重庆邮电大学研究生学位论文修改情况表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 学号 | S221201028 | 姓名 | 吕九峦 | 导师姓名 | 韦庆杰 |
| 专业/领域 | 软件工程 | | | | |
| 学位论文题目 | 联邦半监督学习方法及其在样本生成中的应用研究 | | | | |
| 学位论文修改前存在的主要问题（论文评阅专家提出）  【专家1总体意见】：论文选题聚焦于联邦学习与半监督学习结合的新型研究方向，兼顾理论创新与实际应用，尤其针对当前在数据隐私保护背景下分布式学习场景中样本稀缺与样本对齐难题提出了实际解决方案。  选题具有较强的前沿性与研究价值，切合智能制造、金融风控、医疗健康等实际需求，体现出较高的理论意义和工程应用潜力。  论文对国内外联邦半监督学习及表格数据生成方法进行了较为系统的梳理，并对不同方法从算法特性、隐私保护、性能表现等方面进行了对比分析。作者在联邦学习、半监督学习、PU学习以及生成模型（如TabDDPM, VF-GAIN等）方面展现出扎实的理论基础，能结合多领域方法构建联邦环境下的半监督模型，理论逻辑清晰、推导合理、结构完整。  论文提出了VFPU方法实现了在无共享标签与无负样本的前提下，基于PU策略构建多方协作机制；提出的FedPSG-PUM创新性地引入特征相关性分析，分别对高/低相关性特征采用伪标签生成与表格数据合成，实现缺失样本的“重构”。这两种方法体现出良好的实际效果。  论文整体结构严谨，算法描述清晰，实验设计合理，结果分析充分，展现出作者具备较强的问题建模、算法设计、实验实现与效果分析能力。  论文整体结构严谨，算法描述清晰，实验设计合理，结果分析充分，展现出作者具备较强的问题建模、算法设计、实验实现与效果分析能力。但仍需要注意细节方面的把控。  【专家1修改意见】：1. 建议图中统一使用中文注释，避免中英文混排，以提升整体排版的一致性与专业性。  2. 请优化正文排版结构，避免出现个别字符或词语单独占据一行的情况，以提升阅读体验和版面美观度。  3. 请仔细检查文章参考文献的引用位置，如正文中的[10],[14]和[22],[23]等  4. 建议统一全篇公式后的标点符号使用规范。当前第二章的公式后无标点，第三章有标点，第四章则又无标点，存在不一致问题。  5. 请仔细检查全文中英文术语的全称与缩写使用规范。缩写应在首次出现时给出全称与缩写，如“纵向联邦学习（Vertical Federated Learning, VFL）”，后续再出现时仅使用缩写“VFL”，避免重复说明。  6. 请注意中英文括号使用的一致性规范。例如，第57页中“RMSE”等使用了英文括号，而其他部分使用中文括号，建议全文统一。  【专家2总体意见】：论文针对联邦学习环境下未标记数据稀缺和对齐样本不足的问题进行研究，提出了创新的联邦半监督学习方法及其样本生成策略。论文选题合理，具有理论价值与现实意义。  论文首先分析了联邦学习环境中未标记数据难以有效利用的问题，针对未标记数据稀缺的情况，提出了基于正样本和未标记数据的纵向联邦学习方法，通过反复抽样和迭代训练，有效识别可靠的正样本，解决了未标记数据的有效利用问题，提升了联邦半监督学习模型的泛化性能；然后分析了纵向联邦学习中参与方对齐样本数量有限的问题，针对这一问题，提出了融合联邦半监督学习与数据生成技术的样本生成方法，通过特征关联性分析和数据合成技术，有效解决了非对齐样本的数据缺失问题，提升了模型对样本缺失的鲁棒性与泛化能力。  综上所述，论文从联邦半监督学习及数据生成的角度切入，针对联邦学习中未标记数据利用和对齐样本不足问题进行了研究，论文逻辑结构合理，综述清晰，基础理论知识与专业知识扎实，研究工作量较为饱满，对所提模型有相应实验验证，整体而言，达到了硕士毕业论文的基本要求。  【专家2修改意见】：（1）论文指出在纵向联邦学习的实际应用中，存在一种特殊的正例-未标记（PU）学习场景，即“目标方仅持有正类样本，而其他参与方仅拥有未标记样本”。这一设定具有一定现实意义，但目前描述尚不具体，建议补充具体的生活或行业应用实例，以增强说服力。例如，可考虑医疗、金融等场景中目标变量难以标注或仅部分正类样本可获取的典型案例，以论证该问题的实际存在性和研究价值。毕竟在多数现实任务中，数据集通常至少包含部分负类或混合样本，即使类别不平衡。  （2）第1.1节“研究背景与意义”应聚焦于研究问题的背景、挑战与现实意义，避免在此处介绍作者提出的具体方法或技术路线，以保持章节内容的逻辑清晰性与结构规范性。  （3）论文中提出：“为更好地应对 A 方特征与目标特征间低相关性（即 ???? ≤ ??）给半监督学习带来的预测性能下降问题，FedPSG-PUM 引入了生成模型技术……”，该部分描述较为模糊，建议进一步明确具体使用了哪类生成模型，如何建模低相关性特征的分布规律，以及数据合成的具体过程和策略。应补充算法流程图或伪代码，并结合实验示例，增强方法描述的完整性与可复现性。  （4）论文第3章介绍了多方纵向联邦半监督学习方法，第4章则介绍了基于联邦半监督学习的参与方样本生成方法。然而，两章之间的承接关系与逻辑关联未在章节导语中清晰说明。建议在第4章开头增加对前一章工作的简要总结以及本章与前一章的关联关系，从而理清两部分工作的衔接逻辑。 | | | | | |
| 学位论文内容所做的具体修改和充实情况  已根据两位专家的修改意见对论文进行了认真修改和完善，现将具体修改情况如下：  对专家1意见的修改  1. 图中注释统一为中文  已对论文中所有图表进行全面检查，将混排的英文注释统一替换为中文注释。特别是第3章和第4章中的模型架构图、算法流程图等关键图表，现已确保注释语言一致，提升了整体排版的专业性。  2. 优化正文排版结构  已全面检查论文排版，对存在个别字符或词语单独占据一行的情况进行了调整。主要通过重新调整段落格式、适当调整句式结构，确保了文字排版的美观度和阅读流畅性。  3. 检查参考文献引用位置  已仔细检查并修正了文中所有参考文献的引用位置，特别是[10]、[14]和[22]、[23]等引用在正文中的准确位置，确保引用格式规范统一。  4. 统一公式后的标点符号  已对全文公式后的标点符号使用进行了统一规范处理。根据学术规范，选择了在所有章节的公式后统一不添加标点符号（逗号或句号），消除了第二、三、四章之间的不一致问题。  5. 规范英文术语的全称与缩写  已全面检查英文术语的使用规范，确保每个专业术语缩写在首次出现时给出全称与缩写，如"纵向联邦学习（Vertical Federated Learning, VFL）“，后续再出现时仅使用缩写"VFL”。已修改的术语包括FL、VFL、SSFL、PU learning等多个关键概念。  6. 统一中英文括号使用  已统一全文中英文括号的使用规范，包括第57页中"RMSE"等技术指标的标注，将所有括号统一为中文括号，确保了全文格式的一致性。  对专家2意见的修改  1. 补充PU学习场景的具体实例  在第1.1节和第3.2节中，补充了纵向联邦学习中PU学习场景的具体实例  这些具体实例有力论证了该问题设定的现实存在性和研究价值。  2. 调整研究背景与意义章节内容  已调整第1.1节"研究背景与意义"内容，聚焦于研究问题的背景、挑战与现实意义的阐述，将原有对具体技术方案的介绍内容移至后续相关章节，确保了内容的逻辑清晰性与结构规范性。  3. 明确生成模型技术细节  在第4.3节中，已大幅增强了对FedPSG-PUM方法中生成模型技术的描述：  明确说明了采用的是基于扩散模型的TabDDPM生成技术以及VF-GAIN纵向联邦填补技术，详细阐述了其适用于表格数据的优势；  补充了低相关性特征建模的具体过程，包括特征相关性计算方法及阈值选择依据；  新增了数据合成的详细步骤与参数设置，包括模型的训练与采样过程；  增加了算法的详细伪代码及完整的流程图，提高了方法描述的完整性与可复现性。  4. 加强章节间的逻辑关联  明确阐述了第4章工作与第3章的承接关系：第3章解决了未标记数据有效利用问题，而第4章进一步解决了参与方对齐样本不足的挑战。这一调整使两章内容的逻辑关联更加清晰，增强了论文整体结构的连贯性。 | | | | | |
| 硕士生签字 日期 2025.05.13  导师签字 日期 2025.05.13 | | | | | |