武漢理工大學

本科生毕业设计（论文）开题报告

学 生 姓 名：胡姗

导师姓名、职称：马小林 副教授

所 属 学 院：信息工程学院

专 业 班 级：信息 2001 班

设计（论文）题目：面向隐私保护深度学习的变换数据分类方法

2024 年 2 月 25 日

开题报告填写要求

1．开题报告应根据教师下发的毕业设计（论文）任务书，在教师的指导下由学生独立撰写。

2．开题报告内容填写后，应及时打印提交指导教师审阅。

3．“设计的目的及意义”至少 800 汉字，“基本内容和技术方案”至少 400 汉字。进度安排应尽可能详细。

4．指导教师意见：学生的调研是否充分？基本内容和技术方案是否已明确？是否已经具备开始设计（论文）的条件？能否达到预期的目标？是否同意进入设计（论文）阶段？

撰写内容要求（可加页）：

1．目的及意义（含国内外的研究现状分析）

近年来，深度学习已经成为学术界和产业界的最大热点之一，受到了人们的广泛关注，成为了人工智能应用最重要的驱动力之一。深度学习已广泛应用于医疗[1]、图像处理[2]和金融分析[3]等领域，助推了包括图像分类[4]、语音识别[5]、自然语言处理[6]等多个领域的发展。深度学习是一种用于产生和预测模型的工具，可通过汇聚多个贡献者的数据来获取大量的训练数据，随着训练数据量的增加，模型的准确性会越来越高。在大部分应用场景中，深度学习模型需要分析和处理海量的数据，而这些数据经常会涉及一些个人隐私，由于参与深度学习计算的数据具有极高的机密性和价值，因此应对数据存储、数据训练和数据推理的全流程进行加密，在保护数据隐私的前提下使用深度学习模型具有重要意义[7]。

目前已经有很多隐私保护模型被提出，但很多模型都存在一些安全性问题[8]。在真实应用场景中，数据监管机构可能需要通过数据流分析域内单位或企业是否采用了安全的隐私保护深度学习模型；另一方面，对于攻击者而言，要攻破某个单位或企业采用的深度学习模型，先要分析确定其采用了哪种隐私保护技术，再针对性设计攻击方案。因此，在未知隐私保护模型但截获了其变换后的数据的情况下，分析数据是采用了哪种隐私保护模型，对于提前发现安全漏洞以防御攻击者具有重要意义。但目前隐私保护深度学习领域还缺少这方面的工作。

在国外，研究者们已经充分认识到深度学习与隐私保护之间的紧密关系，针对训练数据的推理攻击,已经涌现出很多具有代表性的防御方案。这些方案大致基于三种隐私保护技术:差分隐私技术[9] (Differential Privacy)、安全多方计算技术[10] (Secure Multi-party Computing)以及同态加密技术[11] (Homomorphic Encryption)，旨在在数据处理和模型训练中兼顾性能和隐私安全。差分隐私机制的核心思路在于对模型输出结果或者处理过程中的中间结果添加适量的随机噪声，从而在概率上保证无法通过分析查询结果判断单个样本的存在与否。其优势在于易于实现且额外开销小，缺点在于噪声的添加会影响准确度，同时，由于深度学习是迭代式训练，迭代次数越多，噪声添加越多，准确度下降越多。另一类解决方案是基于加密技术，微软提出的Cryptonets[12]是其中的代表技术。Cryptonets 充分利

用了同态加密技术，将存储在云端的模型中的计算操作使用同态加密原语重写，使其可以直接在密文上计算并得到加密结果。该方案既保证了模型本身的安全性，因为模型本身被加密操作重写，同时还能保护用户数据及预测结果的隐私性。上述方案都是针对中心化训练的数据隐私保护，即数据所有者相同，对于多参与方而言，存在的挑战更大。目前主流的解决方案是联邦学习[13]，即多个参与方联合各方数据训练一个统一的模型，但保证各方数据隐私性，通常需要引入同态加密和多方安全计算技术。

国内深度学习应用取得显著进展，但对于隐私保护的研究相对较少。许多学者在差分隐私、同态加密、安全多方计算等[14][15]方面进行了一些探索，但在深度学习任务中的实际应用还处于初级阶段。例如沈传年[16]等人对深度学习中的隐私保护技术进行了相关综述，总结了目前在深度学习中常见的隐私保护方法及研究现状,包括基于同态加密的隐私保护技术、差分隐私保护技术等等；何其建[17]提出了一种基于变换层的保护隐私的合作深度学习方案，参与者通过一个服务器,综合各自的数据集合作地训练出一个深度学习模型,同时不披露他们的敏感输入数据；魏立斐[18]等人总结了机器学习常见的安全威胁，如投毒攻击、对抗攻击、询问攻击等，以及应对的防御方法，如正则化、对抗训练、防御精馏等。接着对机器学习常见的隐私威胁，如训练数据窃取、逆向攻击、成员推理攻击等进行了总结，并给出了相应的隐私保护技术，如同态加密、差分隐私。最后给出了亟待解决的问题和发展方向。当前，国内关于变换数据分类中隐私保护的系统性研究相对缺少，亟需深入挖掘和解决这一领域的实际问题。

综上所述，本课题将设计和实现一种面向隐私保护深度学习的变换数据分类方法，采用现有的主流的隐私保护深度学习模型来产生训练数据与标签，再基于训练数据构建分类模型，并分析分类模型的效率，通过实验验证模型的准确率和有效性。为深度学习在变换数据分类任务中的应用提供更为安全可靠的解决方案，促进数字社会的健康发展。

2．研究（设计）的基本内容、目标、拟采用的技术方案及措施

（1）基本内容

本设计的主要内容是学习并熟练运用深度学习框架和隐私保护手段的相关知

识，掌握相关编程知识，运用深度学习模型的理论和原理，结合国内外对隐私保护技术的相关研究，设计和实现一种面向隐私保护深度学习的变换数据分类方法。

（2）目标

1）熟悉常用的隐私保护技术手段，掌握深度学习模型的概念与原理；

2）掌握并复现差分隐私、同态加密等主流的隐私保护深度学习模型，并运用其方法生成模型训练标签数据集；

3）搭建深度学习神经网络模型，划分数据集后进行训练分类网络，对模型评估与调优；

4）集成上述工作，设计出一种完整的可解决隐私保护中变换数据分类问题的理论方案，并搭建神经分类网络对所提方案进行量化的性能分析。

5）完成整体设计，记录详细实现过程，并撰写设计论文。

（3）拟采用的技术方案及措施

1）在计算机上安装 Python、Pycharm 运行环境，配置 pytorch、tensorflow 等深度学习框架，阅读百度飞桨开源深度学习平台文档，学习如何使用在线平台训练深度学习模型。

2）学习深度学习、隐私保护基础理论知识，复现文献中基于图像的隐私保护技术方案，编写 10 种左右隐私保护方案的相关代码。

3）下载 MNIST 数据集，基于复现的隐私保护技术方案，利用代码对数据集中的图像进行处理，生成带有不同隐私保护技术方案标签的图像，作为新的数据集。

4）搭建基于 CNN 卷积神经网络的图像分类深度学习模型，划分数据集为训练集、测试集，训练分类模型，测试模型性能，根据结果对模型调参、调优。

3．进度安排

第 1-3 周：查阅相关中、英文文献资料，确定设计方案，撰写开题报告；

第 4-5 周：完成论文开题工作，完成英文专业文献翻译任务；

第 6-12 周：完成系统方案设计、深度学习网络设计与实现，并撰写论文初稿；

第 13-14 周：修改完善毕业论文；

第 15 周：完成论文答辩工作。

4．阅读的参考文献（不少于15篇，其中近五年外文文献不少于3篇）

[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] 吕园园．基于深度学习算法的图像处理技术[J].长江信息通信,2023,36(12):71-73.

[3] 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.

[4] 刘颖，庞羽良，张伟东，等.基于主动学习的图像分类技术：现状与未来[J].电子学报,2023,51(10):2960-2984.

[5] 台建玮，李亚凯，贾晓启，等.语音识别系统对抗样本攻击及防御综述[J].信息安全学报,2022,7(05):51-64.

[6] 黄治伟，王颖杰，李俊杨.基于自然语言处理的涉恐信息挖掘技术[J].信息安全与通信保密,2023,(11):104-112.

[7] 邓国强．面向机器学习的隐私保护及其应用研究[D].桂林:桂林电子科技大学,2023.

[8] 刘洋.隐私保护机器学习系统下的关键安全技术研究[D].西安:西安电子科技大学,2022.

[9] 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.

[10] Saeed Adelipour，Mohammad Haeri. Privacy-Preserving Model Predictive Control Using Secure Multi-Party Computation. International Conference on Electrical Engineering (ICEE)，Tehran，2023[C]，915-919.

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

[12] Gilad-Bachrach R，Dowlin N，Laine K，et al. Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy. Proceedings of the 33rd

International Conference on Machine Learning，New York，2016[C]. NewYork：ACM,，2016.

[13] Tian Li，Anit Kumar Sahu，Ameet Talwalkar，et al. Federated Learning: Challenges, Methods, and Future Directions[J]. IEEE Signal Processing Magazine，2020，37(3)：50-60.

[14] D. Zhao，P. Zhang，J. Xiang，et al. NegDL: Privacy-Preserving Deep Learning Based on Negative Database. International Conference on Data Intelligence and Security (ICDIS)，Shenzhen，China，2022[C].

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

[16] 沈传年，徐彦婷，陈滢霞.隐私计算关键技术及研究展望[J].信息安全研究,2023,9(08):714-721.

[17] 何其健. 保护隐私的深度学习方案研究[D]. 合肥:中国科学技术大学，2019.

[18] 魏立斐，陈聪聪，张蕾，等.机器学习的安全问题及隐私保护[J].计算机研究与发展,2020,57(10):2066-2085.

5．指导教师意见

指导教师（签名）：_____

年　　月　　日