北 京 邮 电 大 学

本科毕业设计（论文）开题报告

学院	网络空间安全学院	专业	网络空间安全专业	班级	2019211806
学生姓名	卢亭松	学号	2019212443	班内序号	11
指导教师姓名	陆月明	所在单位	网络空间安全学院	职称	教授
设计（论文）题目	（中文）一种基于 FATE 框架的联邦学习方法设计与实现				
	（英文）Design and Implementation of a Federated-Learning Method Based on FATE Framework				

毕业设计（论文）开题报告内容：

一、选题背景与意义

从 1955 年达特茅斯会议开始，人工智能经过两起两落的发展，迎来了第三个高峰期。越来越多的工程与科研实践让我们看到了人工智能迸发出的巨大潜力，也更加憧憬人工智能技术可以在自动驾驶、医疗、金融等更多、更复杂、更前沿的领域施展拳脚。但是，真实的情况却让人失望：除了有限的几个行业，更多领域存在着数据有限且质量较差的问题，不足以支撑人工智能技术的实现；并且，在某些领域，即使动用很多人力来进行数据标注，数据量也依然不够，这是我们面临的现实。

与此同时，数据源之间存在着难以打破的壁垒，在大多数行业中，数据是以孤岛的形式存在的，由于行业竞争、隐私安全、行政手续复杂等问题，即使是在同一个公司的不同部门之间实现数据整合也面临着重重阻力；另一方面，随着大数据的进一步发展，重视数据隐私和安全已经成为了世界性的趋势，新的隐私保护法规的建立在不同程度上对人工智能传统的数据处理模式也提出了新的挑战。因此，想要将分散在各地、各个机构的数据进行整合是十分困难且成本巨大的。

如何在满足数据隐私、安全和监管要求的前提下，设计一个机器学习框架，让人工智能系统能够更加高效、准确地共同使用各自的数据，则是联邦学习希望解决的问题。联邦学习作为未来 AI 发展的底层技术，依靠安全可信的数据保护措施连接数据孤岛的模式，将在保障隐私信息及数据安全的前提下加速人工智能技术的创新发展、促进全社会智能化水平提升，十分具有研究的意义。

二、研究内容和拟解决的主要问题

2.1 研究的基本内容

为了准确地了解对比现有联邦学习架构与传统 AI 架构在性能、准确度等方面的差异，研究将基于开源项目 FATE (Federated Al Technology Enabler)，搭建 FATE 联邦学习集群，在联邦场景实现两

种以上主流机器学习场景（计算机视觉、自然语言处理等）的样例算法，与传统机器学习进行对比，从性能（通信损耗时间、计算损耗时间等）、准确率（准确率、召回率、F1-score 等）、安全性测试（投毒攻击测试、对抗攻击测试等）等层面进行分析。

研究内容 1：联邦学习的定义、规范与价值机制；联邦学习在不同数据场景下的联邦方式；现有的联邦学习开源实现；联邦学习面临的安全问题以及针对性攻击手段。

联邦学习的定义为：在进行机器学习的过程中，各参与方可借助其他方数据进行联合建模。各方无需共享数据资源，即数据不出本地的情况下，进行数据联合训练，建立共享的机器学习模型。

联邦学习系统需要保证：|联邦学习模型的效果-传统方法模型的效果|<有界正数。

联邦学习的价值机制：联邦学习技术基于“合作共赢”的价值机制，对于商业利益而言极具价值。在这样一个联邦机制下，各个参与者的身份和地位相同，而联邦系统帮助大家建立了“共同富裕”的策略，能够带动跨领域的企业级数据合作、催生基于联合建模的新业态和模式、降低技术提升成本和促进创新技术发展。

联邦学习在不同场景下的联邦方式：

1. 横向联邦学习：在两个数据集的用户特征重叠较多而用户重叠较少的情况下，将数据集按照横向（即用户维度）切分，并取出双方用户特征相同而用户不完全相同的那部分数据进行训练。这种方法叫做横向联邦学习。
2. 纵向联邦学习：在两个数据集的用户重叠较多而用户特征重叠较少的情况下，把数据集按照纵向（即特征维度）切分，并取出双方相同而用户特征不完全相同的那部分数据进行训练。这种方法叫做纵向联邦学习。
3. 联邦迁移学习：在两个数据集的用户与用户特征重叠都较少的情况下，不对数据进行切分，而可以利用迁移学习来克服数据或标签不足的情况。这种方法叫做联邦迁移学习。

现有的联邦学习开源实现：目前业界中主要的联邦学习框架有 FATE、TensorFlow Federated、PaddleFL、Pysyft 等。

联邦学习面临的安全威胁以及其攻击手段：

1. 投毒攻击：投毒攻击主要是指在训练或再训练过程中，恶意的参与者通过攻击训练数据集来操纵机器学习模型的预测。联邦学习中，攻击者有两种方式进行投毒攻击：数据投毒和模型投毒。数据投毒是指攻击者通过对训练集中的样本进行污染，如添加错误的标签或有偏差的数据，降低数据的质量，从而影响最后训练出来的模型，破坏其可用性或完整性；而模型投毒不同于数据投毒，攻击者不直接对训练数据进行操作，而是发送错误的参数或损坏的模型

来破坏全局聚合期间的学习过程，比如控制某些参与方 U_i 传给服务器的更新参数 δ_i，从而影响整个学习模型参数的变化方向，减慢模型的收敛速度，甚至破坏整体模型的正确性，严重影响模型的性能。

2. 对抗攻击：对抗攻击是指恶意构造输入样本，导致模型以高置信度输出错误结果。从攻击环境来说，对抗攻击可以分为黑盒攻击和白盒攻击。若知道机器学习模型中的参数与内部结构，攻击者可以把所需的干扰看作一个优化问题计算出来。这种情况下的对抗攻击属于白盒攻击。而另一种常见的情境下，攻击者不知道任何模型的信息，只能跟模型互动，给模型提供输入然后观察它的输出，这种情形下的对抗攻击属于黑盒攻击。对抗攻击还可以根据攻击目的分为目标攻击和非目标攻击。根据干扰的强度大小分为无穷范数攻击、二范数攻击和零范数攻击等。对抗攻击可以帮助恶意软件逃避检测，生成投毒样本，已经被攻击者广泛应用于图像分类、语义分割、机器识别以及图结构等多个领域，成为系统破坏者的一个有力攻击武器。
3. 隐私泄露：联邦学习方式允许参与方在本地进行数据训练，各参与方之间是独立进行的，其他实体无法直接获取本地数据，可以保证一定的隐私安全，但这种安全并不是绝对安全，仍存在隐私泄露的风险。比如恶意的参与方可以从共享的参数中推理出其他参与方的敏感信息。恶意的服务器可以识别更新的参数的来源，甚至进一步通过参与方多次反馈的参数推测参与方的数据集信息，这可能造成参与方的隐私泄露。

支撑指标点：☑2.3 ☐3.3 ☐4.1 ☐4.2 ☐4.3 ☐10.1 ☐10.2 ☐10.3 ☑12.1 ☑12.2

毕业要求指标点 2.3：针对已建立的网络空间安全领域复杂工程问题的抽象模型，通过文献检索与资料查询获取相关知识，分析论证模型的合理性，获得有效结论。

毕业要求指标点 12.1：能够认识不断探索和学习的必要性，具有自主学习和掌握自主学习的方法，具有拓展与更新知识的能力。

毕业要求指标点 12.2：具有终身学习的知识基础和意识，能够针对个人或职业发展需要，采用合适的方法，自主学习，适应社会发展。

研究内容 2：在计算机视觉领域选择基于 CNN 的 mnist 手写数字识别作为样例算法；在自然语言处理领域选择基于 CNN 的中文主题分类作为样例算法；对其分别进行传统机器学习实现与联邦机器学习实现。

基于 CNN 的 mnist 手写数字识别：MNIST 数据集是代表标准和技术数据集的改良研究所的缩

写，是一个包含 60,000 张 0 到 9 之间的手写单个数字的 60,000 个小正方形 28×28 像素灰度图像的数据集。任务是将给定的手写数字图像分类为 10 个类别之一，代表从 0 到 9 的整数值，包括 0 到 9。

基于 CNN 的中文情感分析：使用数据集为清华 NLP 组提供的 THUCNews 新闻文本分类数据集。其中包含体育, 财经, 房产, 家居, 教育, 科技, 时尚, 时政, 游戏, 娱乐 10 个分类。数据格式为带有标注的文本串。

支撑指标点：☐2.3 ☑3.3 ☑4.1 ☑4.2 ☐4.3 ☑10.1 ☐10.2 ☐10.3 ☐12.1 ☐12.2

毕业要求指标点 3.3：综合考虑各种工程因素，给出解决方案，能够利用软件模块，进行网络空间安全领域系统的整体设计与开发。

毕业要求指标点 4.1：能够针对网络空间安全领域的复杂工程问题明确其研究目标，根据目标研究确定需要的实验数据及技术路线，完成实验方案的设计。

毕业要求指标点 4.2：能够选择合适的技术手段，构建实验系统，安全地开展实验，正确采集、整理实验数据。

毕业要求指标点 10.1：具有良好的表达能力，能够就专业问题进行清晰的书面和口头表达，并能与同行进行有效沟通和交流。

研究内容 3：与传统机器学习进行对比，从准确率（准确率、召回率、F1-score 等）、性能（通信损耗时间、计算损耗时间等）、安全性测试（投毒攻击测试、对抗攻击测试等）等层面进行分析。

准确率：依据准确率、召回率等标准对传统机器学习实现与联邦机器学习实现分别进行分析与评估，从多个角度综合量化考察联邦学习与传统机器学习在准确率方面的差异，研究改进和调整办法。为了排除检测数据集不平衡性的干扰，评价异常检测主要使用以下指标：准确率、召回率、f1-score。

性能、安全性：收集不同场景、不同算法下联邦学习在性能、安全性等方面的损耗。统计联邦学习的通信时间损耗、同数据量下收敛速度对比；对比投毒攻击、对抗攻击、隐私泄露等场景下联邦学习与传统机器学习的鲁棒性。

支撑指标点：☐2.3 ☐3.3 ☐4.1 ☐4.2 ☑4.3 ☐10.1 ☑10.2 ☑10.3 ☐12.1 ☐12.2

毕业要求指标点 4.3：能够对实验结果进行分析和解释，通过信息综合得到合理有效的结论。

毕业要求指标点 10.2：熟练掌握一门外语，具备一定的国际视野，能够在跨文化背景下进行沟通和交流。

毕业要求指标点 10.3：能够就网络空间安全领域复杂工程问题与业界同行及社会公众进行有效沟通和交流，撰写报告和设计文稿、陈述发言等。

2.2 拟解决的问题

由于联邦学习引入了安全多方计算、同态加密等技术，必然也会引入一定的代价，例如通信、计算等性能上的损耗、不同场景下数据异构对准确率的影响等。目前关于这些代价仍没有较为细致的对比数据，而准确地了解对比现有联邦学习架构与传统 AI 架构在性能、准确度、安全性等方面的差异，能够帮助我们进一步优化改进联邦学习。

三、研究方法及措施

1. 学习机器学习、隐私保护相关的知识，对研究背景进行调研。

2. 基于 FATE 搭建联邦学习集群实验环境，同时搭建传统机器学习场景，并进行编程实验。

3. 通过控制变量对比的方法分析二者在性能、准确度、安全性等方面的差异，找到现有联邦学习需要优化的方向

4. 通过阅读源码、理解联邦学习实现架构等方式找到优化点并给出不同使用场景下、不同算法下的使用建议

四、研究工作的步骤与进度：

秋季学期 17-18 周：学习机器学习、隐私保护相关的知识，对研究背景进行调研。查找并阅读联邦学习、机器学习隐私保护的相关论文，为后续的实验以及论文撰写打下基础。

春季学期 1-2 周：开始 FATE 集群的搭建，完成开题报告的撰写。

春季学期 3-4 周：构建论文框架，完成 FATE 集群搭建。

春季学期 5-6 周：完成论文前两章的撰写，确定两种以上的机器学习场景及代表性问题，设计相应的联邦机器学习算法。

春季学期 7 周：总结上一阶段的工作，完成中期检查。

春季学期 8-9 周：完成两种以上的联邦机器学习算法的传统实现与联邦环境实现，能够在 FATE 集群上进行训练。

春季学期 10-11 周：在 FATE 平台上完成预测与评估，对比分析算法在集中式环境与联邦环境的差异，从准确率，安全性，通信效率等进行分析。

春季学期 12-13 周：根据上述的实验结果，完成论文的撰写。

春季学期 14 周：完成论文，整理本毕设课题的全部成果。

五、主要参考文献：

[1] 谭作文, 张连福. 机器学习隐私保护研究综述. 软件学报, 2020, 31(7): 2127-2156.

[2] Abreha HG, Hayajneh M, Serhani MA. Federated Learning in Edge Computing: A Systematic Survey.

Sensors. 2022; 22(2):450. https://doi.org/10.3390/s22020450

[3] K. M. Ahmed, A. Imteaj and M. H. Amini, “Federated Deep Learning for Heterogeneous Edge Computing,“ 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA), 2021, pp. 1146-1152, doi: 10.1109/ICMLA52953.2021.00187.

[4] Y. Liu, J. J. Q. Yu, J. Kang, D. Niyato and S. Zhang, “Privacy-Preserving Traffic Flow Prediction: A Federated Learning Approach,“ in IEEE Internet of Things Journal, vol. 7, no. 8, pp. 7751-7763, Aug. 2020, doi: 10.1109/JIOT.2020.2991401.

[5] Liu JC, Goetz J, Sen S, Tewari A. Learning From Others Without Sacrificing Privacy: Simulation Comparing Centralized and Federated Machine Learning on Mobile Health Data. JMIR Mhealth Uhealth, 2021;9(3):e23728

[6] 胡健龙. 联邦学习在车联网数据共享与保护技术中的研究[D]. 电子科技大学,2022.DOI:10.27005/d.cnki.gdzku.2022.004716.

[7] M. Nasr, R. Shokri and A. Houmansadr, Comprehensive Privacy Analysis of Deep Learning: Passive and Active White-box Inference Attacks against Centralized and Federated Learning. In Proceedings of 2019 IEEE Symposium on Security and Privacy (SP), 2019, pp. 739-753, doi: 10.1109/SP.2019.00065.

[8] 王慧超. 机器学习中的数据隐私保护研究[D]. 中国科学技术大学,2021.DOI:10.27517/d.cnki.gzkju.2021.001722.

[9]Wang, X.; Wang, J.; Ma, X.; Wen, C. A Differential Privacy Strategy Based on Local Features of Non-Gaussian Noise in Federated Learning. Sensors 2022,22(2424). https://doi.org/10.3390/s22072424

[10] 师兆森. 联邦学习中的隐私保护技术研究. 电子科技大学,2022. DOI:10.27005/d.cnki.gdzku.2022.000987.

[11] 王健宗. 联邦学习算法综述. 大数据, 2020, 6(6): 2020055-1- doi: 10.11959/j.issn.2096-0271.2020055.

[12] 邱鑫源,叶泽聪,崔翛龙,高志强. 联邦学习通信开销研究综述. 计算机应用, 2022, DOI: 10.11772/j.issn.1001-9081.2021020232

[13] 周俊,方国英,吴楠. 联邦学习安全与隐私保护研究综述. 西华大学学报, 2020, doi：10.12198/j.issn.1673－159X.3607

[14] 周传鑫,孙奕,汪德刚,葛桦玮. 联邦学习研究综述. 2021. DOI: 10.11959/j.issn.2096−109x.2021056

允许进入论文撰写环节：是 □ 否 □	指导教师	

日期	年　月　日	签字	