

本次报告内容时间阶段	2023.6.3-2023.6.15	项目负责人	揭会顺
报告日期	2023.6.16	报告者	杨瑾、李宇飞、揭会顺
项目整体阶段	各模块的前期方法调研与现有方法测试与验证		
目前所处阶段	自动化特征提取模块进行现有方法测试与验证；模型训练任务调度算法设计和调度算法实现；模型选择、训练、部署的方法调研与测试		
具体进度	<p>自动化特征提取模块：</p> <p>基于遗传算法的自动化特征提取技术设计，基于遗传算法的自动化特征提取技术实现，数据集获取与方法验证。</p> <p>已经完成前期工作调研。目前正在公开数据集上测试调研中所涉及的 Baseline 方法。根据测试结果总结提炼特征工程任务的特点与适用性问题，为后续自动化特征提取算法框架设计与验证做好基础准备工作。主要包括：测试方法，测试数据，测试指标等。</p> <p>模型训练任务调度模块：</p> <p>已初步完成模型训练任务调度算法的设计。目前已确定模型调度算法的基本框架为基于元强化学习的调度方法。目前正在进行算法的代码实现工作。</p> <p>模型选择、训练、部署模块：</p> <p>对模型选择方法进行进一步调研和尝试，基于现有的模型训练和模型部署方法进行测试和验证。</p>		
预计结束时间	2023.6.15	实际结束时间	2023.6.15

本周项目进展	<p>自动化特征提取：本周完成了 Featuretools、Boruta、XGboost、tsfresh 以及 pycraret 等 5 个现有自动化特征提取工具使用方法的实践，并在公开数据集上进行了验证。测试指标包括:Precision、Recall、F1-score、Accuracy 以及 AUC；决策特征选择与否的中间属性包括：特征重要性、选择命中次数 (hits) 以及特征相关性等关键属性。同时特征选择过程也分为模型选择前和模型选择后两种模式。</p> <p>模型训练任务调度：本周已经完成了调度算法强化学习部分的实现。算法将使用近似策略优化算法，对调度算法中的行动者和批评者进行在线优化，从而优化模型所产生的策略。</p> <p>模型选择训练部署：测试搭建血缘图谱以建立问题到模型的关系映射，搭建集群环境完成现有模型部署和监控方法的测试。</p>
下周项目进度计划	<p>自动化特征提取：基于已完成的调研方法进行算法框架设计，尽可能多的考虑到不同因素对特征选择过程的影响。同时，基于上述测试流程，确定实验设计方法、数据集及测试指标，为后续算法验证做好准备。实践并测试现有基于遗传算法的特征选择算法框架，分析与项目需求的差异，并做出适应性调整，同时与下游工作做好对接。</p> <p>模型训练任务调度：完成调度算法强化学习部分的单元测试。</p> <p>模型选择训练部署：完成模型训练工作流定义到现有模型训练方法的测试。</p>
项目风险或故障	暂无项目风险故障

基于神经架构搜索与参数调优的

自动化机器学习平台