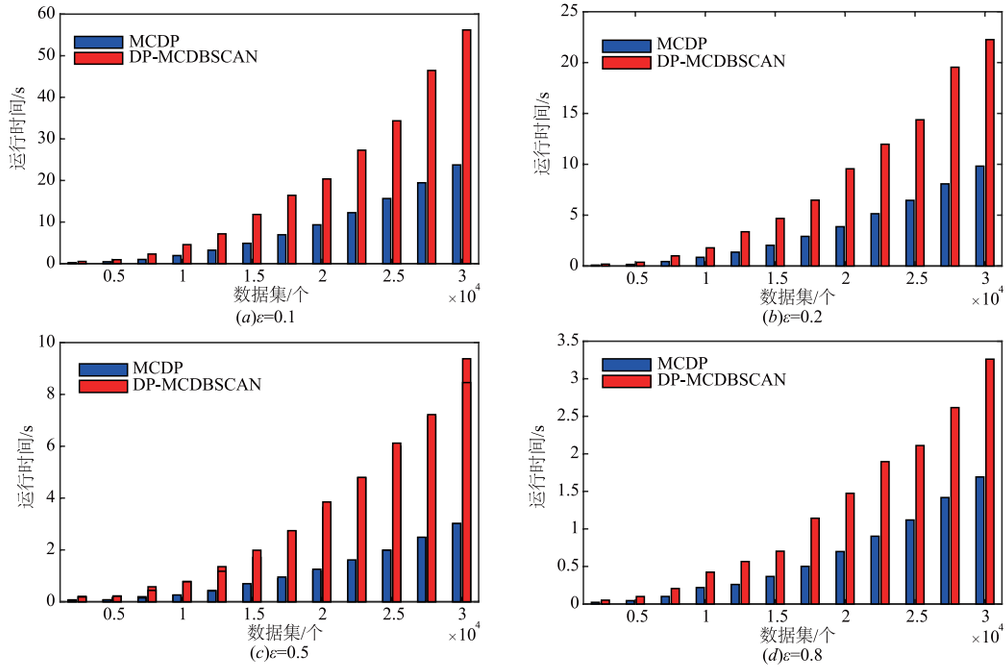

图3 隐私处理效率（与DP-MCDBSCAN对比）

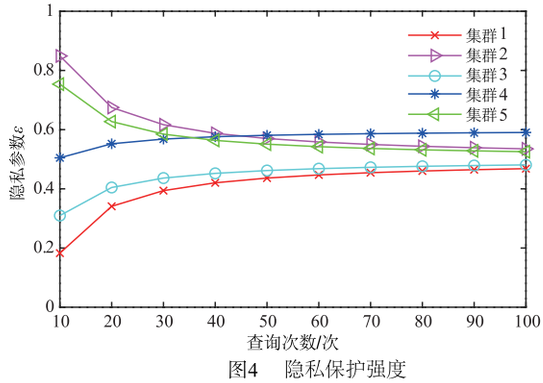

图4 隐私保护强度

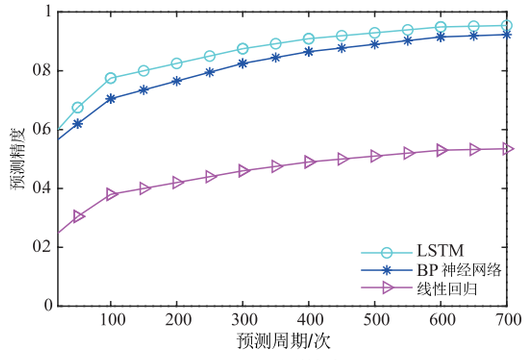

图5 预测精度

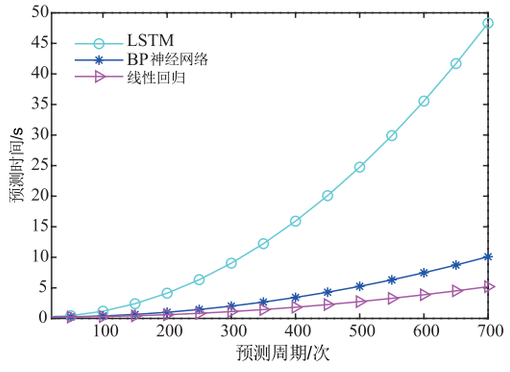

图6 预测效率

5 结束语

本文提出了一种基于神经网络预测的多集群分布式差分隐私数据发布方案 MCDP，通过建立分布式模型显著缓解单服务器的数据处理压力，可以有效地处理海量数据的隐私保护问题. 同时，MCDP 存储的是经过差分隐私处理后的统计数据，不搜集个体的隐私数据，防止用户数据泄漏. 此外，集群之间不同隐私参数的选择保证方案的灵活性，借助神经网络进行隐私参数预测可以实现更准确的预测效果. 实验结果表明 MCDP 在隐私处理效率、隐私保护强度、预测精度和预测效率等方面都具有明显优势.

参考文献

[1] YU S，ZHOU W，GUO S，et al. A feasible IP traceback framework through dynamic deterministic packet marking[J]. IEEE Transactions on Computers，2016，65(5)：1418－1427.

[2] 杨高明，朱海明，方贤进，等. 局部差分隐私约束的关联属性不变后随机响应扰动[J]. 电子学报，2019，47(5)：105－111.

Yang Gao-ming, Zhu Hai-ming, Fang Xie-jin, et al. Invariant post-random response perturbation for correlated attributes under local differential privacy constraint[J]. Acta Electronica Sinica, 2019, 47(5): 105 – 111. (in Chinese)

[3] 鲜征征,李启良,黄晓宇,等. 融合显/隐式信任协同过滤算法的差分隐私保护[J]. 电子学报,2018,46(12):236 – 245.
Xian Zheng-zheng, Li Qi-liang, Huang Xiao-yu, et al. Differential privacy protection for collaborative filtering algorithms with explicit and implicit trust[J]. Acta Electronica Sinica, 2018, 46(12): 236 – 245. (in Chinese)

[4] FU A M, YU S, ZHANG Y Q, et al. NPP: a new privacy-aware public auditing scheme for cloud data sharing with group users[J]. IEEE Transactions on Big Data, DOI: 10.1109/TBDATA. 2017. 2701347, 2017.

[5] OGANIAN A, DOMINGO-FERRY J. Local synthesis for disclosure limitation that satisfies probabilistic k-anonymity criterion[J]. Transactions on Data Privacy, 2017, 10(1): 61 – 81.

[6] SORIA-COMAS J, DOMINGO-FERRER J, SANCHEZ D, et al. t-closeness through microaggregation: strict privacy with enhanced utility preservation[J]. IEEE Transactions on Knowledge and Data Engineering, 2015, 27(11): 3098 – 3110.

[7] DWORK C. Differential privacy[J]. Lecture Notes in Computer Science, 2006, 26(2): 1 – 12.

[8] YIN C, XI J, SUN R, et al. Location privacy protection based on differential privacy strategy for big data in industrial Internet of Things[J]. IEEE Transactions on Industrial Informatics, 2017, 14(8): 3628 – 3636.

[9] QU Y Y, YU S, GAO L X, et al. Big data set privacy preserving through sensitive attribute-based grouping[A]. International Conference on Communications (ICC)[C]. Paris, France: IEEE, 2017. 1 – 6.

[10] DRAKONAKIS K, ILIA P, IOANNIDIS S, et al. Please forget where I was last summer: the privacy risks of public location (meta) data[A]. 26th Annual Network and Distributed System Security Symposium (NDSS)[C]. San Diego, USA: ISOC, 2019. 1 – 15.

[11] YE Q Q, HU H B, MENG X F, et al. PrivKV: key-value data collection with local differential privacy[A]. Symposium on Security and Privacy (S&P)[C]. San Francisco, USA: IEEE, 2019. 1 – 15.

[12] DUAN Y, YOUDAO N E, CANNY J, et al. P4P: practical large-scale privacy-preserving distributed computation robust against malicious users[A]. 19th USENIX Security Symposium (USENIX)[C]. Washington, USA: ACM, 2010. 1 – 15.

[13] NI L, LI C, WANF X, et al. DP-MCDBSCAN: differential privacy preserving multi-core DBSCAN clustering for network user data[J]. IEEE Access, 2018, 6: 21053 – 21063.

[14] LV C, XING Y, ZHAN J, et al. Levenberg-marquardt backpropagation training of multilayer neural networks for state estimation of a safety critical cyber-physical system[J]. IEEE Transactions on Industrial Informatics, 2017, 14(8): 3436 – 3446.

[15] LICHMAN M. UCI Machine Learning Repository [EB/OL]: http://archive. ics. uci. edu/ml, 2013.

[16] GREFF K, SRIVASTAVA R K, KOUTNIK J, et al. LSTM: a search space odyssey[J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, 28(10): 2222 – 2232.

作者简介

陈　思　女,1987 年生于湖北襄阳. 现为南京理工大学计算机科学与工程学院博士研究生. 主要研究方向为大数据隐私保护.
E-mail: chensi@ njust. edu. cn

付安民(通信作者)　男,1981 年生于湖北咸宁. 现为南京理工大学计算机科学与工程学院副教授、博士生导师. 主要研究方向为人工智能安全技术、机器学习与隐私保护.
E-mail: fuam@ njust. edu. cn

柯海峰　男,1993 年生于湖北鄂州. 南京理工大学计算机科学与工程学院硕士研究生. 主要研究方向为隐私保护.
E-mail: 14700533@ qq. com

苏　铓　女,1987 年生于内蒙古翁牛特旗. 现为南京理工大学计算机科学与工程学院副教授. 主要研究方向为云安全、访问控制与权限管理.
E-mail: sumang@ njust. edu. cn

孙怀江　男,1968 年生于陕西西安. 现为南京理工大学计算机科学与工程学院教授、博士生导师. 主要研究方向为神经网络与机器学习.
E-mail: sunhuaijiang@ njust. edu. cn