

# 李淇

电话: 13037523298  
籍贯: 河南省商丘市

性别: 男  
政治面貌: 共青团员

邮箱: liqi2022@iie.ac.cn  
所在院校: 中国科学院信息工程研究所

## 项目经历

### 基于长尾分布的深度学习系统构建

- 数据增强: 针对尾部数据远远小于头部数据的问题, 使用颜色变换、重采样、随机裁剪、添加高斯噪声等数据增强技术增加尾部类的数量, 防止模型在训练时过度拟合头部类别的数据, 忽视尾部类别的数据;
- 模型训练: 为了解决类别不平衡的问题, 我们采用了以下方法来训练模型, 总体的目标是使模型更关注数据量少的类别, 从而提高模型的性能, 减少长尾效应的影响:
  1. 类别损失加权: 对不同类别的损失进行加权, 确保模型更关注数据量较少的类别, 从而平衡不同类别之间的影响;
  2. 固定原型中心: 我们使用手工固定的原型中心作为类别的参考点, 以便更好地区分不同类别之间的特征;
  3. 知识迁移的方法: 从头部类中学习通用知识, 然后迁移到尾部少样本类别中。
- 实现分析: 为了验证模型对尾部类别的学习效果, 我们使用均衡的测试集来评估模型的性能。我们将不同算法进行对比, 分析不同算法在各个类别上的准确率和召回率。结果表明,
  1. 类别损失加权的方法理论上可以通过给头部类别更低的权重、给尾部类别更高的权重, 从而反向抵消长尾效应。然而, 由于对难样本的挖掘不充分, 该方法的性能较差;
  2. 知识迁移的方法虽然性能最优, 但需要大量精力来设计知识转移所需的特殊模块, 由于不同任务之间的特征分布和数据分布可能存在差异, 导致知识迁移的效果不佳, 所以不能有效地泛化到多个任务。
  3. 固定原型中心的方法借助于神经网络结构理论, 理论上可以达到性能的上界, 并且不需要额外的参数或复杂的优化机制。但是由于尾部类缺少足够的语义信息, 导致模型的性能次优。

### 基于联邦学习的模型分类系统构建

- 数据非独立同分布划分: 使用迪利克雷概率分布模拟现实场景中的数据数量分布倾斜和标签分布倾斜两种情况, 首先将数据集按照类别和数量进行划分, 使得不同客户端的数据具有不同的数量和类别, 然后对数据进行清洗和归一化, 提高数据的质量和可用性。
- 模型并行训练: 首先服务器给每个客户端下发初始的卷积网络模型, 其次客户端基于本地数据进行训练并将收敛后的本地模型上传到服务器; 然后服务器基于聚合机制如平均、加权平均、梯度裁剪等机制聚合上传的本地模型; 重复步骤1-3, 直到模型收敛或达到预设的迭代次数。我们构建了12种经典的联邦学习算法, 主要分为本地模型纠偏和全局模型纠偏两大类。
- 性能对比: 在不同的数据分布情况下, 对算法的通信量、训练速度和准确率进行了对比。实验发现12个算法在独立同分布下都能有效的工作, 而随着分布越来越极端, 本地纠偏的方法如FedProx, FedDyn, Moon等会出现不收敛的现象; 全局纠偏的方法如FedBE、FedDF、FedGen等算法可以借助公共数据集和知识蒸馏等方法动态调整梯度的方向和大小, 有效的缓解数据非独立同分布的问题, 并且通信量、训练速度和准确率均优于其他算法;
- 可视化分析: 使用TSNE模型将分布在不同客户端的同一类的数据映射到同一块特征空间中, 以观察它们的分布情况。结果显示, 全局纠偏的方法虽然全局模型的性能更好但是存在特征偏移问题, 即不同客户端的数据特征无法很好地聚集在一起; 而本地纠偏的方法性能较差但是在训练时会让本地表征逐渐靠近全局知识来缓解特征偏移问题。

### 利用语义增强解决异构联邦学习中的特征倾斜问题

- ICME(CCFB)已录用

- 论文动机: 在联邦学习中, 数据的异质性是一个关键挑战, 显著影响了本地和全局模型的性能。之前的研究侧重于解决数据异质性问题, 但却忽视了特征偏移问题。我们引入了语义正则化项和自适应本地聚合, 以减轻本地知识不足的缺点, 并增强表征的语义信息。此外, 基于神经崩溃理论, 我们初始化了一个分类器的简单等角紧框架结构, 并在本地训练过程中保持其稳定性, 以实现不同客户端之间特征空间的对齐, 解决特征偏移问题。
- 实验分析: 我们的方法在12个评价指标中有11个达到了SOTA。为了进一步说明方法的有效性, 我们与FedAvg和最相关的方法FedETF进行了可视化分析, 结果表明我们的算法在分类效果上表现最优, 能够进一步保证类内方差较小、类间方差较大的分类效果。此外, 消融实验中的每个模块也侧面论证了我们方法的有效性。

## 荣誉奖项

本科阶段: 2017年英语竞赛二等奖; 2018年一等奖学金、三好学生; 2019年国家励志奖学金、优秀共青团员; 2020年国家奖学金; 2021年优秀毕业生等奖项  
博士阶段: 2023年中国科学院优秀共青团员; 2024年三好学生; 2024MM(CCFA)在投

## 自我评价

我具备较强的学习能力和自主学习意识, 热衷于探索新的知识和领域。在学习和工作中, 我总能保持乐观积极的心态, 不畏困难和挑战, 并且善于寻找和利用各种资源, 以找到最佳的解决方案。在团队合作中, 我总能与他人友好相处, 尊重他人的意见和建议, 并积极参与讨论和交流, 致力于为目标和发展做出突出的贡献。