浙 江 大 學

硕士专业学位论文中期进展报告

中文论文题目： 面向差分隐私机器学习算法的研究与评估

英文论文题目： Research and Evaluation System Development for Differentially Private Machine Algorithms

姓名：

指导教师：

专业学位类别：机械或电子信息

专业学位领域：工业设计工程或软件工程或人工智能

所在学院：软件学院

提交日期 2023 年 7 月 10 日

目录

1 已完成的论文工作及取得的阶段性成果

1.1 总体工作进展

当前毕业论文的写作进展非常顺利。该论文的题目是《面向差分隐私机器学习算法的研究与评估系统实现》。研究的目标是对现有差分隐私机器学习算法进行综合评测和分析，探索有利于提升差分隐私机器学习可用性的改进方法。同时设计和实现一个差分隐私算法测试框架，以支持对 DPML 算法的可用性和隐私性，该框架将集成多个差分隐私训练算法、数据集及模型结构，可以让使用者方便的调用并训练得到满足差分隐私的模型，也可以让研究者对实验设置和新算法进行基准测试。

在论文的初期阶段，我进行了广泛的文献调研，研究了差分隐私的基本原理和顶会中提出的差分隐私机器学习算法。我了解了差分隐私的概念、定义和应用场景，并对当前的研究进展和挑战进行了深入分析，并提出了初步的分类体系。接下来，我复现了各个文章中的算法代码，并调试代码以确保算法效果。目前正在搭建算法评测框架，将不同改进思路的差分隐私机器学习算法融合进入算法评测框架，并开始进行大规模实验，目前实验进度过半。与此同时，我在对评测系统进行功能完善和寻找分析实验结果的方法。我计划进一步框架的代码架构，使其具有高可扩展性，并对实验结果进行详细的解读和讨论。我也在撰写毕业论文，包括引言、相关工作等。

总体而言，当前毕业论文的写作进展符合预期。通过深入研究和系统实现，我相信该论文能如期完成，并将为差分隐私机器学习算法的研究和应用提供有价值的贡献。

1.2 各章节完成情况及取得的阶段性成果

摘要 未开始撰写。

第一章，绪论 已撰写完成。内容包括课题背景、意义、国内外研究现状。

第二章，相关理论基础 已撰写完成。内容包括机器学习训练流程，差分隐私，差分隐私机器学习和成员推理攻击相关原理。

第三章，差分隐私算法的分类研究 已撰写完成 50%。目前完成了初步分类体系。

第四章，评测框架设计与实现 已撰写完成 50%。

第五章，算法评测结果及分析　未开始撰写。还在做实验，对目前已完成的实验进行了分析，总结了部分实验发现，未开始撰写论文

第六章，总结与展望　未开始撰写。

2 下一步论文工作安排

2.1 重难点分析及拟采取的解决措施

难点 1. 如何实现评测框架的高可扩展性？ 不同差分隐私机器学习算法的实现差异巨大，在机器学习各个阶段的修改都有，比较难抽象出统一的接口，以实现后续对新算法的扩展。

解决措施 借鉴现有类似工作的实现，学习并应用软件工程中的设计模式以增加代码的可扩展性。仔细分析现有算法，将算法分为多个阶段，使用统一的接口进行交互，将算法细节隐藏在模块内部。

难点 2. 差分隐私训练算法时间开销大 差分隐私机器学习因为逐样本梯度裁剪，导致训练时间开销大，可能会导致项目延期。

解决措施 使用并行训练，增加实验机器数量，编写任务提交脚本，提高机器使用效率。

3　目前取得的成果

正在进行中