**重庆邮电大学硕士学位档案（一）**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 论文答辩情况 | | | | |
| 答辩日期：2025年5月18日 | | | 地点：江南分院104 | |
| 答辩记录：（包括提问及回答要点）  问题1、研究现状存在的挑战要分析是什么？  答：现有纵向联邦学习方法过度依赖大量标记和对齐数据，而实际应用中往往面临高昂的标注成本和领域专家稀缺的情况。尤其当只有一方拥有标签信息时，联合模型训练的复杂度显著提高。同时，不同参与方的数据分布存在显著差异，导致模型难以有效学习。尽管RSCFed采用加权距离聚合、FedMatch通过客户端间一致性来应对异质性，但在复杂应用场景中的效果仍有待提升。  问题2、在VFPU框架中，如何保证数据隐私不被泄露？  答：在VFPU框架中，采取了多层次的隐私保护机制：首先，参与方仅共享中间计算结果（如梯度、预测值）而非原始特征数据；其次，对于共享的中间结果，我们采用了安全聚合协议，确保任何参与方都无法从接收到的信息中推断出其他方的私有信息；此外，对于特别敏感的场景，还设计了差分隐私机制，通过向共享值添加校准噪声来进一步增强隐私保护。实验表明，这些机制能在保障模型性能的同时，有效防止隐私泄露风险。  问题3、FedPSG-PUM的核心创新点是什么？与现有方法相比有何优势？  答：FedPSG-PUM的核心创新在于将联邦半监督学习与生成模型技术结合，提出分阶段处理策略：跨方特征相关性分析：通过隐私保护的Spearman秩相关计算，筛选高相关性特征，指导后续生成策略。联邦生成模型填补：对低相关性特征，采用VF-GAIN等联邦生成模型，学习多方数据分布以填补缺失值。  问题4、FedPSG-PUM的局限性及未来方向？  答：联邦半监督学习需多轮迭代，VF-GAIN的对抗训练增加通信成本；τ和α需通过验证集调优，可能影响泛化性；引入元学习或贝叶斯优化自动选择τ和α。 适配图像、文本等多模态数据，提升方法普适性。  问题5、算法中正例比例估计环节对最终性能有多大影响？  答：正例比例估计是VFPU算法的关键环节，对最终性能有显著影响。我们的实验表明，当估计误差在±10%范围内时，模型性能下降相对可控；但若误差超过15%，模型性能将明显下降。  问题6、如何评估模型在联邦环境下的通信效率？  答：从三个维度评估联邦环境下的通信效率：首先，通信轮次数量，即达到模型收敛需要的总通信次数；其次，单轮通信量大小，即每次交互传输的数据量；第三，整体时间消耗，包括通信延迟和计算时间。实验结果表明，与基准方法相比，VFPU在保持相似精度的情况下，通信轮次减少了约20%，整体训练时间缩短了25-30%。 | | | | |
| 答辩秘书签名： 2025年5月18日 | | | | |
| 答 辩 委 员 会 成 员 | | | | |
| 成员 | 姓 名 | 职 称 | | 工 作 单 位 |
| 主席 | 廖晓峰 | 教授 | | 重庆大学 |
| 委员 | 龙林波 | 副教授 | | 重庆邮电大学 |
| 委员 | 尹学辉 | 副教授 | | 重庆邮电大学 |
| 委员 | 郑申海 | 讲师（高校） | | 重庆邮电大学 |
| 委员 | 刘歆 | 副教授 | | 重庆邮电大学 |
| 秘书 | 夏士超 | 讲师（高校） | | 重庆邮电大学 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 学位论文答辩委员会决议 | | | | | | |
| 建议授予学位  投 票 结 果 | 实到委员数： | | 同意人数： | | 不同意人数： | |
| 答辩成绩 | 优： | 良： | 中： | | 及格： |
| 答辩委员会决议：  吕九峦同学的硕士论文《联邦半监督学习方法及其在样本生成中的应用研究》，以数据隐私保护需求日益增加与分布式数据环境广泛存在，联邦学习面临标记数据稀缺、数据分布异构以及参与方对齐样本不足等挑战为背景，旨在解决纵向联邦学习场景中如何有效利用未标记数据和非对齐样本，同时保障数据隐私安全的关键问题，其选题具有重要的理论研究意义与工程应用价值。  该论文针对联邦学习中标记数据稀缺、数据分布异构以及参与方对齐样本不足等问题，提出了基于正样本和未标记数据的纵向联邦学习方法，以及结合联邦半监督学习与数据生成的样本生成方法。通过大量对比实验，验证了所提方法在保护数据隐私的前提下能够有效利用未标记数据提升模型性能，即使在样本缺失率较高情况下依然保持显著优势。  该生的论文文字通顺，结构完整，阐述清楚，工作量饱满；其论文工作反映出该生掌握了扎实的基础理论和专业知识，具有较强的独立从事科学研究工作的能力。答辩时论述清楚，思路清晰，能准确回答问题。  综上所述，答辩委员会认为该论文达到硕士学位论文水平，投票结果为：（5）票通过，（0）票不通过，（0）票弃权（全票为5票）。建议授予该生工学硕士学位。 | | | | | | |
| 答辩委员会主席： 2025年5月18日 | | | | | | |

**重庆邮电大学硕士学位档案（二）**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 答辩委员会成员 | | | | | | | | |
| 成员 | 姓 名 | | 职 称 | | 工 作 单 位 | | | |
| 主席 | 廖晓峰 | | 教授 | | 重庆大学 | | | |
| 委员 | 龙林波 | | 副教授 | | 重庆邮电大学 | | | |
| 委员 | 尹学辉 | | 副教授 | | 重庆邮电大学 | | | |
| 委员 | 郑申海 | | 讲师（高校） | | 重庆邮电大学 | | | |
| 委员 | 刘歆 | | 副教授 | | 重庆邮电大学 | | | |
| 秘书 | 夏士超 | | 讲师（高校） | | 重庆邮电大学 | | | |
| 建议授予学位  投 票 结 果 | | 实到委员数： | | | 同意人数： | | 不同意人数： | |
| 答辩成绩 | | 优： | 良： | 中： | | 及格： |
| 答辩委员会决议：  吕九峦同学的硕士论文《联邦半监督学习方法及其在样本生成中的应用研究》，以数据隐私保护需求日益增加与分布式数据环境广泛存在，联邦学习面临标记数据稀缺、数据分布异构以及参与方对齐样本不足等挑战为背景，旨在解决纵向联邦学习场景中如何有效利用未标记数据和非对齐样本，同时保障数据隐私安全的关键问题，其选题具有重要的理论研究意义与工程应用价值。  该论文针对联邦学习中标记数据稀缺、数据分布异构以及参与方对齐样本不足等问题，提出了基于正样本和未标记数据的纵向联邦学习方法，以及结合联邦半监督学习与数据生成的样本生成方法。通过大量对比实验，验证了所提方法在保护数据隐私的前提下能够有效利用未标记数据提升模型性能，即使在样本缺失率较高情况下依然保持显著优势。  该生的论文文字通顺，结构完整，阐述清楚，工作量饱满；其论文工作反映出该生掌握了扎实的基础理论和专业知识，具有较强的独立从事科学研究工作的能力。答辩时论述清楚，思路清晰，能准确回答问题。  综上所述，答辩委员会认为该论文达到硕士学位论文水平，投票结果为：（5）票通过，（0）票不通过，（0）票弃权（全票为5票）。建议授予该生工学硕士学位。 | | | | | | | | |
| 答辩委员会主席： 2025年5月18日 | | | | | | | | |

**重庆邮电大学硕士研究生学籍表**

|  |
| --- |
| 答辩委员会决议（包括评语、投票或讨论结果）：  吕九峦同学的硕士论文《联邦半监督学习方法及其在样本生成中的应用研究》，以数据隐私保护需求日益增加与分布式数据环境广泛存在，联邦学习面临标记数据稀缺、数据分布异构以及参与方对齐样本不足等挑战为背景，旨在解决纵向联邦学习场景中如何有效利用未标记数据和非对齐样本，同时保障数据隐私安全的关键问题，其选题具有重要的理论研究意义与工程应用价值。  该论文针对联邦学习中标记数据稀缺、数据分布异构以及参与方对齐样本不足等问题，提出了基于正样本和未标记数据的纵向联邦学习方法，以及结合联邦半监督学习与数据生成的样本生成方法。通过大量对比实验，验证了所提方法在保护数据隐私的前提下能够有效利用未标记数据提升模型性能，即使在样本缺失率较高情况下依然保持显著优势。  该生的论文文字通顺，结构完整，阐述清楚，工作量饱满；其论文工作反映出该生掌握了扎实的基础理论和专业知识，具有较强的独立从事科学研究工作的能力。答辩时论述清楚，思路清晰，能准确回答问题。  综上所述，答辩委员会认为该论文达到硕士学位论文水平，投票结果为：（5）票通过，（0）票不通过，（0）票弃权（全票为5票）。建议授予该生工学硕士学位。 |
| 主席签字： 2025年5月18日 |