Times New Roman

题目（中英文题目一致)字体为2号黑体(全文除特别声明外，外文统一用Times New Roman)

作者名[1)] 作者名[2),3)] 作者名[3)](*字体为3号仿宋*作者)*

[1)](单位全名部门(系)全名, 市(或直辖市) 国家名邮政编码) *字体为6号宋体*单位

[2)](单位全名部门(系)全名, 市(或直辖市) 国家名邮政编码)*中英文单位名称、作者姓名须一致*

[3)](单位全名部门(系)全名, 市(或直辖市) 国家名邮政编码)

论文定稿后，作者署名、单位无特殊情况不能变更。若变更，须提交签章申请，国家为中国可以不写，省会城市不写省名，其他国家必须写国家名。

摘 要 在预测服务市场中，模型训练者需要从数据拥有者处购买数据，随后训练机器学习服务模型，并为服务定价作为商品出售给模型使用者。现有市场为了激励数据所有者参与交易，允许数据拥有者将扰动后的数据售予服务提供者，同时为服务提供者提供可选的扰动认证服务，以减少数据拥有者添加过量扰动的可能性。现有的预测服务定价方法并未考虑服务提供者与数据拥有者关于数据扰动认证的博弈。本文研究了基于机器学习的预测服务在恶意扰动下的利润最大化定价问题，综合考虑了需求多样性和隐私保护性。首先推导了服务提供者在可认证扰动下的最优解，然后开发了基于利润比保障的有效框架以解决信息不完全的优化问题。该框架不仅能高效获得收益损失有界的次优服务价格，还能通过最大似然估计有效估计服务质量函数。基于真实数据集的实验表明，相较于现有最优方法，我们的工作在理论验证和实际效能方面均表现出显著优势。

关键词 数据交易；利润最大化；非完全信息博弈；可验证性需求；隐私保护

中图法分类号 TP DOI号: *投稿时不提供DOI号

Title *（中英文题目一致)字体为4号Times New Roman,加粗* Title

NAME Name-Name[1)] NAME Name[2)] NAME Name-Name[3)] *字体为5号Times new Roman*Name

[1)](Department of ****, University, City ZipCode, China) *字体为6号Times new Roman* Depart.Correspond

[2)](Department of ****, University, City ZipCode)*中国不写国家名*

[3)](Department of ****, University, City ZipCode, country)*外国写国家名*

Abstract In the prediction service market, model trainers need to purchase data from data owners, train machine learning service models, and price the services as commodities for sale to model users. Existing markets incentivize data owners to participate by allowing them to sell perturbed data to service providers while offering optional perturbation verification services to reduce the likelihood of data owners adding excessive perturbations. Current

*收稿日期： - - ；最终修改稿收到日期： - - .*投稿时不填写此项*. 本课题得到… …基金中文完整名称(No.项目号)、… …基金中文完整名称(No.项目号)、… … 基金中文完整名称(No.项目号)资助.**作者名1(通信作者)**，性别，xxxx年生，学位(或目前学历)，职称，是/否计算机学会(CCF)会员（提供会员号),主要研究领域为*****、****.E-mail: **********.**作者名2（通信作者)**，性别，xxxx年生，学位(或目前学历)，职称，是/否计算机学会(CCF)会员（提供会员号),主要研究领域为*****、****.E-mail: **********. **作者名3（通信作者)**，性别，xxxx年生，学位(或目前学历)，职称，是/否计算机学会(CCF)会员（提供会员号),主要研究领域为*****、****.E-mail: **********.(给出的电子邮件地址应不会因出国、毕业、更换工作单位等原因而变动。请给出所有作者的电子邮件) 第1作者手机号码(投稿时必须提供，以便紧急联系，发表时会删除): … …, E-mail: … …*此部分6号宋体*

prediction service pricing methods do not account for the game-theoretic interactions between service providers and data owners regarding data perturbation verification. This paper studies the profit-maximization pricing problem for machine learning-based prediction services under malicious perturbations, jointly considering diverse demand and privacy preservation. We first derive the optimal solution for service providers under verifiable perturbations, then develop an effective framework with a profit ratio guarantee to address the optimization problem under incomplete information. This framework not only efficiently obtains sub-optimal service prices with bounded revenue loss but also accurately estimates the service quality function via maximum likelihood estimation. Experiments on real-world datasets demonstrate that, compared to state-of-the-art methods, our work achieves significant advantages in both theoretical validation and practical efficiency.

KeywordsData trading; profit maximization; incomplete information; availability requirement; privacy preserved.

1 引言

数据与数据驱动的服务在各领域具有广泛的运用，尤其是数据驱动的预测服务，包括股票预测、图像识别等。其具有零知识服务特性，意即使用者可以在不具备计算资源，不掌握机器学习的知识的前提下能使用预测服务。相较于传统的机器学习应用更加经济和方便。因此使用基于机器学习的预测服务市场欣欣向荣。

为了进行预测服务，服务提供者需要从数据拥有者处购买数据，使用数据训练机器学习模型，并将模型作为服务出售。期间服务提供者需要决定购买的数据量，测试模型的质量，为机器学习模型的服务定价，以获取最高的利润。

然而在购买数据时，服务提供者不知道所购买的数据对模型的预测结果有何影响，数据拥有着也没动机公布数据质量，市场中也存在不同类型的服务购买者。除此之外，数据拥有者提供的数据可能包含隐私，不能直接用这种隐私数据训练服务模型。

为了解决数据隐私的问题，可以使用基于查询的数据出售方案进行数据交易。基于查询的数据出售方式能支持广泛的运用途径，适合不同场景下的模型训练问题，同时支持订制数据的准确性与隐私性，能让训练师有所选择，保持平衡。该技术通过在原始数据中添加扰动，使用差分隐私等手段保护数据拥有者的隐私。在基于查询的数据出售方案中，服务提供者决定所购买数据的精密度系数，数据拥有者根据精密度系数选择所添加的扰动和收取的费用，在为原始数据添加扰动后，将扰动后数据发送给服务提供者。

添加扰动的方法有两种，一种是请可信第三方添加，另一种是让数据拥有者自行添加。可信第三方难以找到，而且费用很高。倘若让数据拥有者去添加噪声的话，拥有者有动机去添加过量噪声，保护自己隐私的同时减少了服务提供者训练的模型质量，损害了服务提供者的利益。

因此在基于查询的数据出售方案中引入了基于同态加密的认证系统。若数据拥有者提供的数据相当可疑，服务提供者可以对该数据发起认证。一旦发起认证，借助引入诚实但好奇的第三方，服务提供者便可以得知数据拥有者是否添加了过量扰动，若添加了过量扰动，数据拥有者便需要为自己的恶意行为赔偿服务提供者。若数据拥有者没有添加过量扰动，则其数据隐私遭到泄露，服务提供者需要为认证过程赔偿数据拥有者。

因此，本文首次考察了在可认证的隐私保护的，有不同种类的服务购买方的不完全信息的数据市场，主要贡献如下所示：1、我们是第一个考虑服务市场模型中存在恶意扰动者。2、我们对一个三方市场进行建模，考察其中各个角色的需求与行动，并定量分析他们的最优策略为何。3、我们设计了一套行之有效的市场机制，能保证数据拥有者

倾向于按照协议添加扰动。同时市场机制满足个体理性、群体理性和免套利性。

2 相关工作

2.1 数据定价

较低的出售价格能吸引更多的买家，而更高的售价可以提升单件物品的利润，该如何平衡两者以获得最高收入？收入最大化，又称利润最大化问题，是传统经济学与博弈论领域已得到充分研究的问题，其一大成果便是数据定价。瓦尔拉斯均衡解决了完全竞争市场中卖家的收入最大化问题，单边拍卖也能从购买方角度提供解决方案。然而以上工作君基于边际效益，数字产品具有边际成本近乎为零的特点，其复制成本低廉的特性使得大数据能免费传播。原则上而言，最有效的数据交易价格就是零，这使得以上的定价方案在数据市场中均不适用[14]。

已有的数据定价方案较为零散，适用于不同的场景，包括SQL查询[12][13][14]、线性统计（linear aggregate query）模型[15][16][17][18]、机器学习模型[19][20]、机器学习分类[21][22]、物联网数据等[23]。这些工作缺少利润的定量计算与优化，或者对市场的信息透明度有着强假设，因此不适用于我们的场景。

2.2 代理人模式

使用经典的委托-代理人模式，由[1]首次提出，假设数据代理人（data agent）是可信的第三方，让出售者提交数据，代理人扰动数据以达到免套利的效果。在这个机制中，数据购买方向数据拥有者的隐私进行赔偿后便可以计算添加的扰动多寡，从而对扰动的合法性发出质疑。当数据拥有者提高扰动时，能减少自己的隐私损失，同时让数据的可用性降低，因此能相应减少数据的价格，以增加销量。这就构成了一个多变量优化问题。之后有许多工作试图解决该优化问题以最大化代理人的收益，[2]的工作大多基于公开信息博弈，假设代理人知道数据拥有者的损失函数进行优化，[3]则取消了上述限制，根据之前的交易结果推断下次交易时的策略。再者，数据之间通常并非独立的，而是有一定的相关性，相关数据的隐私分析更为困难。为了解决上述问题，[5][6]引入了差分隐私技术，计算数据拥有者的隐私损失，并在时序数据市场[7]、关联子查询数据市场[8]等领域进行考察，在最大化数据代理人t的收益的基础上保护数据拥有者的隐私。

2.3 数据拥有者模式

在实际操作中，一个完全可信的第三方难以寻找且代价太高[9]，因此出现了由数据拥有者自行添加扰动的出售模式。[10]假设数据使用者和数据拥有者是商业联盟，并计算合适的隐私以最大化双方收益总和。然而市场中数据拥有者和数据使用者实际上是对立关系，数据拥有者有动机添加过量扰动，从而保护自己的隐私，对数据使用者而言则减少了数据可用性[9]。为了让数据拥有者能按照协议添加合适的扰动，需要为数据拥有者引入负反馈机制，即允许数据使用者对数据进行质疑。然而在数据交易市场中，数据使用者无法获得未扰动的原始数据，使得质疑困难重重。

3 市场模型

3.1 可认证扰动

对于$\epsilon-$差分隐私而言，其渐近误差边界与$\frac{1}{\epsilon}$成正比，与模型的训练时长$\frac{1}{n^{0.5}}$成正比**?**。

本文所考虑的模型如上图所示，其创新点在于考察了预测模型市场中，数据拥有者可能存在的造假问题。模型一共包含4方角色，分别是数据拥有者(Data Seller)、服务提供者(Service Provider)以及服务消费者(Service Consumers)以及第三方(Third Party)。

在服务提供者向数据拥有者购买数据的过程中，服务提供者需要决定所购买的数据的隐私扰动系数ϵ，使用同态加密技术生成公钥pk和私钥sk，并公布公钥pk。第三方随即根据隐私扰动系数ϵ生成相应强度的扰动$\eta(\epsilon)$，使用服务提供者的公钥加密得到$pk(\eta(\epsilon))$，并将其发送给数据拥有者。

数据拥有者接收$pk(\eta(\epsilon))$)后，使用服务提供者的公钥加密数据商品d_i得到$pk(d_i)$，并决定此次交易使用的策略s_{do_i}选择添加的扰动量ϵ。若选择诚实策略$s^h_{do_i}$，则选择从第三方接收$pk(\eta(\epsilon))$，并使用同态加算法计算得到$pk(d_i+\eta(\epsilon)) = pk(d_i)\oplus pk(\eta(\epsilon))$,此时添加的扰动量$\eta = \eta(\epsilon)$；若选择恶意扰动策略，则生成过量扰动$\eta_m$，使用服务提供者的公钥计算$pk(\eta_m)$，并使用同态加算法计算得到$pk(d_i+\eta_m) = pk(d_i)\oplus pk(\eta_m)$,此时添加的扰动量$\eta = \eta_m$。并将$pk(d_i+\eta(\epsilon))$发送给服务提供者。服务提供者接受密文后，解密$pk(d_i+\eta)$得到$d_i+\eta$，随即根据隐私扰动系数$\epsilon$，向数据拥有者支付价格$p_{do_i}(\epsilon)$，并决定是否质疑数据拥有者添加了过量扰动.如若决定质疑，则向第三方发起认证要求。

当服务提供者发起认证要求之后，第三方将扰动密文$pk(\eta(\epsilon))$发送给服务提供者，服务提供者解密得到$\eta(\epsilon)$，计算$d_p = d_i+\eta-\eta(\epsilon)$，计算其哈希值$H(d_p)$，并将$H(d_p)$发送给第三方。同时服务提供者需要将数据原文的哈希值$H(d_i)$发给第三方。第三方通过比较$H(d_i)$和$H(d_p)$是否相等，就能判断数据提供方是否进行了恶意扰动。无论认证结果如何，数据原文都会暴露给服务提供者。因此若数据提供方没有添加过量噪声，质疑结果$\phi=0$，服务提供方就需要给数据提供方补偿。反之，若$\phi=1$，数据提供方需要为自己的恶意扰动付出代价，给服务提供方补偿。

若未发起认证或者数据拥有者没有添加恶意扰动，服务提供者便使用密文训练模型，评估模型质量q，提供预测服务，并决定预测服务的价格p。

定义：预测模型为了分析服务提供者训练预测模型的过程，首先需要考察模型训练过程。给定数据集S_i，服务提供者使用该数据训练，获取预测模型$h(S_i)$。服务消费者给定数据x，模型h关于x的预测结果$h(x)$的正确率为$q_h(x)$。

为简洁起见，我们记使用数据集S_i购买数据量n_i训练成的集合$h(S_i)$的准确率为$q(S_i)$。不失一般性，假设训练用数据越多，模型质量越好，即$q(x)$为关于x单调递增的函数，$q'(x)>0$。由于模型质量关于训练用数据的边际效益递减，当随着训练用数据规模的增大，添加等量数据对模型质量的提升效果逐渐变小，因此我们假设$q''(x)<0$。

文中记号

记号	*记号的含义*
DO=$\{do_i\}_{i=1}^m$	数据拥有者
SC=$\{sc_j\}_{j=1}^n$	服务消费者
SP	服务提供者
ϵ	扰动系数
$p_{do_i}(\epsilon)$	数据拥有者i的要价
$g_{do_i}(\epsilon)$	数据拥有者i的隐私损耗
p	预测服务单价
$q_i(\epsilon)$	预测服务质量
$f_j(q)$	服务购买者j的收益
$d_j(q)$	服务购买者j的购买量

服务提供方完成训练后，需要为机器学习服务进行定价，确定单位服务的单价p。在本文中，我们使用线性模型拟合服务提供者的收益于支付价格，即服务消费者的收益与模型质量q成正比，需要支付的价格与购买量d成正比。鉴于服务购买方的购买量和单价成反比，服务提供方需要选取合适的价格p，以最大化其利润。

为了解决上述问题，先考察服务消费者在给定服务单价下的策略，为此需要分析其效用，定义其效用函数U_{sc_j}如下：

$$U_{sc_j}(p,q,d) = q\cdot f_j(d,q) - p\cdot d \tag{1}$$

其中$f_j()$为服务消费者j的效用函数，体现了服务市场中j类型的服务消费者的市场特征。d为服务消费者j所购买的服务量，购买d份量的服务后，j的收益为$q\cdot f_j(d,q)$。p为服务提供者设定的单位价格，购买d份预测服务时需要向服务提供者支付$p\cdot d$的费用。

可以看到，当f和p给定时，c为关于d的函数。如前文所述，我们假设服务消费者的效用函数a均

为凸函数，那么b也是凸函数。由凸函数的性质可知，U_{sc_j}达到最大值时当且仅当一阶导为零，即满足：

$$q \cdot f_j'(d) - p = 0 \tag{2}$$

上式表明理性的服务购买者j所购买的服务数量d_j 应如下所示：

$$d_j(p,q) = \max\{0, f_j'^{-1}(p/q)\} \tag{3}$$

可以推断出，在模型质量为q，定价为p的情况下，第j种数据购买方的所提供的交易额为$p \cdot d_j$. 设第j种服务消费者在市场中的占比为 π_j，那么服务提供者的总收益为：

$$r_{SP}(p,q) = p \cdot \sum_j \pi_j \cdot d_j \tag{4}$$

下面考察数据拥有者和服务提供者的交易，在数据市场中，服务提供者在数据拥有者中挑选数据，支付价格，购买定量数据，处于冲突状态。因此我们假设数据拥有者和服务提供者的交易遵循完全信息非合作博弈模型。该博弈模型要求参与者在进入市场时即公布自己的效用函数,同时在知道对手效用函数的前提下, 以自身收益最大化为目标选择策略。下面分析数据拥有者和服务提供者的纯策略。

在数据市场博弈模型中，数据拥有者可以选择诚实地添加扰动($s_{do_i}^h$)，或者恶意添加扰动($s_{do_i}^m$)两种策略，以下简记为(h)和(m)。服务提供者可以选择相信数据提供者诚实添加扰动从而不采取认证(s_{SP}^b)，或者质疑数据提供者添加过量扰动并采取认证(s_{SP}^v)两种策略，以下简记为(b)和(v)。我们记数据拥有者的收益函数为$u_{do_i}(x,y)$，服务提供者的收益函数为$u_{SP}(x,y)$，其中x为数据拥有者的策略，y为服务提供者的策略。数据拥有者采用诚实策略的概率为$p_{do_i}^h$，采用恶意扰动的概率为$p_{do_i}^m$.服务提供者采用相信的概率为p_{SP}^b，采用质疑的概率为p_{SP}^v。

那么数据拥有者在服务提供者选择相信策略的收益为：

$$\begin{aligned} U_{do_i}(p_{do_i}, p_{SP}) = p_{do_i}^h \cdot u_{do_i}(h,b) \\ + p_{do_i}^m \cdot u_{do_i}(m, s_{SP}^b) \end{aligned} \tag{5}$$

那么数据拥有者在服务提供者选择相信策略的收益为：

$$\begin{aligned} U_{do_i}(p_{do_i}, p_{SP}) = p_{do_i}^h \cdot u_{do_i}(h,v) \\ + p_{do_i}^m \cdot u_{do_i}(m, s_{SP}^v) \end{aligned} \tag{6}$$

那么服务提供者在数据拥有者选择诚实扰动策略的收益为：

$$\begin{aligned} U_{SP}(p_{do_i}, p_{SP}) = p_{SP}^b \cdot u_{SP}(h,b) \\ + p_{SP}^v \cdot u_{SP}(h,v) \end{aligned} \tag{7}$$

那么服务提供者在数据拥有者选择恶意扰动策略的收益为：

$$\begin{aligned} U_{SP}(p_{do_i}, p_{SP}) = p_{SP}^b \cdot u_{SP}(h,b) \\ + p_{SP}^v \cdot u_{SP}(h,v) \end{aligned} \tag{8}$$

下面考察服务提供者和数据拥有者在纯策略下的效用$u_{SP}(s_{do_i}, s_{SP})$和$u_{do_i}(s_{do_i}, s_{SP})$， 我们记服务提供者的要价为$p_{do_i}$，记数据拥有者的隐私损耗（privacy loss）在扰动系数 ϵ下为$g(\epsilon)$，那么有：

若数据拥有者选择诚实添加扰动，服务提供者选择不发起认证质疑，那么服务提供者依协议向数据拥有者支付购买费用，而无需支付认证费用或者赔偿费用，因此数据拥有者i的效用为：

$$u_{do_i}(h,b) = p_{do_i} - g(\epsilon) \tag{9}$$

服务提供者的效用为：

$$u_{SP}(h,b) = r_{SP}(p,q) - p_{do_i} \tag{10}$$

其中$r_{SP}(p,q)$为服务提供者为服务消费者提供服务所获得的总收益。

若数据拥有者选择诚实添加扰动，服务提供者选择发起认证质疑，那么服务提供者依协议向数据拥有者支付购买费用，还要支付赔偿费用，因此数据拥有者i的效用为：

$$u_{do_i}(h,v) = p_{do_i} + c_{do_i} - g(0) \tag{11}$$

服务提供者的效用为：

$$u_{SP}(h,v) = r_{SP}(p,q) - p_{do_i} - c_{do_i} \tag{12}$$

其中c_{do_i}是服务提供者向数据拥有者i支付的费用。

若数据拥有者选择恶意添加过量扰动，服务提供者不选择发起认证质疑，那么服务提供者依协议向数据拥有者支付购买费用，还几乎不能获得收益，因为训练模型的数据被扰动了。在该情形下，数据拥有者i的效用为：

$$u_{do_i}(m,b) = p_{do_i} \tag{13}$$

服务提供者的效用为：

$$u_{SP}(m,b) = -p_{do_i} \tag{14}$$

若数据拥有者选择恶意添加过量扰动，服务提供者选择发起认证质疑，那么数据拥有者还需要向服务提供者支付赔偿。在该情形下，数据拥有者i的效用为：

$$u_{do_i}(m,v) = p_{do_i} - c_{SP} \tag{15}$$

服务提供者的效用为：

$$u_{SP}(m,v) = c_{SP} - p_{do_i} \tag{16}$$

其中c_{SP}为数据拥有者在添加恶意扰动并被服务提供者认证成功时，需要支付给服务提供者的补偿。

服务提供者采用相信策略的概率为p_{SP}^b，采用认证策略的概率为p_{SP}^v，那么数据拥有者的期望收益为：

$$U_{do_i} = \begin{cases} p_{SP}^b \cdot u_{do_i}(h,b) + p_{SP}^v \cdot u_{do_i}(h,v), S_{do_i} = h \\ p_{SP}^b \cdot u_{do_i}(m,b) + p_{SP}^v \cdot u_{do_i}(m,v), S_{do_i} = m \end{cases} \tag{17}$$

其中h意为数据拥有者采纳了诚实扰动策略，m意为采纳了恶意扰动策略。

记数据拥有者采用诚实扰动的概率为$p_{do_i}^h$，采用恶意扰动的概率为$p_{do_i}^m$，那么服务提供者的期望收益为：

$$U_{SP} = \begin{cases} p_{do_i}^h \cdot u_{SP}(h,b) + p_{do_i}^m \cdot u_{SP}(m,b), S_{SP} = b \\ p_{do_i}^h \cdot u_{SP}(h,v) + p_{do_i}^m \cdot u_{SP}(m,v), S_{SP} = v \end{cases} \tag{18}$$

最后，我们考察数据拥有者和服务提供者的博弈达到纳什均衡的状态时的两者策略。此时无论服务提供者采用何种混合策略都不影响数据拥有者的效益，同时无论数据拥有者采用何种混合策略都不影响服务提供者的效益，因此有：

$$\begin{cases} u_{do_i}(S_{do_i} = s_{do_i}^h) = u_{do_i}(S_{do_i} = s_{do_i}^m) \\ u_{DC}(S_{DC} = s_{DC}^b) = u_{DC}(S_{DC} = s_{DC}^v) \end{cases} \tag{19}$$

可以得出数据拥有者的纳什均衡策略为：

$$\begin{cases} p_{do_i}^h = \frac{c_{SP}}{c_{SP}+c_{do_i}} \\ p_{do_i}^m = \frac{c_{do_i}}{c_{SP}+c_{do_i}} \end{cases} \tag{20}$$

特别地，在数据拥有者采用纳什均衡混合策略时，服务提供者的期望收益和纯策略相等，因此其期望收益为:

$$\begin{aligned} U_{SP}(p_{do_i}, p_{SP}) &= p_{do_i}^h \cdot u_{SP}(h,b) + p_{do_i}^m \cdot u_{SP}(m, s_{SP}^b) \\ &= \frac{c_{SP}(r_{SP} - p_{do_i}) - c_{do_i} p_{do_i}}{c_{SP} + c_{do_i}} \\ &= \frac{c_{SP} r_{SP}}{c_{SP} + c_{do_i}} - p_{do_i} \end{aligned} \tag{21}$$

接下来考察服务提供者的纳什均衡策略，此时有：

综上所述，在数据市场中，最大化服务提供者利润的优化模型如下所示：

$$\frac{c_{SP}r_{SP}}{c_{SP}+c_{do_i}}-p_{do_i} \tag{22}$$

其中：

$$r_{SP}(p,q)=p\cdot\sum_j \pi_j\cdot d_j \tag{23}$$

$$d_j(p,q)=\max\{0,f_j'^{-1}(p/q)\} \tag{24}$$

$$c_{do_i}\geq 0 \tag{25}$$

$$p_{do_i}\geq 0 \tag{26}$$

$$p\geq 0 \tag{27}$$

接下来对优化目标进行数学上的分析，记$C_{sp}=\eta\cdot\frac{\eta}{\eta+1}x\cdot D(x)-Q^{-1\prime}(q)=0Q^{-1\prime}(q)=\frac{\eta}{\eta+1}x\cdot D(x)$

令上式=0，解得

$$q=(Q^{-1\prime})^{-1}(\frac{\eta}{\eta+1}x\cdot D(x)) \tag{31}$$

记$y=\frac{\eta}{\eta+1}x\cdot D(x)$,有：

$$U_{SP}=q\cdot y-Q^{-1}(q) \tag{32}$$

且$q=(Q^{-1\prime})^{-1}(y)$

那么$y=Q^{-1\prime}(q)$，又因为$Q^{-1}(q)=p_{do_i}$，$y=\frac{1}{f'(p_{do_i})}$，因此$p_{do_i}=Q'^{-1}(\frac{1}{y})$，得出：

$$\begin{aligned}Q^{-1}(q)&=p_{do_i}=Q'^{-1}(\frac{1}{y})\\ q&=Q(Q'^{-1}(\frac{1}{y}))\end{aligned} \tag{33}$$

因此原式转换如下：

$$\begin{aligned}U_{SP}&=q\cdot y-Q^{-1}(q)\\ &=y\cdot Q(Q'^{-1}(\frac{1}{y}))-Q'^{-1}(\frac{1}{y})\end{aligned} \tag{34}$$

记$z=Q'^{-1}(\frac{1}{y})$，那么$\frac{1}{y}=Q'(z)$，$y=\frac{1}{Q'(z)}$,有：

$$\begin{aligned}U&=\frac{1}{Q'(z)}Q(z)-z\\ U'(z)&=-\frac{Q''(z)}{Q'(z)^2}Q(z)\end{aligned} \tag{35}$$

由于$Q(z)>0$，$Q'(z)>0$，$Q''(z)<0$，因此$U'(z)>0$恒成立，因此求原式最大值等价于求z最大值

由于

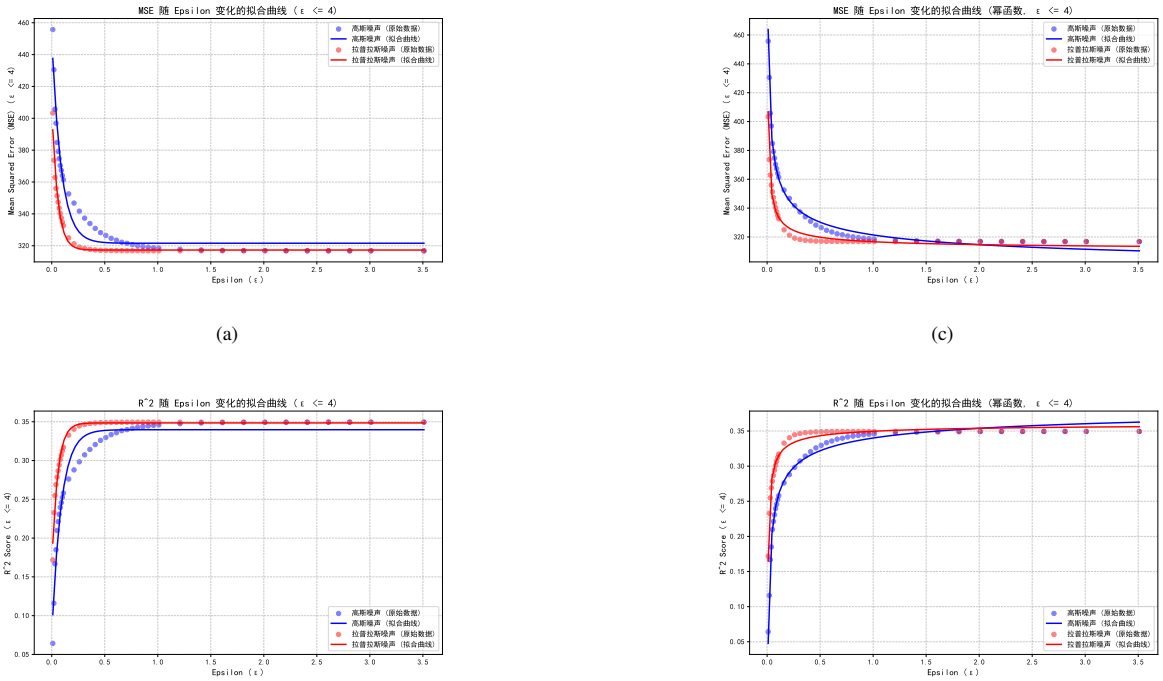

Fig. 1 跨栏图片示例

4 数学实验

4.1 实验设置

为了全方位考察保护隐私的数据市场机制，本次实验使用YEARMSD和MINST数据集模拟数据拥有者拥有的数据，模拟数据拥有者和服务提供者在数据出售阶段的博弈。各个数据集的详细信息如下表**??**所示：

Table 1 数据集详细信息

数据集	训练方案	训练集大小	测试集大小
YearMSD	线性回归	386609	128836
MINST	深度学习	60000	10000

为定量模拟数据拥有者和服务提供者在数据购买与模型训练阶段的博弈，本次实验考察了不同噪声添加下的数据所训练的模型质量，定量考察隐私保护与模型质量之间的关系，并进行函数拟合，结果如图**??**和**??**所示：

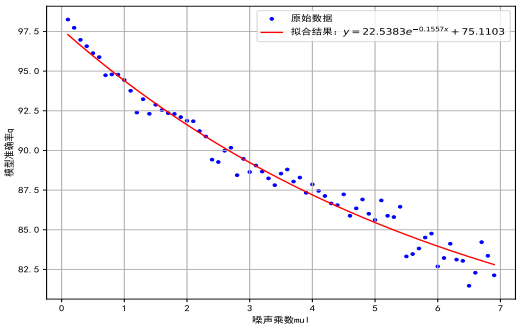

Fig. 2 MINST模型质量与噪声乘量

首先考察图**??**，图中横轴为训练用数据所添加的噪声乘量mul，纵轴为训练出来的模型质量，评判标准为准确率，单位为1%。实验结果表明，随着噪声乘量mul提高，原始数据所添加的扰动强度提高，数据的隐私保护效果提升，相应的所训练的深度学习模型的预测准确率降低，模型的质量逐步下降，，准确率从最初的97.5%下降到82.5%，说明隐私保护与模型质量不可兼得。同时注意到质量-噪声函数总体呈现下凸函数形，随着噪声的增加，准确率下降的速度越来越慢。是因为噪声具有边际递减特性：已添加的噪声越多，再添加等量的噪声对模型的影响程度越来越小。

同时我们考察了YearPredictionMSD模型质量与高斯差分以及拉普拉斯差分的关系，见图**??**。图**??**和图**??**描述了模型训练结果的均方误差（Mean Squared Error，MSE）损失函数关于噪声添加系数ϵ的关系，横轴为所添加的噪声添加系数ϵ，纵轴为MSE。MSE通过计算预测值和真实值之间的误差平方的平均值，来衡量模型的预测性能。MSE 越小，表示模型的预测越准确；MSE 越大，表示模型的预测误差越大，模型质量越差。图**??**和图**??**描述了模型训练结果的决定系数(R^2系数)关于噪声添加系数ϵ的关系，横轴为所添加的噪声添加系数ϵ，纵轴为R^2系数。R^2系数衡量模型对数据的拟合程度，取值范围通常在 0 到 1 之间。越接近 1，说明模型拟合效果越好。试验结果表明ϵ越低，说明差分隐私的保护越严格，数据所添加的扰动强度越高，数据的隐私保护效果越好，所训练的模型质量越差。随着ϵ的增加，噪声逐渐减弱，模型的质量逐步增加，并最终趋于一个定值。且从图**??**和图**??**中可以看出，使用负指数函数$y = a - b \cdot e^{-c \cdot x} \Phi a, b, c > 0 \Psi$对添加拉普拉斯噪声的模型质量进行拟合结果最好，因此使用拉普拉斯噪声的负指数函数拟合结果作为$q - \epsilon$的关系函数进行后续实验。

4.1.1 对比方案

实验引入了自保策略和3种定值策略，作为对比方案。

服务提供者购买数据后，将用YEARMSD和POWER数据集训练线性回归模型，使用SUSY训练逻辑回归模型，MINST数据集训练深度学习网络的识图模型。

服务购买者将根据模型的质量和自己的效益函数，购买预测服务。

4.2 数据交易模拟实验结果

4.2.1 案例说明

在深度学习模型的实验中，假定数据购买方的有5类，购买方的收益函数为$f = a_j \cdot (1 - e^{-b_j \cdot x})$，$0 < j < 6$。其中x为模型质量，$a_i$均服从$(50, 100)$的均匀分布，$b_j$均服从$(0.5, 1.5)$的均匀分布。根据前述实验拟

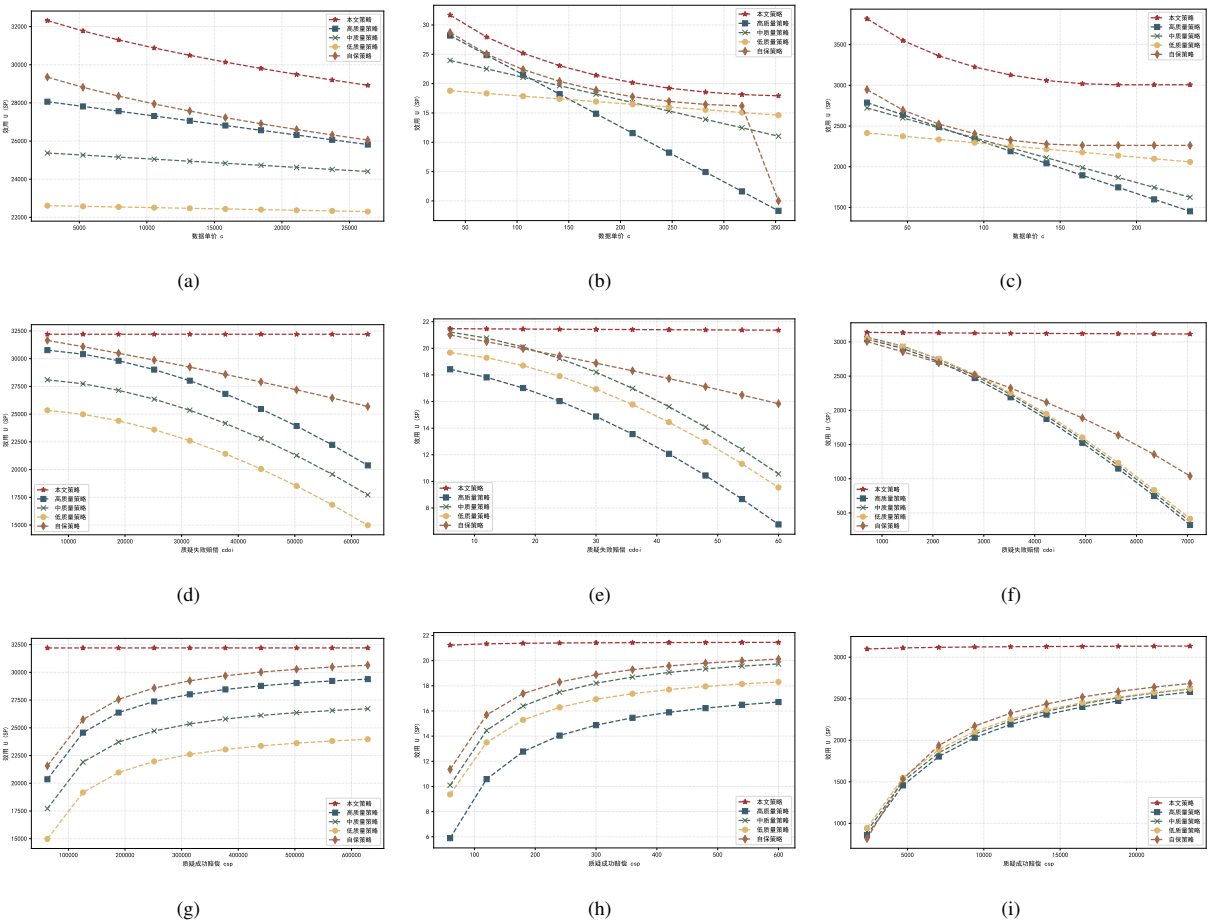

Fig. 3 跨栏图片示例

合所得的质量函数为$0.884 - 0.59 \cdot e^{-0.114 \cdot x}$，其中x为隐私扰动强度。

4.2.2 效益评估

下面考察深度学习模型的实验结果，其中使用深度学习计算的实验结果详见**??**，其中横轴为出售者的隐私敏感度，纵轴为服务提供者的收益图。图中可以看出，出售者用户隐私敏感度的增加，购买数据的成本急剧上升，传统的最高成本策略所获得的收益很快就不足以支付购买成本，净利润成为了负值。说明在数据市场中需要

Fig. 5 MINST模型质量与噪声乘量

4.2.3 模型准确率评估

我们采取了以下三个指标作为实验的测量指标：

(1)数据拥有者的诚信程度。

(2)服务提供者的总收益。

(3)服务的预测准确度，即服务质量。

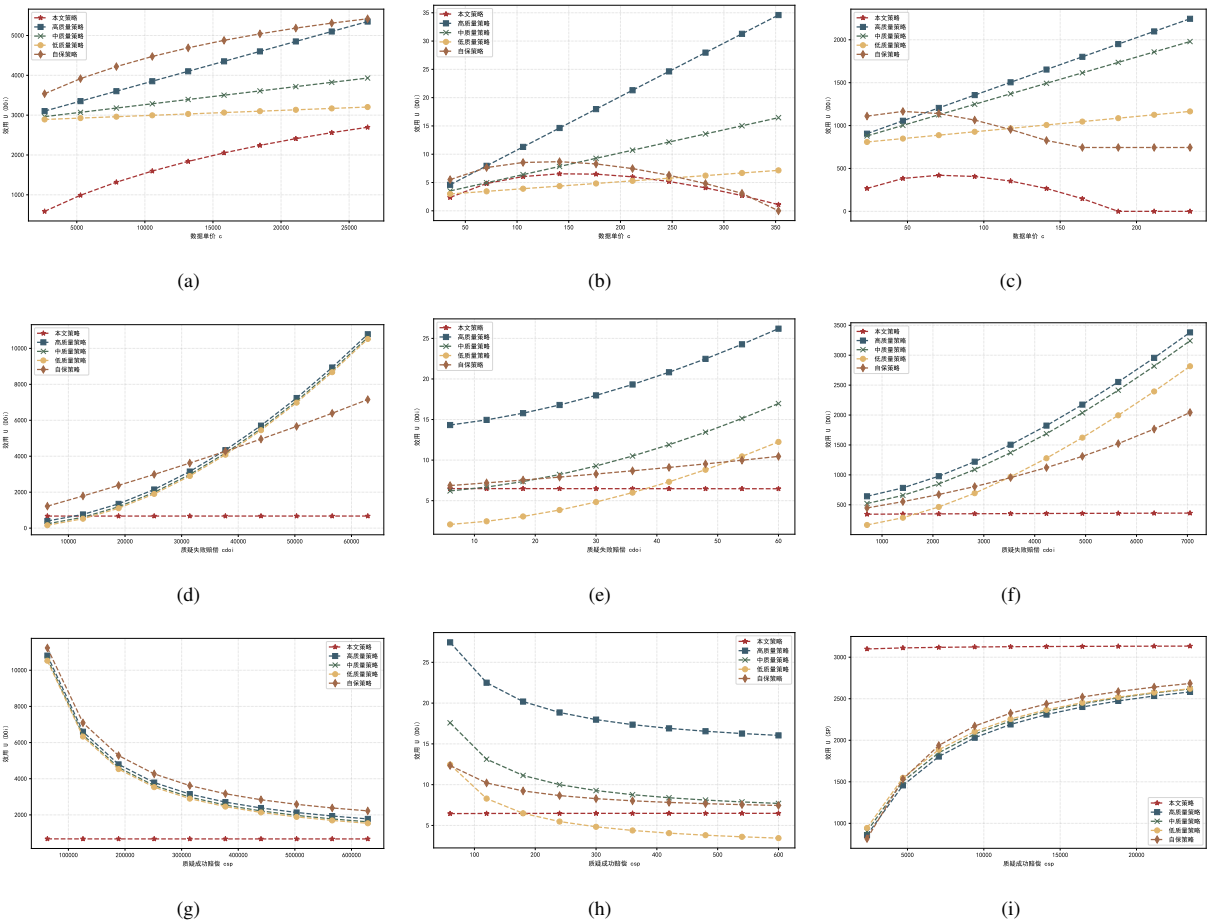

Fig. 4　跨栏图片示例

Fig. 7　This is a picture.

Fig. 6　图片说明 *字体为小5号，图片应为黑白图，图中的子图要有子图说明*

Fig. 8　第一个图片

Fig. 9　第二个图片

5　一级标题*字体为4号黑体*标题1

这是一个CTEX的utf-8编码例子，这里是楷体显示，这里是宋体显示，**这里是黑体显示**，这里是仿宋显示。另外一种方式：这是楷体显示，*but*英文和数字是斜体*abcABC123*，**这是黑体显示abcACB123**，这是仿宋显示abcABC123。

5.1　关于数学部分

数学、中英文皆可以混排。You can intersperse math, Chinese and English (Latin script) without adding extra

environments.

5.2 对投稿的基本要求

(1) 研究性论文主体应包括引言(重点论述研究的科学问题、意义、解决思路、价值、贡献等)、相关工作(为与引言部分独立的一个章节)、主要成果论述、关键实现技术、验证(对比实验或理论证明)、结论(结束语)等内容；系统实现或实验应有关键点的详细论述，以便读者能够重复实现论文所述成果。实验应有具体的实验环境设置、全面细致的数据对比分析。

(2) 综述应包括引言、问题与挑战、研究现状分析、未来研究方向、结论等内容。以分析、对比为主，避免堆砌文献或一般性介绍、叙述。

(3) 定理证明、公式推导、大篇幅的数学论述、原始数据，放到论文最后的附录中。

稿件提交时的基本要求：

(1) 本模板中要求的各项内容正确齐全，无遗漏；

(2) 语句通顺，无中文、英文语法错误，易于阅读理解，符号使用正确，图、表清晰无误；

(3) 在学术、技术上，论文内容正确无误，各项内容确定。

5.3 二级标题 *字体为5号黑体*标题2

5.3.1 三级标题 *字体为5号宋体*标题3

正文部分, 字体为5号宋体 正文文字

正文文字要求语句通顺，无语法错误，结构合理，条理清楚，不影响审稿人、读者阅读理解全文内容。以下几类问题请作者们特别注意：

1) 文章题目应明确反映文章的思想和方法；文字流畅，表述清楚；

2) 中文文字、英文表达无语法错误；

3) 公式中无符号、表达式的疏漏，没有同一个符号表示两种意思的情况；

4) 数学中使用的符号、函数名用斜体；

5) 使用的量符合法定计量单位标准；

6) 矢量为黑体，标量为白体；

7) 变量或表示变化的量用斜体；

8) 图表规范，量、线、序无误，位置正确(图表必须在正文中有所表述后出现，即…如图1所示)(注意纵、横坐标应有坐标名称和刻度值)。

9) 列出的参考文献必须在文中按顺序引用，即参考文献顺序与引用顺序一致，各项信息齐全(格式见参考文献部分)；

10) 首次出现的缩写需写明全称，首次出现的符号需作出解释。

11) 图的图例说明、坐标说明全部用中文或量符号。

12) 图应为矢量图。

13) 表中表头文字采用中文。

14) 公式尺寸：

标准：10.5磅

下标/上标：5.8磅

次下标/上标：4.5磅

符号：16磅

次符号：10.5磅

15) 组合单位采用标准格式，如：“pJ/bit/m^4”应为 “pJ/(bit·m^4)”

定理1.　******. *定理内容.*

[“定义”、“假设”、“公理”、“引理”等的排版格式与此相同，详细定理证明、公式可放在附录中]

证明.　*证明过程.* [“例 x”等的排版格式相同]

证毕.

示例图片

(请插入当时做图时的矢量版，如有当时的文件，例如Visio,origin,matlab, smartdraw,Execl,powerpoint 等各种软件作的图，图字用6号宋体，外文Times new roman，**图中文字尽量用翻译成中文**)

如插入图为截图，必将原作图文件(如*.vsd,*.opj, *.fig,*.sdr,*.eps,*.emf, *.wmf,*.ps 等后缀名)随修改稿压缩后传过来排版。

图X　图片说明 *字体为小5号，图片应为黑白图，图中的子图要有子图说明*

表X　表说明 *表说明采用黑体*

示例表格	*第1行为表头,表头要有内容*

过程X.　过程名称

《计算机学报》的方法过程描述字体为小5号宋体，IF、THEN等伪代码关键词全部用大写字母，变量和函数名称用斜体

算法Y.　算法名称.

输入：… …

输出：… …

《计算机学报》的算法描述字体为小5号宋体, IF、THEN等伪代码关键词全部用大写字母，变量和函数名称用斜体

5.3.2　参考文献

这是参考文献示例。参考文献应遵循GB/T 7741-2015标准。引用文献1 **?**，文献2 **?**，文献3-5 **???**。

致　谢　*致谢内容.* 致谢

参 考 文 献

附录X.

附录内容置于此处，字体为小5号宋体。附录内容包括：详细的定理证明、公式推导、原始数据等

First A. Author *计算机学报第1作者提供照片电子图片，尺寸为1寸。英文作者介绍内容包括：出生年,学位(或目前学历),职称,主要研究领域(与中文作者介绍中的研究方向一致).* *字体为小5号Times New Roman*

Second B. Author *英文作者介绍内容包括：出生年,学位(或目前学历),职称,主要研究领域(与中文作者介绍中的研究方向一致)。* *字体为小5号Times New Roman*

Background

论文背景介绍为英文，字体为小5号Times New Roman体

论文后面为400单词左右的英文背景介绍。介绍的内容包括：

本文研究的问题属于哪一个领域的什么问题。该类问题目前国际上解决到什么程度。

本文将问题解决到什么程度。

课题所属的项目。

项目的意义。

本研究群体以往在这个方向上的研究成果。

本文的成果是解决大课题中的哪一部分，如果涉及863\973以及其项目、基金、研究计划，注意这些项目的英文名称应书写正确。