



数字中国创新大赛
DIGITAL CHINA INNOVATION CONTEST

2025 数字中国创新大赛

Digital China Innovation Contest, DCIC 2025

基于人工智能的闽江流域水情预报



自我介绍



魏晋，美国哥伦比亚大学机器人与控制专业毕业，现就职于福建水口发电集团有限公司。公司“三力”协同与智慧平台数字化研究专家团队成员。重点研究水资源知识图谱开发和数字孪生流域构建等攻关项目，致力于探索数字技术在水利领域的深度应用，助力公司数字化转型。

赛题介绍

- 赛题名称
 - 基于人工智能的闽江流域水情预报
- 出题单位
 - 国网福建省电力有限公司
- 赛题背景
 - 闽江，福建省最大独流入海河流，流域降水丰沛。在新型电力系统和新阶段水利高质量发展的环境下，如何缓解流域防汛压力，提升水资源利用率是水电企业高质量发展的重中之重。目前，基于传统水文学原理已实现未来9小时甲级精度的洪水预报，但存在遇见期与预见精度上仍有提升空间，因此，通过结合人工智能技术实现流域入库流量的精准预测，对保障水库安全调度、防洪及清洁能源稳定供应具有重大战略意义。
- 赛题任务
 - 本赛题要求选手借助机器学习技术，基于闽江流域上游监测点的历史水位、雨量等数据，生成逐时入库流量预测模型，实现未来9-24小时入库流量预测，提升模型在台风、暴雨等极端天气场景下的泛化能力，为水电站腾库容、保大坝的“黄金调度窗口”提供科学支撑。

赛题解读

- 任务类型：回归预测
- 任务目标：根据各水文站的历史雨量水位数据、历史入库流量数据对未来入库流量进行预测
- 数据集：
 - A榜-雨量水位训练集.csv (2014-2019)
 - A榜-入库流量训练集.csv (2014-2019)
 - A榜-测试集.csv (2020-2021)
 - B榜-入库流量.csv (2020-2021)
 - B榜-测试集.csv (2022-2023)

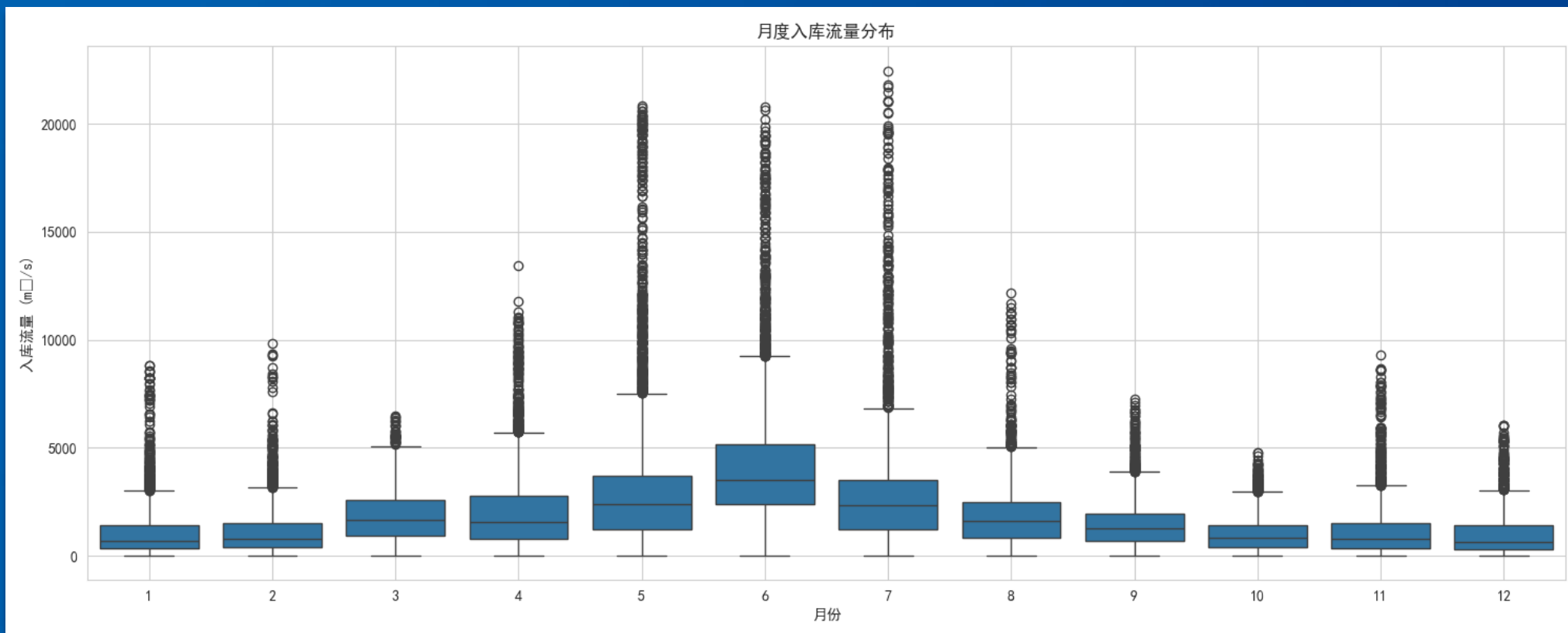
赛题解读

- 输入
 - 各水文站历史雨量水位数据
 - 历史入库流量数据
- 输出
 - 未来入库流量数据（两年）
- 难点解析：
 - 数据量大（训练集3252292行，14列）
 - 早期由于技术原因造成无效数据较多
 - 气候异常导致降水规律难以捕捉
 - 过长的预测序列

解题思路——数据

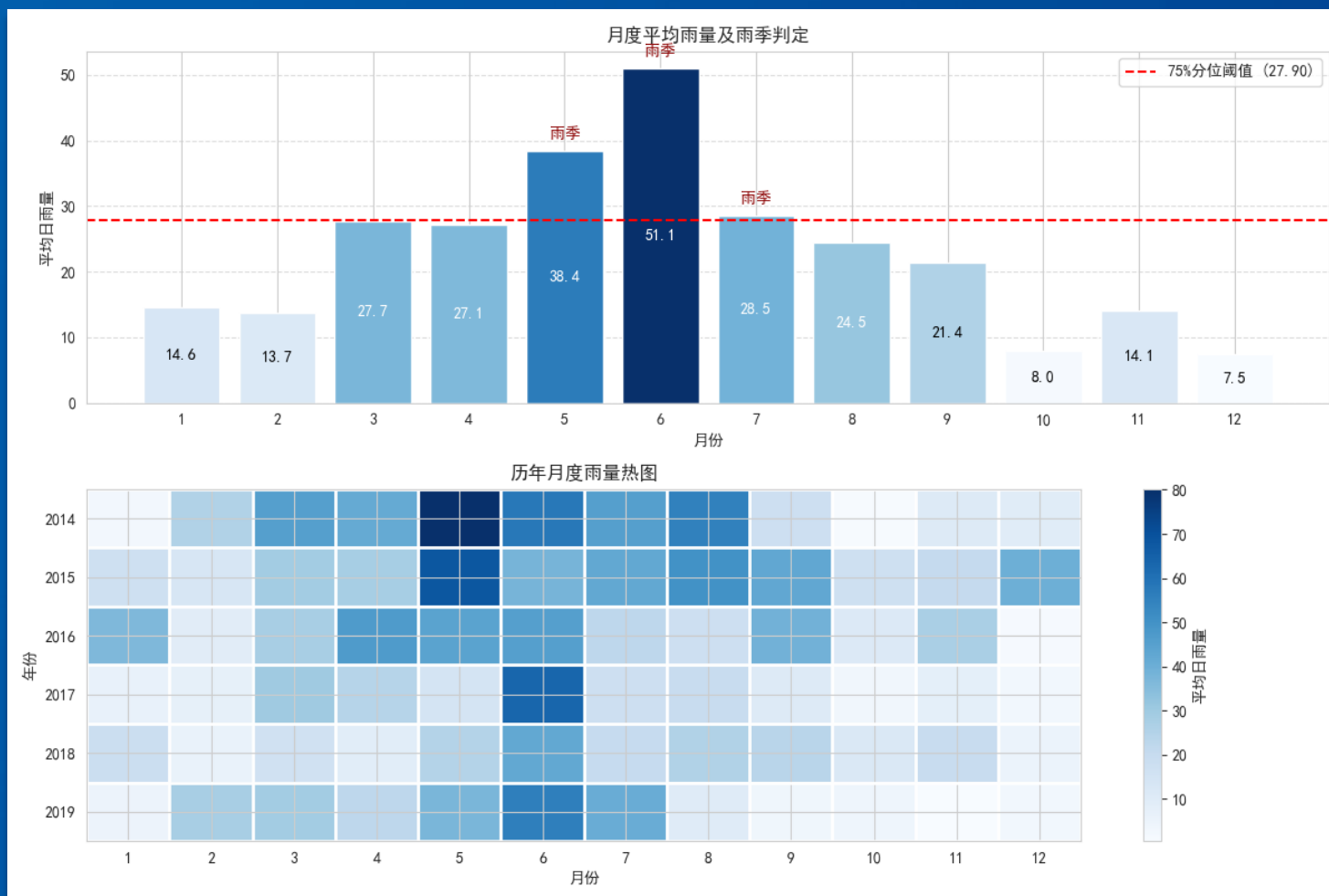
- 数据探索
 - 检查数据完整性：识别缺失值和异常值（如负值）
 - 时间序列可视化：雨量与入库流量的时间分布
 - 相关性分析：不同水文站统计雨量与入库流量的关系
- 数据预处理
 - 缺失值处理：插值或填充
 - 异常值处理：平滑或剔除
 - 时间特征提取：年、月、日、小时、季节等
- 数据集分割
 - 训练集：2014-2019年逐时数据（雨量、入库）
 - 验证集：无
 - 测试集：2020-2021年逐时数据（雨量）

解题思路——数据

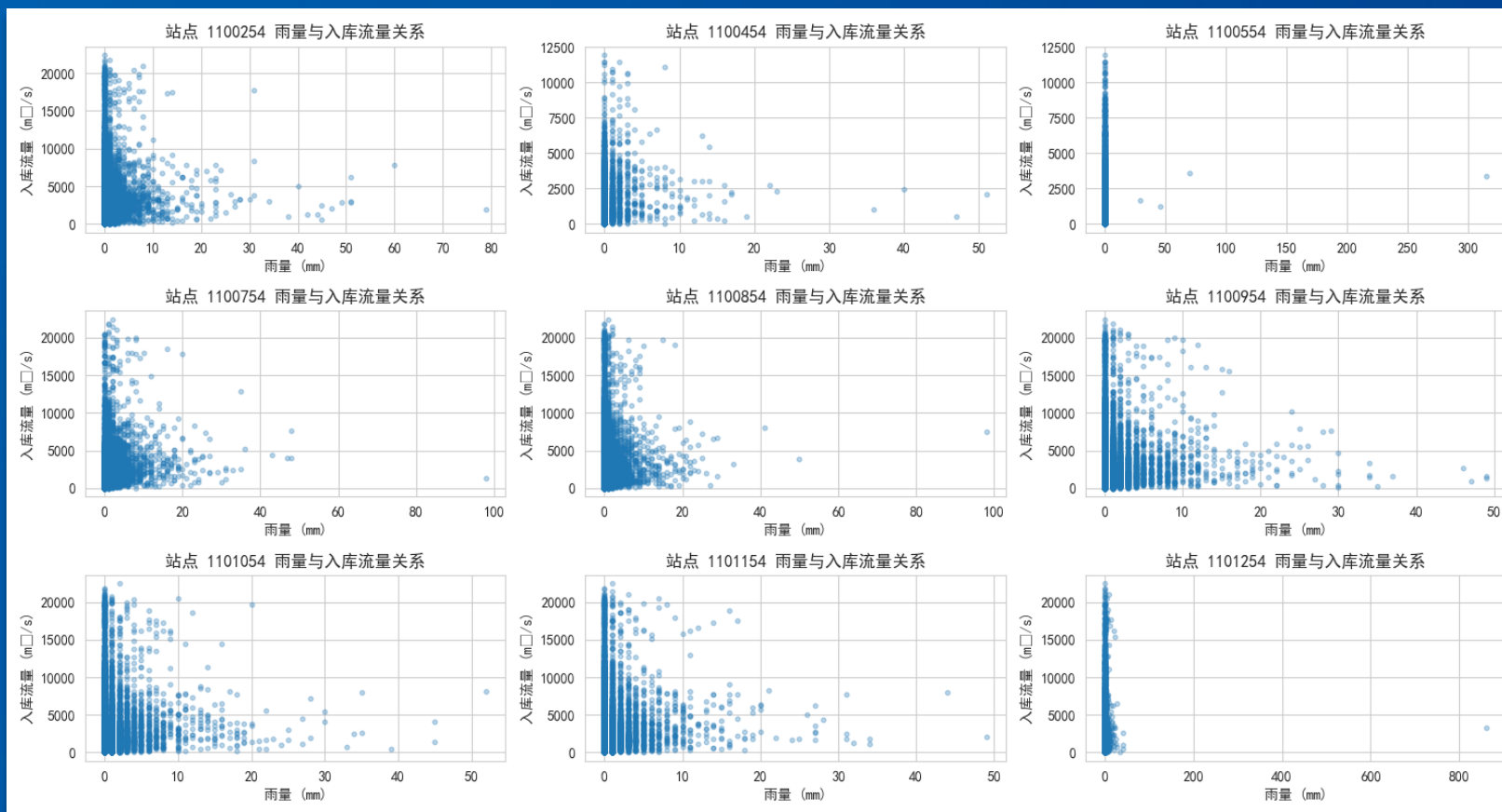


1. 5-7月箱体和中位数较高，可判断为汛期
2. 6-7月离群值较多，说明极端流量情况较多
3. 10-12月的流量较低，可判断为枯水期

解题思路——数据



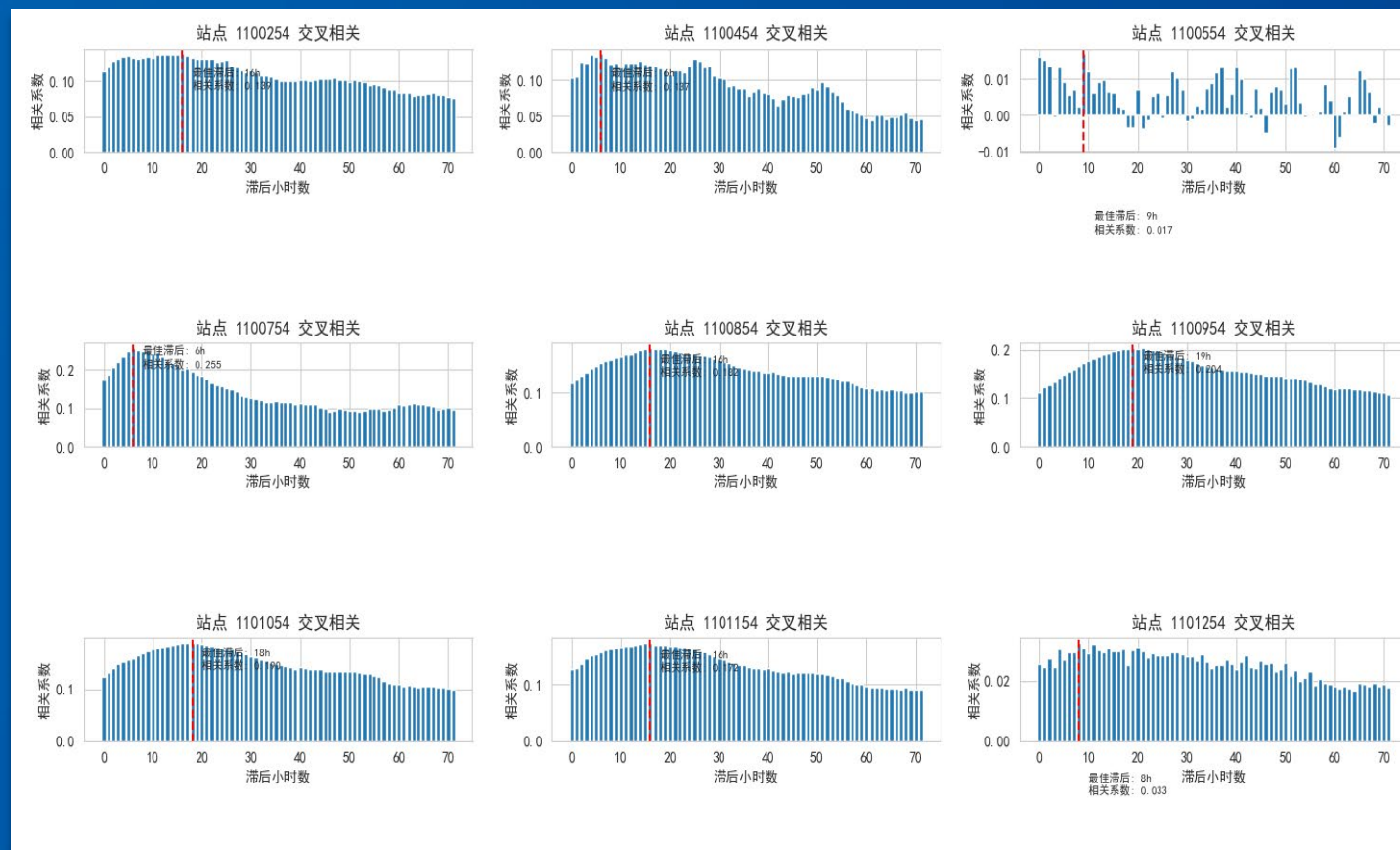
解题思路——数据



解题思路——特征工程

- 时间滞后特征
 - 不同水文站雨量的滞后值 ($t-1$, $t-2$, $t-3$, ..., $t-n$)
 - 入库流量的滞后值 ($t-1$, $t-2$, $t-3$, ..., $t-n$)
 - 确定最佳滞后时间窗口
- 统计特征
 - 滑动窗口平均值 (3h, 6h, 12h, 24h...)
 - 滑动窗口累计值 (尤其针对雨量数据)
 - 滑动窗口的最大值、最小值、方差等
- 空间特征
 - 不同水文站数据的聚合 (平均值、加权平均等)
 - 水文站位置信息整合 (外部数据)
 - 上下游位置关系编码 (外部数据)
- 外部特征
 - 季节性特征
 - 气象数据整合 (外部数据)
 - 水文地质特征 (外部数据)

解题思路——特征工程



解题思路——模型选择

- 基准模型
 - 经典统计模型：ARIMA、SARIMA
 - 多元线性回归
 - 随机森林
- 高级模型
 - 梯度提升树：XGBoost、LightGBM
 - 深度学习模型：LSTM/GRU、Transformer、CNN
 - 混搭模型：CNN-LSTM、CNN-Transformer
- 集成方法
 - 模型集成：投票法、平均法
 - 堆叠集成
 - 分级预测

解题思路——模型选择



解题思路——模型选择

- Temporal Fusion Transformer (TFT)
 - 多模态数据融合
 - 同时处理静态特征（如站点信息、地理信息）和时间动态特征
 - 捕捉长短期依赖关系
 - 利用LSTM单元和多头注意力机制，既能捕捉局部短期依赖，也能识别长期趋势和周期性变化
 - 灵活应对数据缺失与不规则性
 - 针对常见的数据缺失和噪声问题，采用多层门控机制，一定程度提升了鲁棒性
 - 高效的训练和实时预测
 - 兼顾预测精度和计算效率

模型表现

• 配置

- CPU: Intel I9 13900KF
- 内存: 32G DDR5
- GPU: RTX 4090 24GB
- 平台: Linux

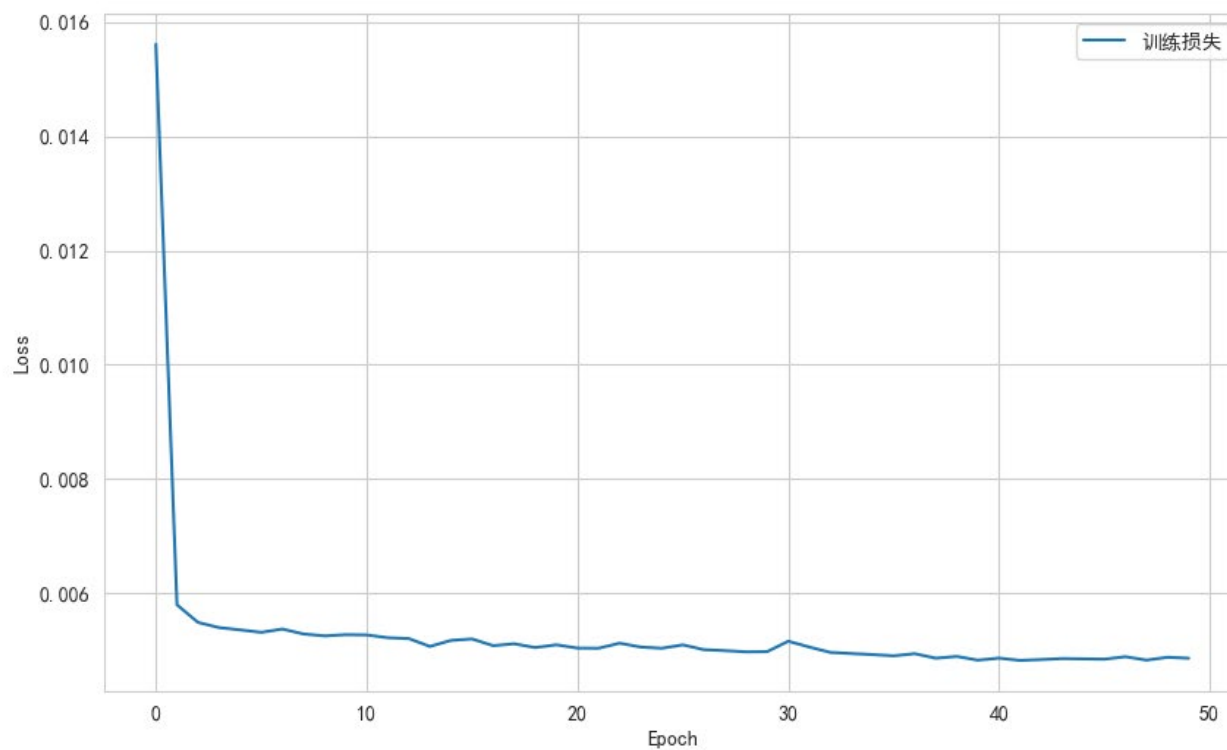
• 表现概况

- 训练 (开启cuda.amp)
 - 速度: 5s/epoch
 - 显存占用: 6G
- 推理 (开启cuda.amp)
 - 速度: 2s (测试集)
 - 显存占用: 4G

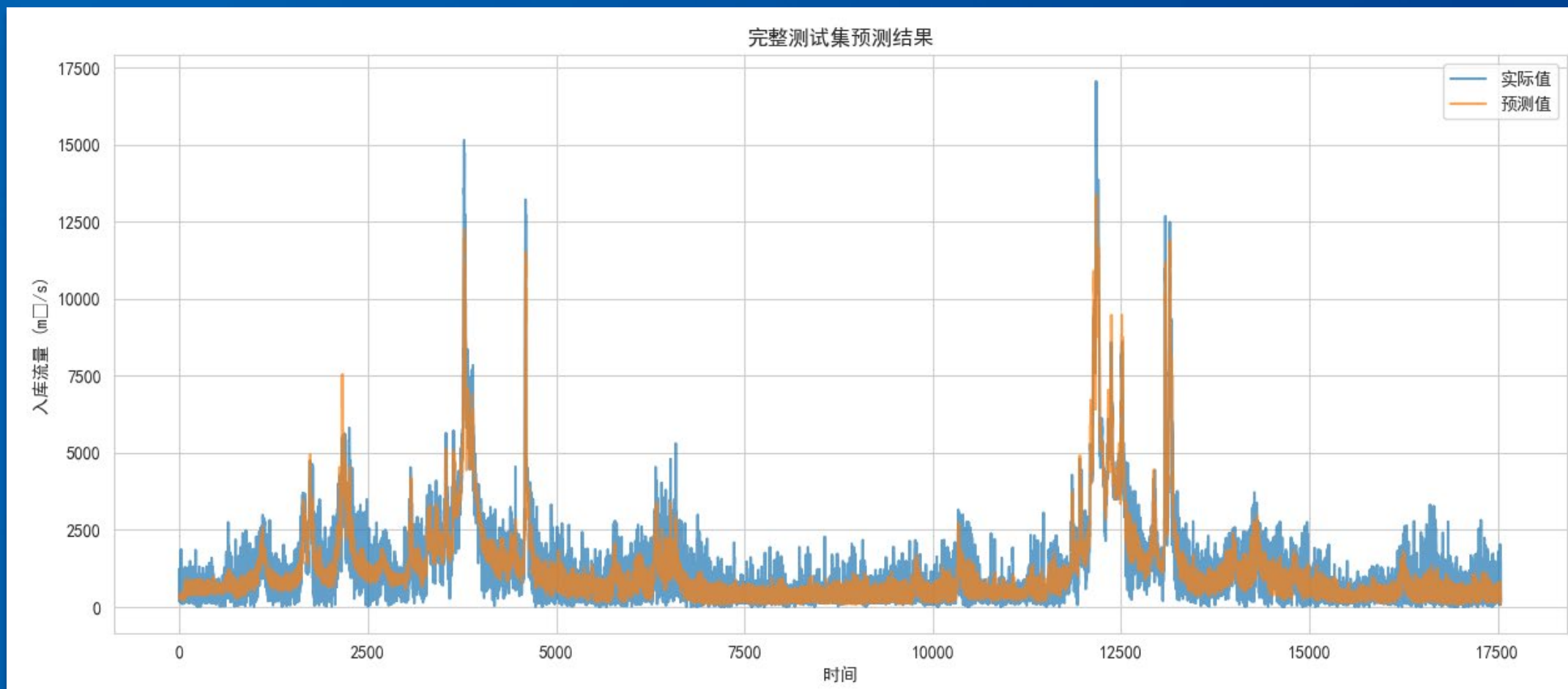
```
TemporalFusionTransformer(  
  (vsn): VariableSelectionNetwork(  
    (var_processors): ModuleList(  
      (0-37): 38 x TimeDistributed(  
        (module): GatedResidualNetwork(  
          (fc1): Linear(in_features=1, out_features=128, bias=True)  
          (elu): ELU(alpha=1.0)  
          (fc2): Linear(in_features=128, out_features=128, bias=True)  
          (dropout): Dropout(p=0.3, inplace=False)  
          (gate): Linear(in_features=1, out_features=128, bias=True)  
          (sigmoid): Sigmoid()  
          (proj): Linear(in_features=1, out_features=128, bias=True)  
          (layer_norm): LayerNorm((128,), eps=1e-05, elementwise_affine=True)  
        )  
      )  
    )  
  )  
  (weight_network): GatedResidualNetwork(  
    (fc1): Linear(in_features=38, out_features=128, bias=True)  
    (elu): ELU(alpha=1.0)  
    (fc2): Linear(in_features=128, out_features=38, bias=True)  
    (dropout): Dropout(p=0.3, inplace=False)  
    (gate): Linear(in_features=38, out_features=38, bias=True)  
    (sigmoid): Sigmoid()  
    (layer_norm): LayerNorm((38,), eps=1e-05, elementwise_affine=True)  
  )  
  (softmax): Softmax(dim=-1)  
)  
  (lstm_encoder): LSTM(128, 128, batch_first=True, bidirectional=True)  
  (lstm_transform): Linear(in_features=256, out_features=128, bias=True)  
  (attn_layers): ModuleList(  
    (0-2): 3 x GatedResidualNetwork(  
      (fc1): Linear(in_features=128, out_features=128, bias=True)  
      (elu): ELU(alpha=1.0)  
      (fc2): Linear(in_features=128, out_features=128, bias=True)  
      (dropout): Dropout(p=0.3, inplace=False)  
      (gate): Linear(in_features=128, out_features=128, bias=True)  
      (sigmoid): Sigmoid()  
      (layer_norm): LayerNorm((128,), eps=1e-05, elementwise_affine=True)  
    )  
  )  
  (temporal_attn): TemporalSelfAttention(  
    (query_proj): Linear(in_features=128, out_features=128, bias=True)  
    (key_proj): Linear(in_features=128, out_features=128, bias=True)  
    (value_proj): Linear(in_features=128, out_features=128, bias=True)  
    (out_proj): Linear(in_features=128, out_features=128, bias=True)  
    (dropout): Dropout(p=0.3, inplace=False)  
  )  
  (output_layer): Sequential(  
    (0): Linear(in_features=128, out_features=128, bias=True)  
    (1): LeakyReLU(negative_slope=0.01)  
    (2): Linear(in_features=128, out_features=1, bias=True)  
  )  
)
```

```
=====  
Total params: 1,206,973  
Trainable params: 1,206,973  
Non-trainable params: 0  
Total mult-adds (Units.GIGABYTES): 17.22  
=====
```

结果展示



结果展示



结果展示

- 模型: TemporalFusionTransformer
- 特征数: 38
- 隐藏层大小: 128
- 注意力头数: 8
- 层数: 3
- Dropout: 0.3
- 批大小: 256
- 学习率: 0.001
- 对数变换: 启用

性能指标

- MSE: 285745.7269
- RMSE: 534.5519
- MAE: 365.1126
- R²: 0.8763
- 竞赛得分: 0.00186723

排名	排名变化	队伍名称	有效提交次数	最高分提交时间	最高得分
	↑ 1	default_0216	30	2025-03-15 15:54	0.00236026
	-	碰碰运气	24	2025-03-15 15:13	0.00190597
	↓ 2	TJU_Baseline	59	2025-03-15 15:16	0.00164538