武漢理工大學

《信息安全技术》
硕士研究生课程论文

学院：计算机与人工智能学院

学号：2024303053

姓名：胡姗

班级：软工专硕2405

深度学习中的隐私保护技术研究

摘　要

深度学习技术在多个领域取得了显著的成功，但其发展也伴随着数据隐私保护的挑战。由于深度学习模型训练需要大规模的数据集，而这些数据通常包含大量敏感信息，因此隐私泄露的风险显著增加。本文深入探讨了深度学习中的隐私保护技术，包括数据加密、差分隐私、数据处理等关键技术。这些技术旨在在保护数据隐私的同时，尽量不影响模型的训练效果和准确性。

关键词：隐私保护；深度学习；信息安全

Abstract

Deep learning technology has achieved remarkable success in many fields, but its development is also accompanied by the challenge of data privacy protection. Since deep learning model training requires large-scale data sets, which usually contain a lot of sensitive information, the risk of privacy leakage increases significantly. This article deeply explores the privacy protection technology in deep learning, including key technologies such as data encryption, differential privacy, and data processing. These technologies are designed to protect data privacy while minimizing the impact on the training effect and accuracy of the model.

Key Words：Privacy-Preserving；Deep Learning；Information Security

目　录

第1章 绪论

1.1 研究背景及意义

近年来，得益于神经网络理论和计算机硬件的进步，基于人工神经网络的深度学习已广泛应用于医疗健康[1]、金融分析[2]和大数据分析[3]等领域。在大多数现实场景中，深度学习的构建和训练需要搜集、剖析、保存和处理海量的信息，同时还需要配备许多的人力消耗与硬件设施，以此获得高精度的模型。然而这些数据可能涉及一些敏感信息，例如用户人脸图像、金融信息、医疗诊断记录等，这些隐私数据的泄露可能会导致不可弥补的财产损失甚至生命安全问题，包括但不限于身份盗窃、个人信息泄露等。因此，提出隐私保护深度学习模型至关重要，在保护用户数据隐私的前提下使用深度学习模型具有重要意义。

1.2 国内外研究现状

随着互联网技术日新月异的进步，隐私泄露愈发受到重视。为了应对隐私泄露导致的种种潜在风险， 各国政府都在积极行动，通过不断发布数据安全和隐私保护的相关法规，增强对企业和个人的规范力度。同时在隐私保护深度学习领域，已经涌现出了许多在深度学习过程中防止用户敏感信息泄露的技术。这些技术主要分为同态加密[5](Homomorphic Encryption，HE)、数据处理[6](Data Processing，DP)、差分隐私[7](Differential Privacy，DP)三大类。各种不同的技术方案旨在数据处理和模型训练以及测试中兼顾性能和隐私安全。

同态加密指的是某项计算操作在函数执行前后结果是完全相同的。基于同态加密的隐私保护方法具有较低的通信复杂度，被广泛应用于隐私保护的深度学习中，并且由于直接在密文上进行相关运算，大多数都可以支持任意的数据集划分形式。Phong L T[8]等人将加法同态加密技术和基于随机梯度下降算法的神经网络结合，在诚实但好奇的云服务器上保护模型训练过程中的梯度参数，所有的参数都加密后存储在云服务器上，在安全性和准确性方面都取得了良好效果。Alessandro Falcetta 与 Manuel Roveri[5]从理论和算法两个方面介绍了 BFV（Brakerski Fan Vercauteren）全同态加密方案及其具体实现，并提出了一种新的隐私保护的方案，并将其应用于卷积神经网络结构的设计中。

数据处理方法是通过对数据属性进行扰乱来实现隐私保护的目标。Lin K P[9]结合随机线性变换和随机扰乱核矩阵的方法，不仅有效保护了数据隐私，还显著降低了计算成本。Sagar Sharma[10]提出了多种图像伪装机制来对图像数据进行隐私保护，有效地抵御图像重建攻击，实现了数据实用性与数据和模型保密性之间的权衡。D Zhao[11]等人提出了一种基于矩阵变换的隐私保护深度学习模型 DLMT，将每个训练图像视为像素矩阵，并将其与随

机矩阵逐像素相加或相乘，变换后的数据与原始数据有明显差异，很难恢复原始数据，并通过参数很好地控制了隐私安全性和准确性之间的平衡。Chen K[12]等人提出了一种新颖的几何数据扰动策略，通过结合旋转、平行和距离等几何变换，对数据实施随机扰动。

差分隐私技术由 Dwork C[13]于 2006 年首次发布，其关键在于巧妙地在处理过程中向模型输出结果或者在数据处理过程的中间结果中引入适量的随机分布的噪声，从概率上确保不能根据分析查询结果来判定某个特定样本的存在性。这一策略有效地保护了数据中的隐私信息，使得数据分析能够在不侵犯个体用户隐私的情况下进行。Jalpesh Vasa 和 Amit Thakkar[14]详细阐述了几种基于深度学习的大数据分析隐私保护方法及其优缺点，并讨论了差分隐私的方法在保护隐私方面的有效性。

1.3 深度学习相关知识

深度学习的发展可以追溯到上世纪 50 年代提出的感知器模型，随后受受限玻尔兹曼机、深度信念网络等模型的启发，深度学习在上世纪 80 至 90 年代初期逐渐开始发展。近年来，随着科学技术的进步，深度学习凭借其卓越的能力，在处理高维、非线性、海量信息方面表现优异，从而广泛应用于处理分类、预测、识别等多样化任务中。

深度学习模拟了人脑神经网络内部的复杂结构，是机器学习的重要分支。同时，深度学习技术本身也在持续发展，传统的集中式学习方法正逐步向目前主流的分布式学习转变，为深度学习领域带来了新的发展动力。

深度学习的基本原理根植于人工神经网络，这种技术旨在模拟人大脑神经元之间的连接方式和工作机制。在人脑神经元内部，信号达到一定阈值即被激活，类似地，人工神经元通过激活函数来模拟这一过程。深度学习的数学原理与机器学习有共通之处，其核心在于神经网络具有强大的函数拟合能力，无论函数复杂程度如何。一般由输入层、隐藏层和输出层共同组成深度学习模型的结构，其中隐藏层可以包含多个节点。在神经网络的层级结构中，输入数据在逐层传递时都会进行一次数学变换，这些变换不断逼近目标函数，最终在输出层得到预测或分类的结果。

深度学习旨在从高维数据中提取复杂特征，并利用这些特征建立一个模型，将输入与输出（如分类类别）联系起来。深度学习架构通常以多层网络的形式构建，因此更抽象的特征是作为低层特征的非线性函数来计算的。本小节将具体阐述深度学习领域中神经网络的基本架构及其学习过程，并探讨卷积神经网络的核心原理。

第2章 深度学习中的隐私威胁

2.1 隐私目标

深度学习的飞速发展与进步，主要得益于对来自不同来源的大量数据进行收集，由于收集到的海量训练数据可能包含高度敏感的个人隐私信息，这些信息可能导致用户个人隐私受到严重威胁。因此模型训练过程存在许多风险，可能会造成严重的安全问题，在训练此类模型时，保护训练数据隐私是非常必要的。本小节主要讨论一些隐私目标：训练数据的隐私、模型的隐私和模型输出的隐私。

（1） 训练数据的隐私

目前我国关于个人信息数据的采集、存储和使用等方面尚缺少明确的法规与管理机制，大部分都靠各部门的自我约束，用户难以知晓自己的隐私信息以什么方式使用，并且数据可能包含一些高敏感数据，如图片、视频、语音等。训练数据中直接包含了隐私信息，攻击者获得训练数据，不用借助任何解密手段就能得到其包含的隐私。

（2） 模型的隐私

模型隐私主要指经过训练后的模型的参数，模型参数通过训练所得，其中包含了数据集的某些特征，模型参数泄露同样会导致敏感信息被非法获得。攻击者通过访问模型，可以逆向推断出模型的训练数据或模型的结构和参数。这种攻击可能暴露模型的知识产权和训练数据的隐私信息。

（3） 模型输出的隐私

攻击者通过模型的输出，从模型的输出中推测出训练数据的特征或模式，推断出关于训练数据的敏感信息。例如，通过模型对特定输入的反应，推断出训练集中可能存在的敏感数据模式。

由于人工智能在各种应用中的贡献，隐私已成为一个至关重要的问题。深度学习主要由数据采集阶段、模型训练阶段和模型预测阶段三部分组成，分别对应着输入层、隐藏层和输出层三部分，因此攻击者能够在深度学习过程的各个阶段进行隐私威胁和隐私攻击，在模型训练和预测两个阶段种，对隐私的威胁进行了总结，如表 2.1 所示。

表 2.1 深度学习中的隐私威胁[15]

阶段	威胁策略	敌手能力	敌手知识
训练	信息泄露	窃取训练数据	有限知识
预测	模型逆向攻击	模型提取	黑盒/白盒
预测	成员推理攻击	询问目标模型	黑盒

2.2 训练阶段的隐私威胁

在模型训练过程中，最大的隐私威胁安全隐患主要为训练数据集的隐私泄露。深度学习的训练方式包括集中式学习与分布式学习。其中，集中式学习即在一台中心服务器上集成所有参与方的训练数据进行学习；分布式学习的方式则是把训练数据和计算分别分配给不同的服务器，最后由中心服务器进行集成。

隐私威胁与深度学习部署结构密切相关。在集中式学习中，模型通过收集大量数据进行训练，在精度上具有很大的优势。然而，它也给集中式服务器带来了高负载，一旦发生攻击，所有的个人数据都将面临风险，如图 2.3 所示。而且深度学习模型和数据的隐私性与深度学习模型的使用方式无关，而与敌手访问托管该模型和数据的系统的程度有关，因此可以看作是一个传统的访问控制问题。

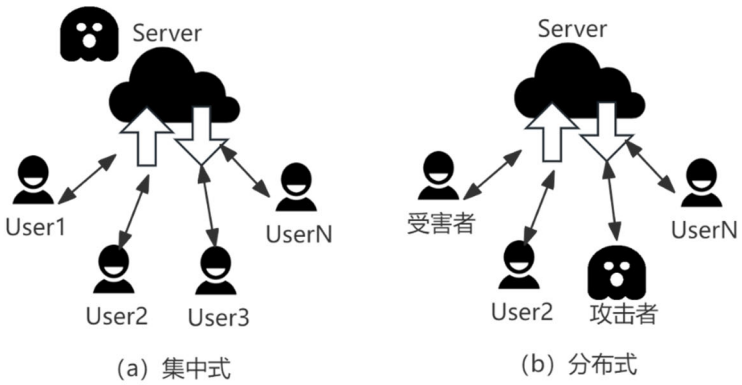

图 2.3 两种不同的模型训练方式

因此分布式协作学习被研究者提出并得到广泛应用，其中本地用户和集中式服务器分别承担部分训练任务，并且只共享参数的子集。如图 2.3(b)所示，如果存在恶意参与者，可以通过生成对抗网络(Generative Adversarial Networks，GAN)对其他参与者进行信息盗取，生成与受害者私有目标训练集分布相同的原型样本。在训练阶段，恶意用户始终处于活跃状态，欺骗受害者发布自己的隐私信息。

2.3 预测阶段的隐私威胁

训练完成的目标模型是深度学习整个过程的核心竞争力，因此对于预测阶段的隐私安全问题，需要予以高度警觉。在模型完成训练之后，通常会被用于预测各种特定的结果，辅助完成更为明智的决策。然而，一旦在预测阶段遭受恶意攻击，其潜在后果可能极为严重。预测阶段存在的安全及隐私威胁主要可以分为成员推断攻击、训练数据提取和模型提取攻击三个方面[16]。

(1) 成员推断攻击(Membership Inference Attack)

成员推断攻击是一种威胁，其中攻击者利用模型提供的预测接口，提供分析预测的结果来推测训练集中是否含有特殊的特征值。这种攻击方式不需要知晓模型的结构，学习方式、参数以及样本的分布情况等，而仅仅只需要知道预测类别的置信度即可实施。如图 2.4 所示，攻击者用数据记录查询目标模型，以获得该记录上的预测，这是一个置信度的向量。然后将向量与目标记录的标签一起传递，构建攻击模型。由于目标模型处理训练样本和未见样本的结果不同，攻击模型可以识别这种差异，并知道该记录是否属于目标模型的训练集。

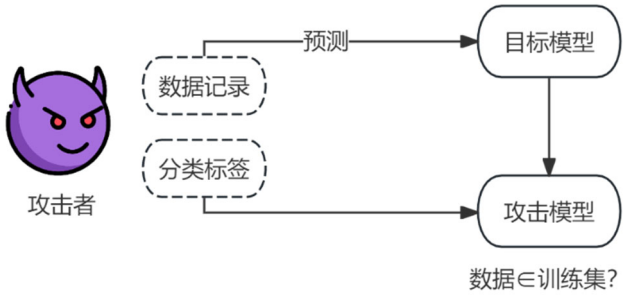

图 2.4 成员推断攻击

(2) 训练数据提取(Training Data Extraction)

训练数据提取，也称模型逆向攻击(Model Inversion Attack)，是一种基于模型预测 API 的攻击方法，通过一系列的查询来获取模型训练数据里的敏感信息的一种攻击方式。通过逆向攻击可以得到局部或者所有的训练样本，或者训练样本的统计特征。由于数据提取攻击造成的隐私数据泄露可能会造成巨大的威胁。

(3) 模型提取攻击(Model Extraction Attack)

模型提取攻击是指对基于私有数据训练的模型进行参数提取。攻击者的目的是模拟出一个模型的函数，相仿于训练出的模型的功能，并且在一个可验证的样本上，它可以达到和目标模型相近的预测效果。由于模型的参数和训练样本之间存在着密切的关系，因此当目标模型参数被泄露时，其隐私问题就会更加凸显出来。在此基础上，持续向目标模型提供样本并对返回结果进行记录，可以模拟重建出目标模型。

第3章 深度学习中的隐私保护

3.1 同态加密

同态加密，其定义如下：假定有一个加密函数F，明文M结果加密函数加密后变成密文M'，明文N加密后变成密文N'，即$F(M)=M'$，$F(N)=N'$，存在F对应的解密函数F^{-1}，能够将F加密后的密文解密为加密前的明文。将M'与N'相加得到P'，如果解密函数F^{-1}对P'解密后的结果等于M和N相加的结果，即$F^{-1}(P')=F^{-1}(M'+N')=M+N$，则$F$是符合同态加密性质的加密函数。同态加密可以使对方无法看到原始数据，也能保持数据的无损丢失，因此在保障数据保密性的同时，还可以保障数据的有效性，图 3.1 为同态加密针对深度学习的隐私保护流程。

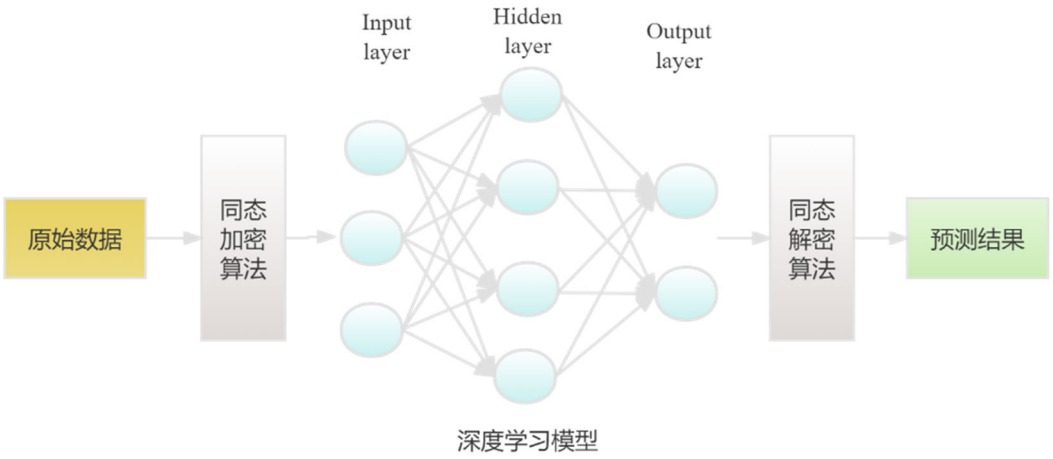

图 3.1 同态加密隐私保护流程

同态加密技术根据同态性质的不同，针对不同的计算模式可以分为下面三种类型：加法同态、乘法同态和全同态。不同之处在于满足的同态性质不同，其中加法同态满足$F(M)+F(N)=F(M+N)$，乘法同态满足$F(M)*F(N)=F(M*N)$，全同态则是指一个加密函数既能满足加法同态同时也满足乘法同态，可以实现加法同态和乘法同态满足的各项数学运算。

同态加密技术的关键在于它可以在密文上进行明文操作，将其解密后得到的结果等同于明文操作，是一种最直接、最有效的保护用户隐私的方法。然而，当前同态密码机制仍有诸多不足，例如只能支持整数，乘法深度固定，无法无限叠加乘法，且运算消耗大。因此，现有的同态加密方案不能简单地应用于深度学习中，需要解决的困难主要在于可行性和效率问题。

3.2 数据处理

数据处理，也称数据扰动。因其直接隐藏敏感信息并且易于操作而备受青睐，目前已经成为数据发布研究和应用领域的基础研究方向。该方法已经广泛应用于互联网、通信和银行等各个领域中，以保护隐私并支持深度学习。近年来，研究人员们一直在探索和改进新的数据扰动技术，这也是隐私保护深度学习领域的热点研究方向。本小节主要讨论以数据处理、数据变换为基础的隐私保护方法。

基于数据处理的隐私保护深度学习方法是指数据持有者在将数据发布参与模型训练之前对数据进行预先处理，以保护其中可能包含的敏感信息。其中，经典的数据扰动方法是添加随机噪声的方法，其简明的定义如下：

将一列数据的初值$(x_1, x_2, ..., x_n)$视为从具有某种分布的随机变量X中随机抽取的。随机化过程通过在原始数据值上添加随机噪声R来改变原始数据，并生成扰动数据列Y。

$$Y = X + R \tag{2.1}$$

扰动之后的数据以及噪声分布对数据访问者可见，它的原理就是在不改变特定的数据或属性的情况下，对敏感数据进行扭曲，从而确保被干扰后的数据仍能保留一定的统计特性，进而用于深度学习任务中。

随着隐私保护技术的不断研究和发展，在不同场景下基于数据处理提出了各种不同的隐私保护方法，这些方法已经被应用于实际场景中。例如基于图像混淆的方法（包括像素化、模糊化）、基于矩阵变换的方法、基于图像分割的方法等等各种变形。例如基于矩阵变换的隐私保护深度学习模型 DLMT[11]，在本地对数据进行预处理以隐藏敏感信息，然后将处理后的数据发送到服务器进行训练和测试，而不是在云服务器上使用原始数据训练模型。

每一个训练和测试数据都必须经过一个相同的随机矩阵R的变换，并且R的维数与原始数据的维数相同。对于一幅图像A，将它看成是一个维数为$W \times H \times C$的像素矩阵，其中W是宽度，H是高度，C是通道数。随机生成一个具有相同维数的矩阵R，用R元素逐元素相加或相乘A。R中的每个值都是区间$[1, MAX_V]$内的随机整数，MAX_V为正整数，将这两种方法分别称为矩阵加法变换(MAT)和矩阵乘法变换(MMT)，(3.2)式、(3.3)式分别表示 MAT 和 MMT，其中$i \in [1, W]$，$j \in [1, H]$，$k \in [1, C]$。

$$A'[i, j, k] = A[i, j, k] + R[i, j, k] \tag{2.2}$$

$$A'[i, j, k] = A[i, j, k] \times R[i, j, k] \tag{2.3}$$

如今，数据处理技术中的一个主要挑战是平衡隐私保护和数据效用，这通常被认为是一对相互冲突的因素。因此在扰动中选择性地保留任务/模型的特定信息将有助于实现更好的隐私保证和更好的数据效用。

3.3 差分隐私

差分隐私技术旨在为用户在不影响整体输出的情况下，提供一种基于概率分布的机制或协议，让使用者可以在不影响整体输出的情况下，让攻击者无从得知数据集上的有关个体的信息，实现隐私保护。

差分隐私技术通过在原始数据或者在其变换过程中加入一些可量化的随机噪音，以实现保护用户数据隐私的目的。该技术确保对相邻样本集（两个样本集 D 和 D' 满足结构和属性相同且只相差一条数据）进行相同的查询操作，其应答结果基本保证相同，即单条数据对整体结果的干扰具有有限性。差分隐私保护模型示意图如图 3.2 所示。

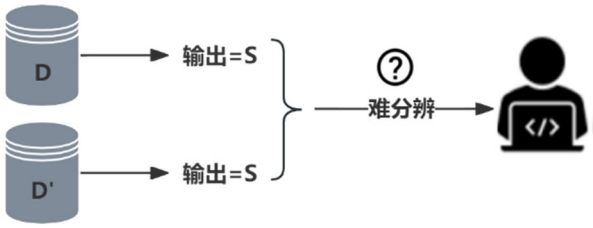

图 3.2 差分隐私保护模型

差分隐私技术在实际应用中，将需要差分隐私保护的数据区分为两种类型：数字型数据、非数字型数据。差分隐私技术以增加相应分布的随机噪声采样来保护用户的隐私，因此噪声的选择与隐私保护效果紧密相关，对于数据为数值类型的时候，通常采取拉普拉斯机制或高斯机制分布；对于数据为非数字类型的时候，通常会采取指数机制分布，通过引入评分函数，获得对应的得分，并将其规范化后作为查询的返回概率。三种实现机制[错误!未找到引用源.]如下所示：

定义 3.1 拉普拉斯机制：给定查询函数 f ： $D \to \mathbb{R}^d$ 作用于数据集 D，如果 M 满足下式，则称扰动机制 M 满足 ε -差分隐私。

$$M(D) = f(D) + Lap(S(f)/\varepsilon) \tag{2.4}$$

其中， $S(f)$ 为查询函数 f 的敏感度 l_1， $Lap(\bullet)$ 代表 $Laplace$ 函数，其分布满足均值为 0，尺度为 $S(f)/\varepsilon$ 的 $Laplace$ 分布。

定义 3.2 高斯机制：当 M 的输出类型为数值时，敏感度可以用 l_2 范数，向函数 f 的输出结果中添加高斯噪声来实现差分隐私，定义如下：

$$M(D) = f(D) + \mathbb{N}(0, S_f^2 \bullet \sigma^2 I) \tag{2.5}$$

式中， $\mathbb{N}(0, S_f^2 \bullet \sigma^2 I)$ 是均值为 0，标准差为 $S_f \cdot \sigma$ 的高斯噪声， I 表示单位矩阵，部分

高斯分布曲线如图 3.3 所示。

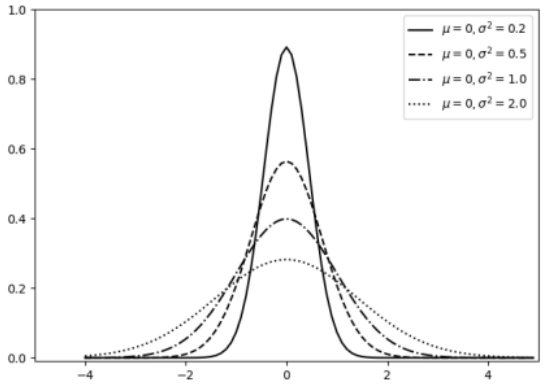

图 3.3 均值=0 高斯分布图

定义 3.3 指数机制：给定随机算法 M，若在数据集 D 上以正比于 $\exp\left(\frac{\varepsilon q(D,r_i)}{2\Delta q}\right)$ 的概率从输入中选择并输出 r_i，代表 M 满足 ε-差分隐私。

$$M(D,q,r_i)\sim\exp\left(\frac{\varepsilon q(D,r_l)}{2\Delta q}\right) \tag{2.6}$$

$$\Delta q=\max_{D,D'}\|q(D,r_i)-q(D',r_i)\|_1 \tag{2.7}$$

式中，$q(D,r_i)$ 代表的是数据集 D 中某一概率输出的结果分数，Δq 为 l_1 敏感度。

通过 2.2 小节可以得知，深度学习下的隐私攻击主要针对输入层、隐藏层和输出层，差分隐私保护的部署也会根据隐私攻击进行相对应的部署，通过对模型或数据添加满足上述不同机制的随机噪声进行扰动，来实现隐私保护。根据添加噪声进行扰动的部分不同，可以分为下面四种：输入扰动、梯度扰动、目标扰动和输出扰动，如图 3.4 所示。

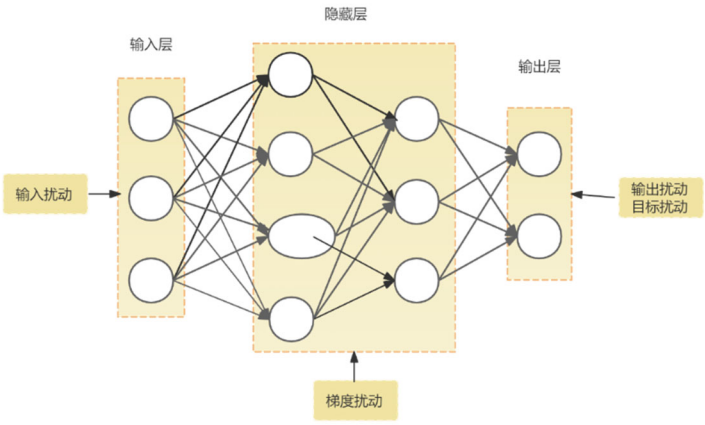

图 3.4 差分隐私保护方式

（1）**输入扰动**：对原始数据直接引入噪声来达到对数据的隐私保护，将扰动后的数据作为训练模型的输入进行训练。输入扰动主要划分成两种类型：加性噪声和乘性噪声，加性噪声是把随机噪声添加到原始数据中，乘性噪声是把随机噪声与原始数据相乘，输入扰动能够在损失较弱数据可用性的情况下提供隐私保护。

（2）**梯度扰动**：对深度学习过程的梯度变化添加噪声来实现数据的隐私保护。梯度反映了训练过程中损失函数对模型参数的变化率，可以通过向梯度添加高斯噪声或拉普拉斯噪声来实现，梯度扰动无需对数据做出改动就可实现隐私保护。

（3）**目标扰动**：对训练过程的目标函数引入随机噪声来产生扰动，进而达到隐私保护的目的。通过将其他差分隐私保护机制与目标扰动结合在一起，可以获得更良好的隐私保护效果，其中，函数机制通过扰动目标函数系数来保护隐私。

（4）**输出扰动**：对深度学习最终结果，即输出模型的输出结果引入噪声来实现隐私保护。能够在一定程度上保护模型的隐私性，以防敌手通过模型攻击分析出敏感信息，因为不需要对原始数据或每个训练样本的梯度进行干扰或存储，所以会显得更加高效。

第4章 总结与展望

伴随着深度学习技术的兴起，人工智能在各行各业掀起了一股新的浪潮，其安全性与隐私问题也日益突出，是深度学习进步路上的挑战之一。目前深度学习隐私保护领域的进展还在初始进程，尚有很多困难有待思考与处理。为此，后续研究可以从以下几个方面进行展开和深入：

（1）要建立健全的评价体系和法制保障体系。在此基础上，提出一种基于隐私泄漏的安全性评价指标和度量准则，并对相应的法律法规进行完善，从而有效地遏制企业或机构的违法行为。

（2）加强对变换数据的分类研究。变换数据的分类问题是当前该领域的难点，未来可以针对不同类型的变换数据设计更加有效的分类方法，提高模型的泛化能力和准确率，对隐私保护模型的相关参数进行提取。

参考文献

[1] Mujeeb Ur Rehman, Arslan Shafique, Yazeed Yasin Ghadi, et al. A Novel Chaos-Based Privacy-Preserving Deep Learning Model for Cancer Diagnosis[J]. IEEE Transactions on Network Science and Engineering,2022,9(6): 4322-4337.

[2] 闫洪举. 基于深度学习的金融时间序列数据集成预测[J].统计与信息论坛,2020,35(04): 33-41.

[3] 冯登国, 张敏, 李昊. 大数据安全与隐私保护[J].计算机学报,2014,36(10): 1-13.

[4] Shokri R, Shmatikov V. Privacy-preserving deep learning[C]. Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. ACM, pp: 1310-1321,2015.

[5] Alessandro Falcetta, Manuel Roveri. Privacy-preserving deep learning with homomorphic encryption: An introduction[J]. IEEE Computational Intelligence Magazine,2022,17(3): 14-25.

[6] D. Zhao, S. Liao, H. Li, et al. Competitor Attack Model for Privacy-Preserving Deep Learning[C], 2023 IEEE/ACM 23rd International Symposium on Cluster, Cloud and Internet Computing Workshops (CCGridW), pp.133-140, 2023.

[7] Abadi M, Chu A, Goodfellow I, et al. Deep learning with differential privacy[C]. Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. ACM, pp.308-318, 2016.

[8] Phong L T, Aono Y, Hayashi T, et al. Privacy-preserving deep learning via additively homomorphic encryption[J]. IEEE Transactions on Information Forensics and Security,2018,13(5): 1333-1345.

[9] Lin K P. Privacy-preserving kernel k-means clustering outsourcing with random transformation[J]. Knowledge and Information Systems,2016,49(3): 885-908.

[10] Sagar Sharma, AKM Mubashwir Alam, Keke Chen. Image Disguising for Protecting Data and Model Confidentiality in Outsourced Deep Learning[C]. In 2021 IEEE 14th International Conference on Cloud Computing (CLOUD), pp. 71–77, 2021.

[11] D Zhao, Y Chen, J Xiang, et al. DLMT: Outsourcing Deep Learning with Privacy Protection Based on Matrix Transformation[C]. In 2023 26th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Rio de Janeiro,Brazil, 2023.

[12] Chen K, Liu L. Geometric data perturbation for privacy preserving outsourced data mining[J]. Knowledge and Information Systems,2011,29(3): 657-695.

[13] Dwork C, Mcsherry F, Nissim K, et al. Calibrating Noise to Sensitivity in Private Data Analysis[C]. Theory of Cryptography Conference. Springer,Berlin, Heidelberg,2006.

[14] Jalpesh Vasa, Amit Thakkar. Deep learning: Differential privacy preservation in the era of big data[J]. Journal of Computer Information Systems,2023,63(3): 608-631.

[15] 宋蕾, 马春光, 段广晗. 机器学习安全及隐私保护研究进展[J]. 网络与信息安全学报,2018,4(8): 1-11.

[16] 唐鹏, 黄征, 邱卫东. 深度学习中的隐私保护技术综述[J]. 信息安全与通信保密,2019 (6): 11-19.