**我按照昨天讨论的内容，重新整理了目前研究的核心思路：**

**P1.现象与痛点：**

数据质量波动会显著拖慢无监督 AutoML 搜索并导致聚类效果不稳定

**P2.关键假设：**

影响具备结构性，可由数据特征–算法特征映射 Φ 预测

**P3.方法设计（三个角度）：**

① 特征增强→更准确预测最优组合

② 搜索裁剪→缩小候选空间

③ 动态调优→运行时修正参数漂移

**P4.理论支撑（可选）：**

证明裁剪后搜索空间上界，并给出 Φ 的时间复杂度

**P5.证据链（用“三步实验”递进式验证）：**

① 影响验证

② 方法有效性

③ 综合增益（速度 × 质量）

**P6.实践价值（可选）：**

在 1 M 行模拟表上提速 8×、损失 <1%

我打算修改的论文结构和写作逻辑如下：

| **章节** | **写作目标** | **对应研究环节** |
| --- | --- | --- |
| 摘要 | 告诉读者 现象 (P1)、核心贡献、最终收益 (P5-P6) | P1, P3, P5 |
| 1 引言 | ✧ 动机示例 → ✧ 问题引入 → ✧ 贡献点列表 | P1, P2, P5 |
| 2 相关工作 | 对齐审稿人知识背景，突出空缺 | P1-P3 |
| 3 问题定义 | 定义数据质量维度、聚类指标、最优方案目标函数 | P1-P2 |
| 4 方法 | 4.1 特征增强 4.2 搜索裁剪 4.3 动态调优  4.4 复杂度与理论上界 | P3, P4 |
| 5 实验设计 | 数据集、评价指标、三段实验设计 | P5 |
| 6 实验结果分析 | 6.1 影响验证 6.2 方法有效性  6.3 综合增益 6.4 讨论 | P5 |
| 7 结论 | 回答开篇悬念，指出局限与扩展 | P6 |
| 附录 | 代码开源链接、完整推导、额外表格 | P4-P5 |