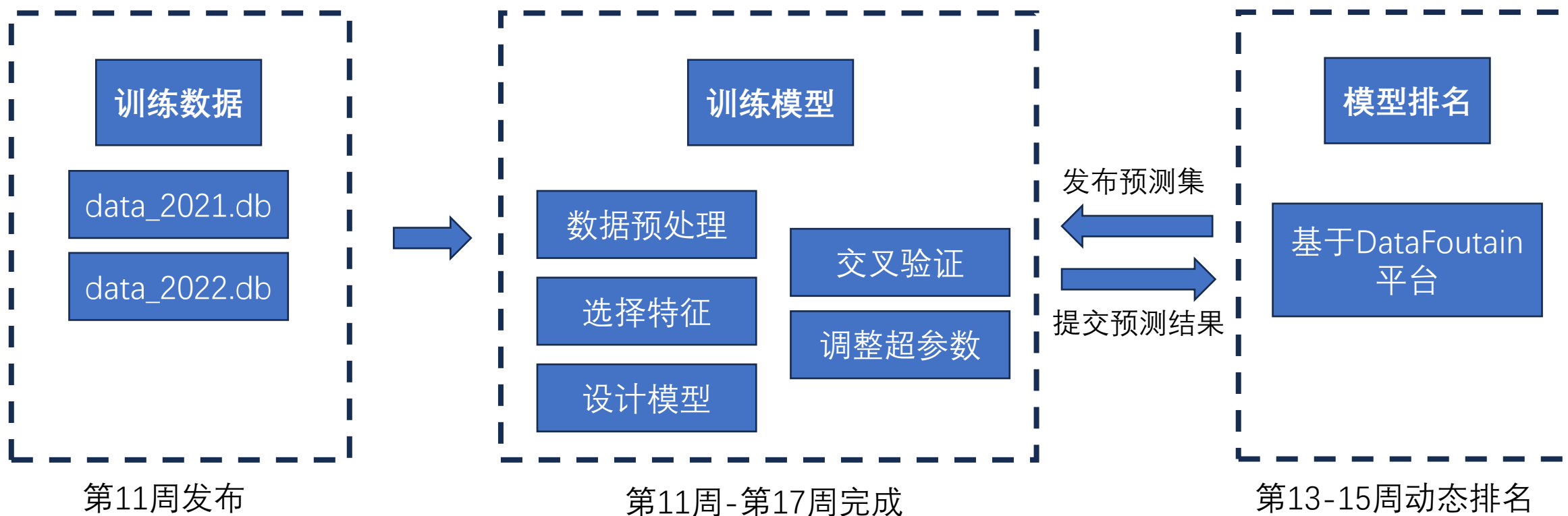


大作业说明

2024.11.20

作业任务

- 基于提供的训练数据，**预测电价**。训练数据通过数据库文件提供。
- 大作业基于DataFountain平台做为期三周的竞赛。



训练数据说明

- 数据量较大，以数据库形式提供，需要使用**数据库语句**进行操作
- 时间颗粒度：5min；空间颗粒度：5个区域
- 数据库文件包含data_2021.db和data_2022.db两年的数据
- 数据库文件包含3个表格（Gen, Demand, Price）
 - 包含每一台机组的投标数据、发电能力（最大装机或新能源预测出力）
 - 包含每一个区域的预测负荷和出清能量价格
- 提供用于训练的数据比较丰富，但是实际进行预测时，机组的**投标数据(price_band, volume_band)将被隐去**（提供样例）
- **自行决定采用哪些特征**（以及是否需要变换）进行训练
- 值得提醒的是，数据中**可能存在缺失或离群点**，同学们需要做判断和处理

| 列名 | 数据说明 |
|-----------------|-------------------------------|
| DUID | 机组编号 |
| type | 机组类型 1：新能源机组 2：其他机组 |
| region | 机组所在区域 |
| time | 运行时间 |
| capacity | 对于新能源机组，为新能源预测出力；对于其他机组，为最大出力 |
| price_band_{i} | 分段投标的价格，i=1-10 (\$/MWh) |
| volume_band_{i} | 分段投标的电量，i=1-10 (MW) |

| 列名 | 数据说明 |
|--------|---------------|
| region | 区域 |
| time | 时间 |
| price | 能量价格 (\$/MWh) |

| 列名 | 数据说明 |
|--------|--------------|
| region | 区域 |
| time | 时间 |
| demand | 区域对应的负荷 (MW) |

模型排名

- 基于DataFountain平台（平台暂未开放）
- 13-15周，为期三周的动态排名，每周一套测试数据（预测12天的电价）
- 平台**每周一发布当周的测试数据**（db文件，格式与训练文件相同），每天可以向平台**提交三次预测结果**（csv文件），平台将生成预测评价指标，根据评价指标进行动态排名
- 上一周测试数据的真实值将会在次周周一同时发布
- 预测评价指标基于MSE，计算方式如下

$$S = \frac{1}{1 + \sqrt{\sum_r^5 w_r \frac{\sum_t^T (\lambda_{r,t} - \hat{\lambda}_{r,t})^2}{\sum_t^T (\lambda_{r,t})^2}}}$$

尽管大作业采用动态排名
但是最终大作业给分不参考该排名