

旧数据的代码

1. data_explore

单纯的可视化数据，没什么意义，结果存到了tbm_visualizations，就是一些图片

2. data_merge

写了详细的注释，就是分别读取文件然后合并到一起

输入：原始数据

输出：合并后的单个excel

3. ml

这个代码集成了很多的模型，读取excel后会遍历所有目标变量，每个模型单独训练一遍，最终每个目标变量取最好的

输入：合并后的单个excel

输出：指标文件、可视化，全部存在results文件夹里，日志在logs里

注意，这个文件也可以用来训练新的数据，改一下路径即可。

4. opt

这个代码就是优化。逻辑就是：

1. 初始化与数据准备

1. 加载预训练模型、特征选择器和 PCA 降维器
2. 识别目标变量（从模型文件名中提取）
3. 定义可调整的操作参数（推力和转速）
4. 加载并预处理数据，包括填充缺失值和筛选特征

2. 特征处理与预测

1. 根据基础特征和调整后的参数（推力和转速）准备输入特征
2. 应用全局预处理、特征选择和 PCA 降维
3. 使用加载的模型预测目标变量（掘进速度、能耗和磨损）

3. 目标函数设计

1. 综合考虑掘进速度、能耗和磨损三个目标
2. 对各目标值进行归一化处理
3. 使用权重控制各目标的重要性
4. 目标函数值越小越好，其中：
5. 掘进速度越大越好（通过 $1 - \text{归一化值}$ 体现）
6. 能耗和磨损越小越好（直接使用归一化值）

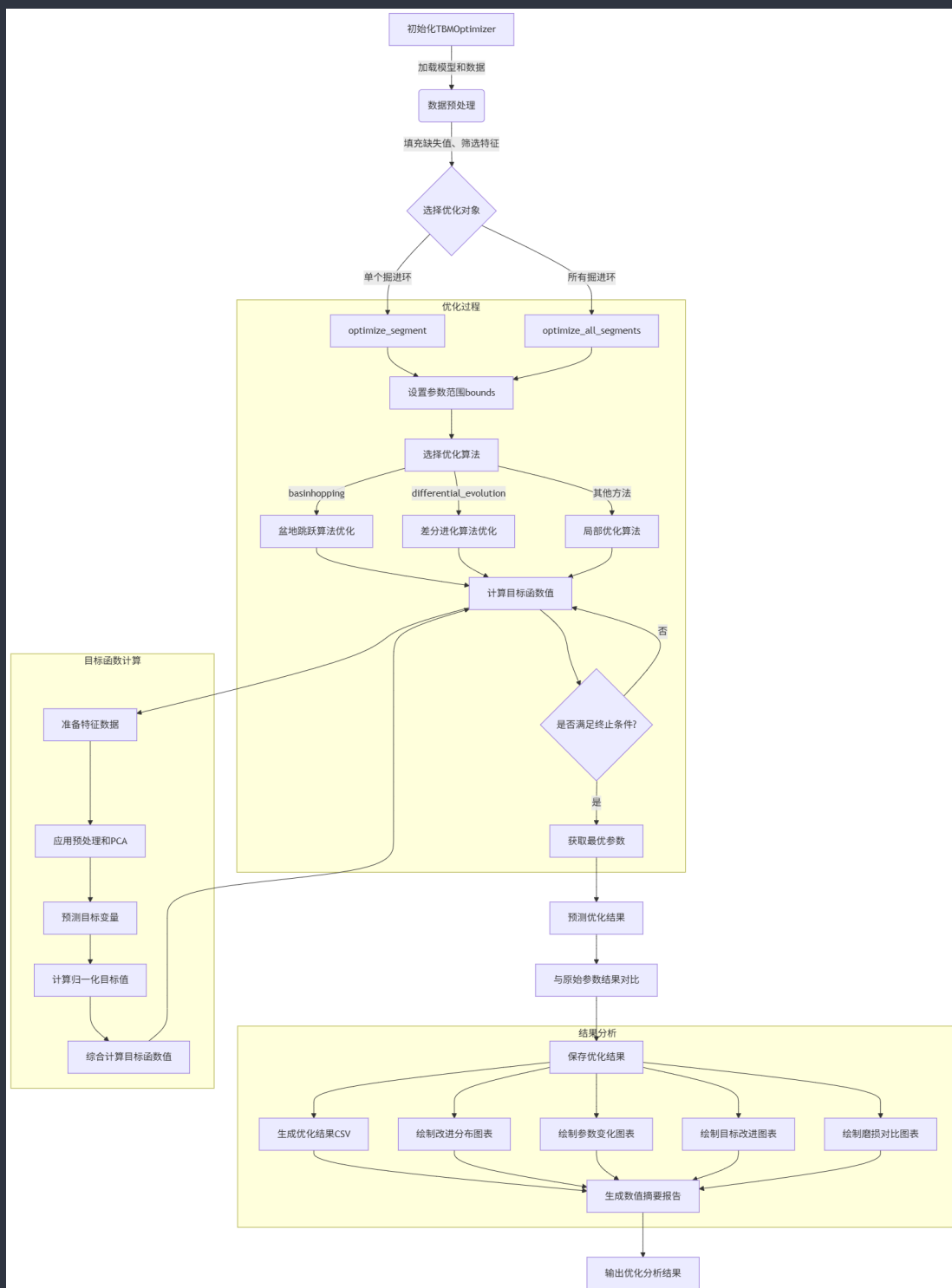
4. 优化过程

1. 支持单个掘进环和所有掘进环的参数优化
2. 提供多种优化算法选择：盆地跳跃算法 (basinhopping)、差分进化算法 (differential_evolution) 和其他局部优化方法
3. 自动设置参数优化范围（基于历史平均值的 0.5-1.5 倍）
4. 优化失败时自动尝试替代方法

5. 结果分析与可视化

1. 保存优化结果到 CSV 文件
2. 生成多种可视化图表：
 1. 优化效果概览（速度、能耗和磨损的改进百分比分布）
 2. 参数变化对比（推力和转速优化前后对比）
 3. 目标函数改进情况
 4. 磨损对比（原始参数、优化参数与实际磨损的对比）
3. 生成数值摘要报告，包含各项指标的改进百分比

具体的流程图：



上面这一套完成了数据分析、合并、构建模型预测。

新数据的代码

1. merge2

合并数据，逻辑和之前一样，不过分组方式有一些改变，具体在注释。

输入：原始数据

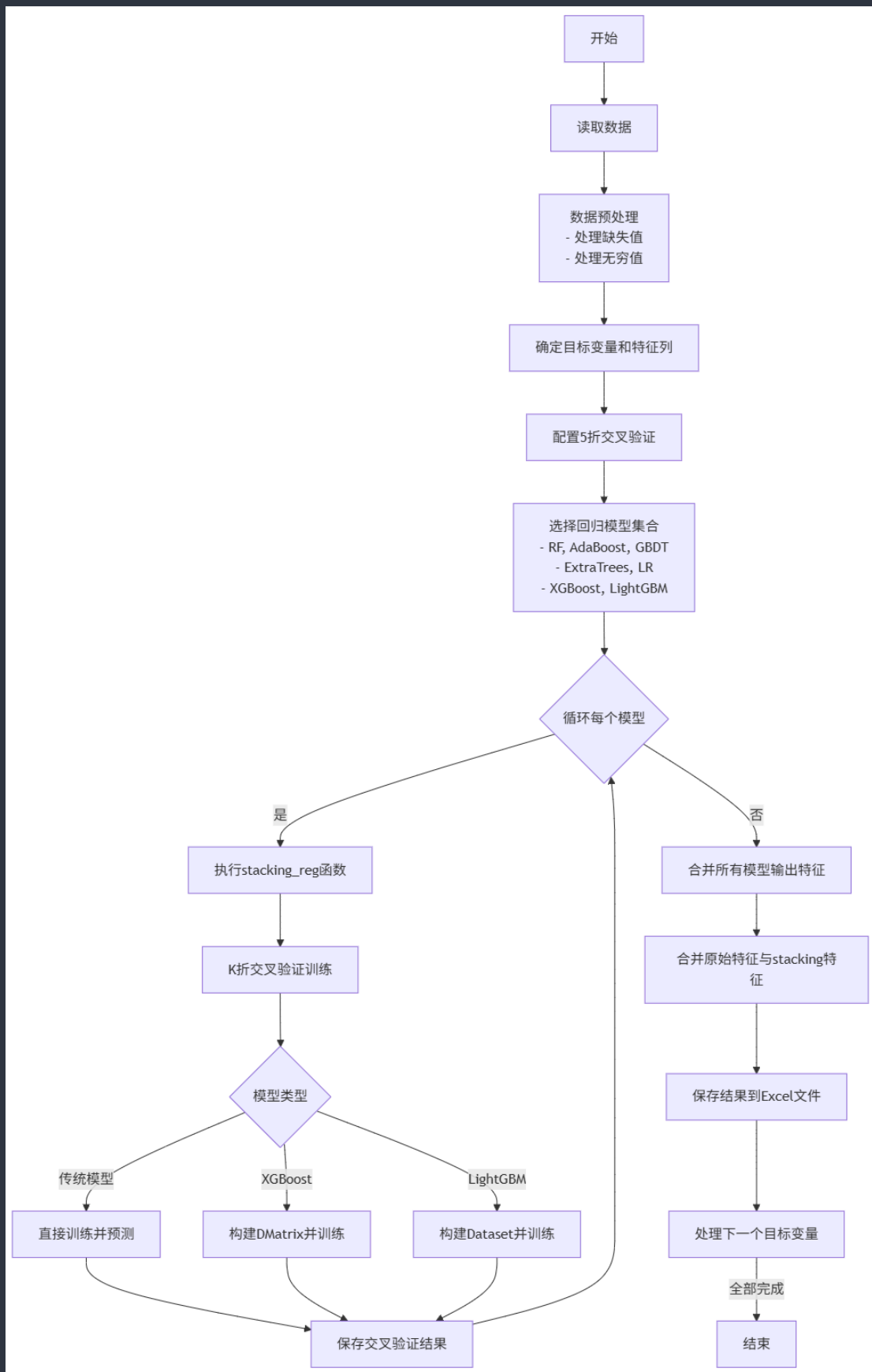
输出：合并后的单个excle

2. stacking

stacking数据特征，相当于使用其它的模型先提取出特征，将这些特征用作预测，这是比赛常用的方法，可以不用这一步，不过我试了一下这个效果好。

输入：合并后的单个excle

输出：每个目标变量一个excle，分别是特征和标签



3. ml2

包含stacking特征的预测。

新数据的优化也可以使用opt，输入内容一模一样，不需要做修改。

注意，优化的一些参数位置我用“parameter”标记，直接搜索就可以找到要改哪些东西来进行调优。