



统计模型和机器学习方法在话务预测的应用

--机票大数据组 侯淑芳

目录

- ◆ 项目背景简介
- ◆ 项目实施和成果
- ◆ 项目尚需改进的问题

◆ 项目背景简介

1.国内预订部现状：

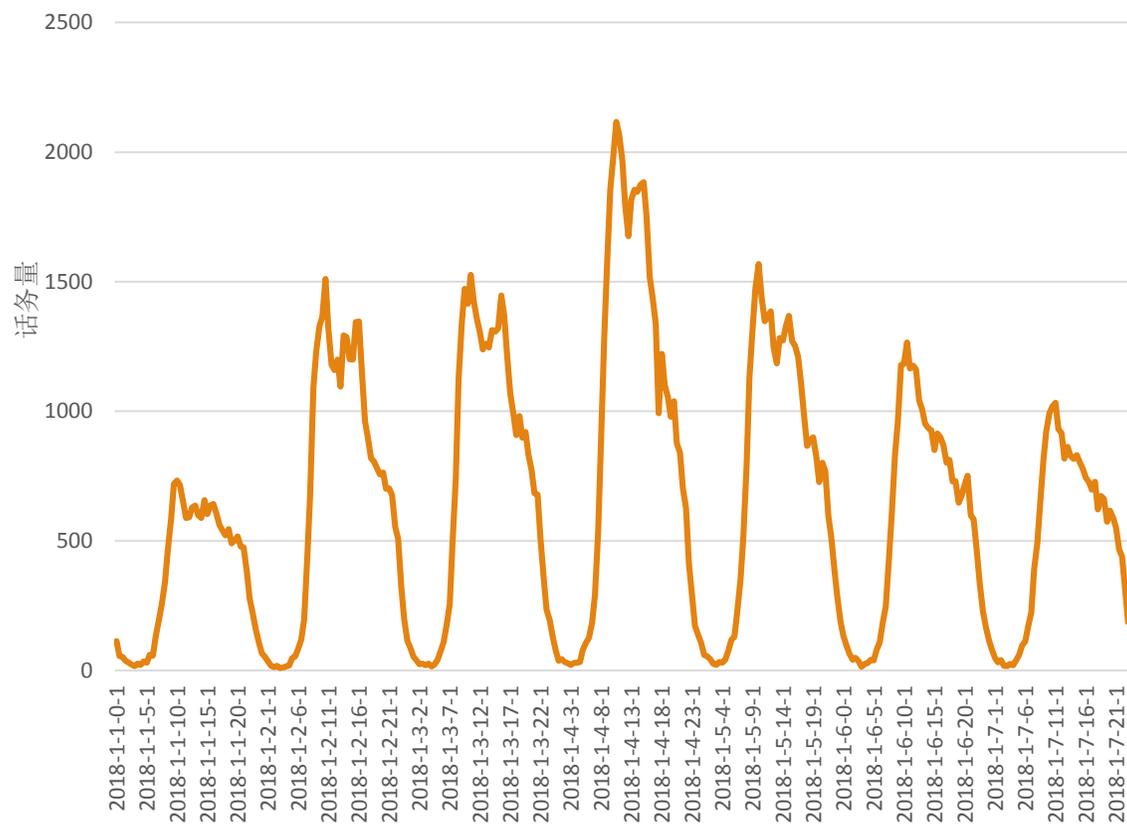
目前国内预订部的话量都是人工预测，在正常情况下，人工预测准确率比较高（92%-95%），**但预订部痛点在于：一旦遭遇突发事件，比如大面积航变，人工预测准确率普遍偏低（60%-70%），由此导致无法精准排班，造成不必要的人力成本浪费的同时也影响了我携的用户体检。**

2.项目目标：

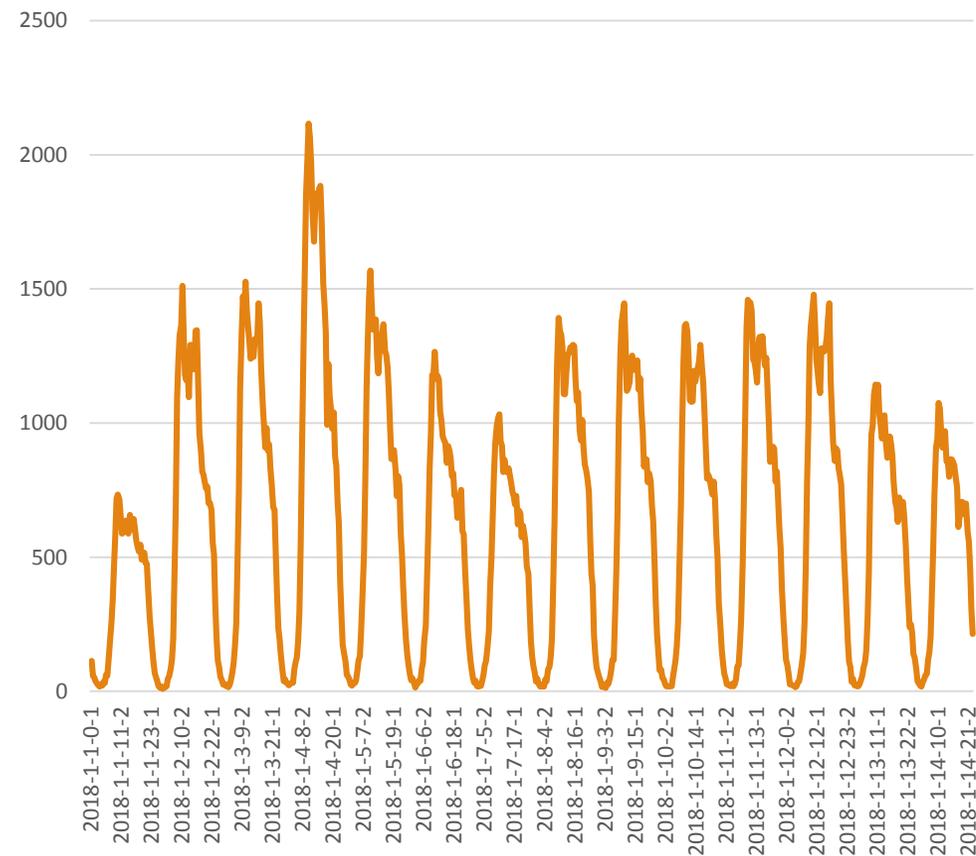
解决预订部的痛点，提升航变时期的预测准确率。本项目主要通过机器学习方法预测未来一周国内预订部每天半小时的话务量，提高话务预测准确度并取代预定部同事手工预测的现状，实现系统自动化预测的同时，保障客服人员的精准排班，在节省人力的同时保证我携对客户的服务质量。

◆ 项目实施—认识你的数据

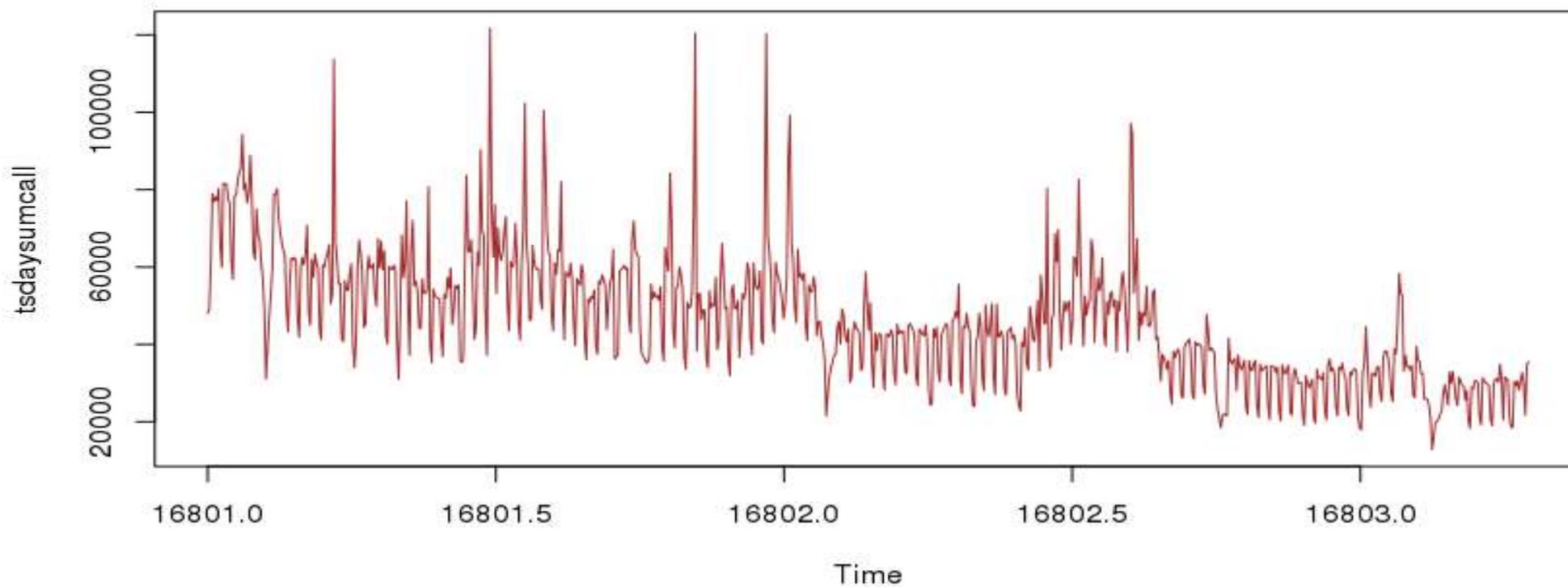
call



call



◆ 项目实施—认识你的数据



◆ 项目实施—认识你的数据

数据表现特征：

- 时间序列特征明显（周期波动比较规律）
- 周期特征多样化
 - 年周期—逐年递减（2015: 30,264,898, 2016: 21,040,231 2017: 14,798,807）
 - 季节特性
 - 周变化趋势
 - 小时周期
 - 节假日波动
- 异常值

◆ 项目实施—模型选择和调优

◆ Msts/TBATS

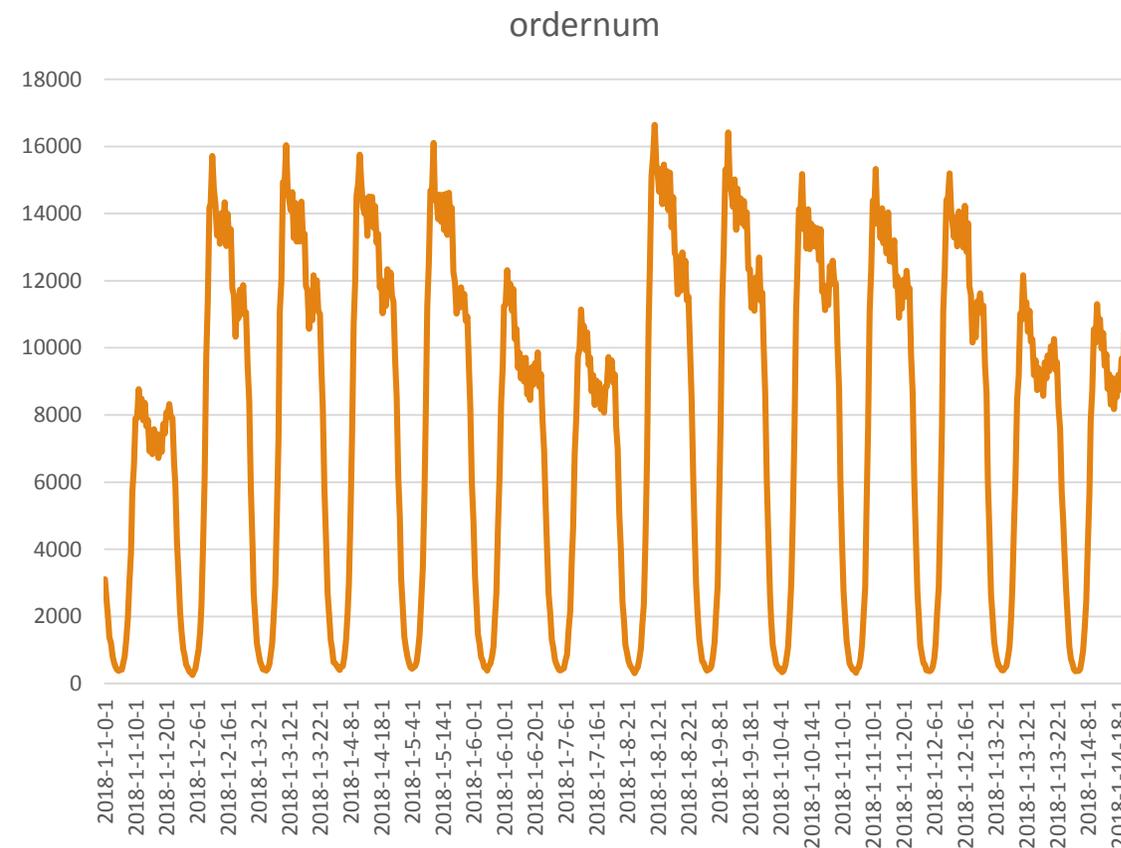
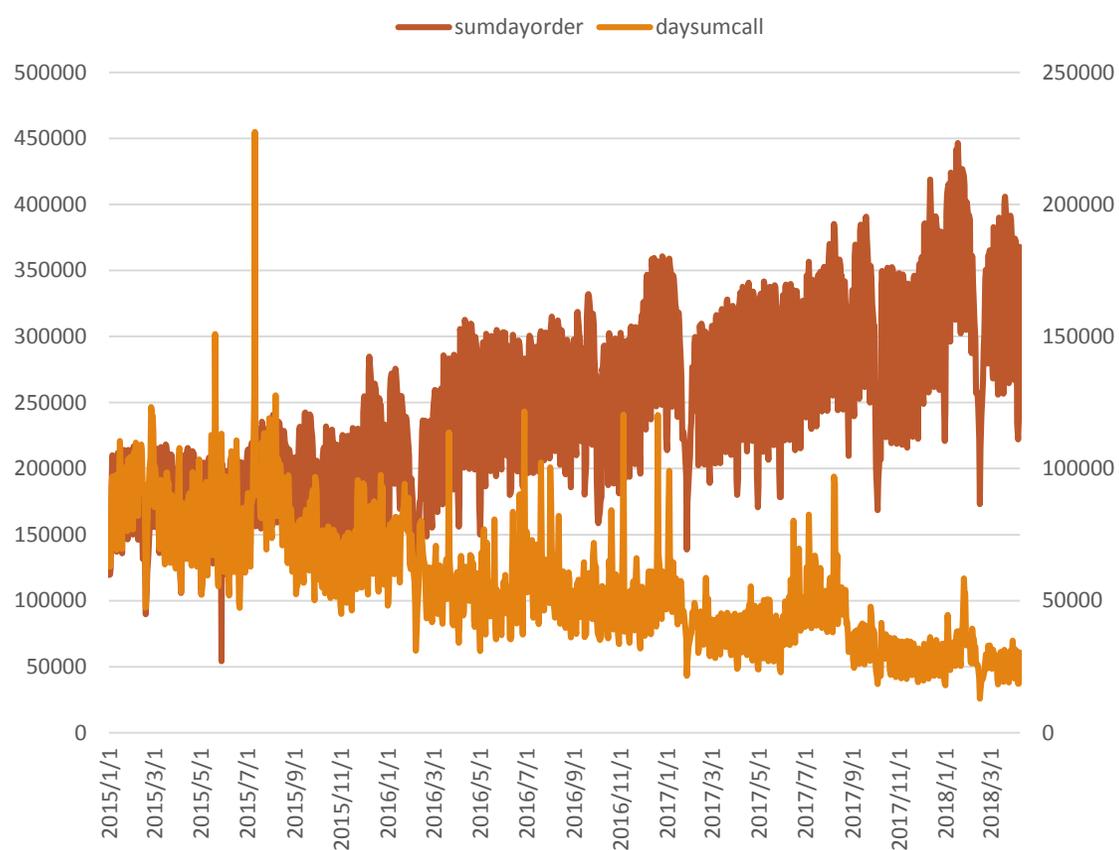
◆ 指数平滑

◆ Arima(arimaX)

$$y_t = a + \sum_{i=1}^M \sum_{k=1}^{K_i} \left[a \sin\left(\frac{2\pi kt}{P_i}\right) + \beta \cos\left(\frac{2\pi kt}{P_i}\right) \right] + ARMA + \text{其他因子} \quad \text{---- (1)}$$

$$(1 - \sum_{i=1}^p \phi_i L^i)(1 - L)^d X_t = (1 + \sum_{i=1}^q \theta_i L^i) \varepsilon_t \quad \text{----- (2)}$$

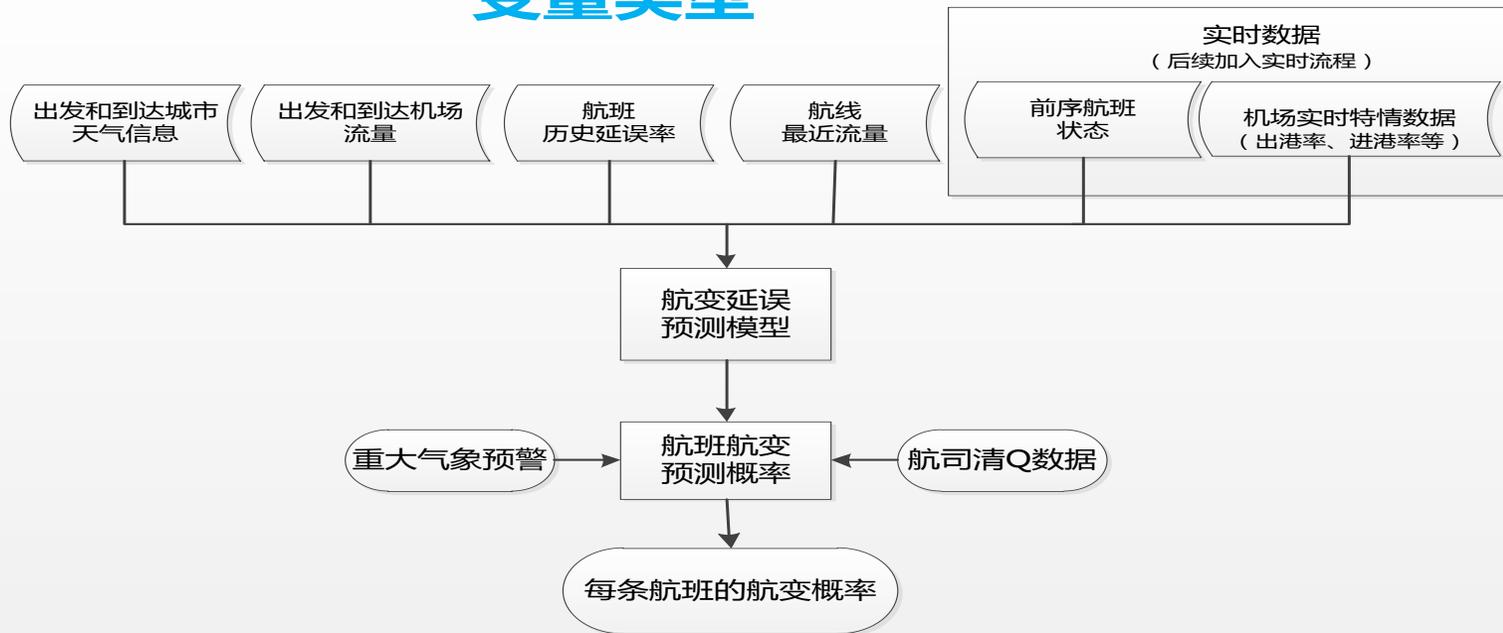
◆ 项目实施—模型选择和调优



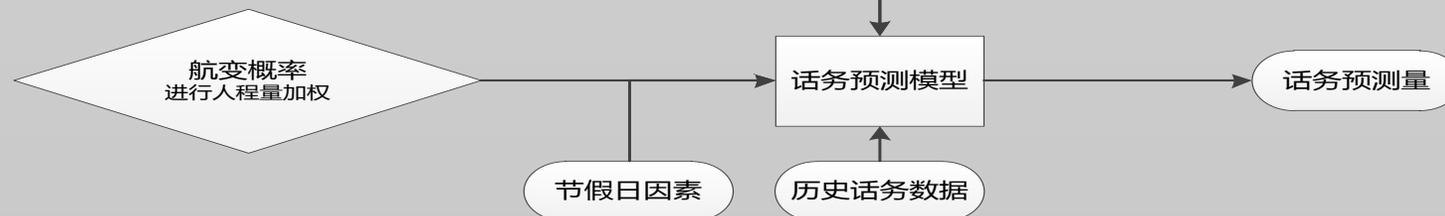
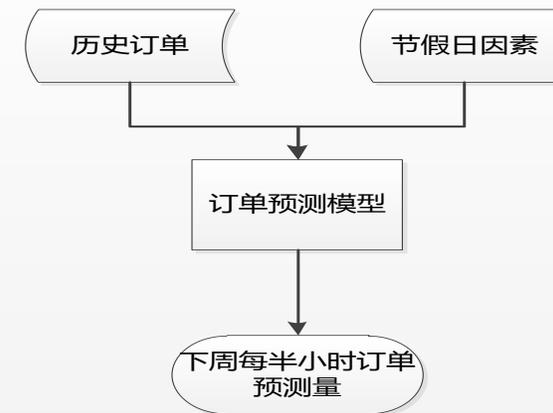
◆项目实现-----预测系统V1.0

国内话务预测系统

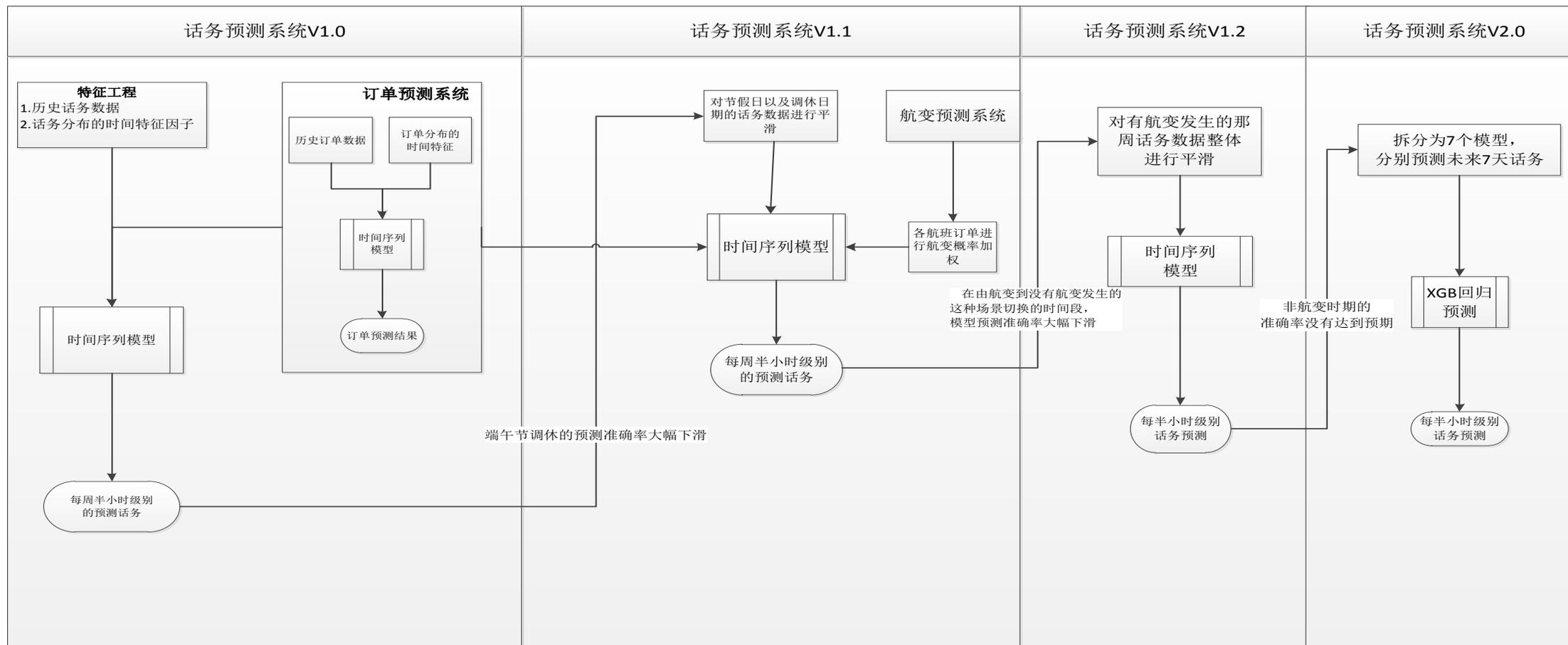
变量类型



订单预测系统

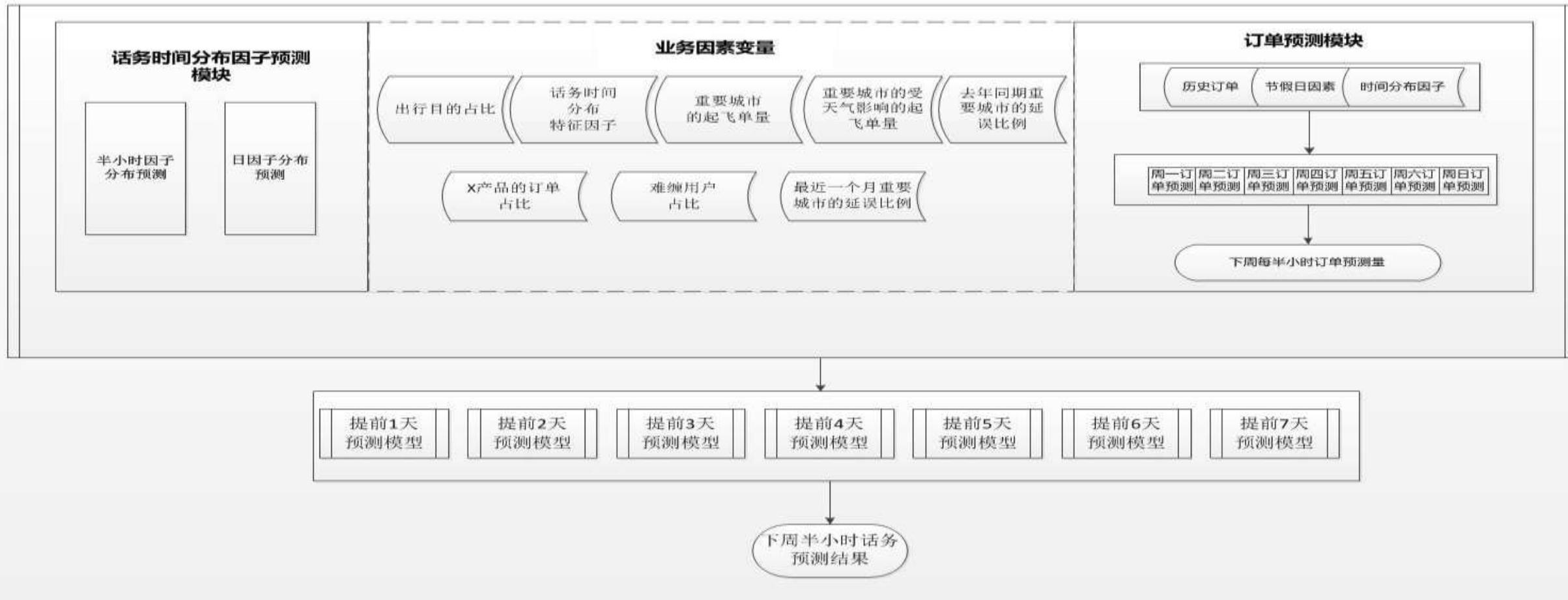


◆ 项目实施---国内话务预测系统迭代开发过程



◆项目实现-----预测系统V2.0

国内话务预测系统V2.0



◆ 项目实施—v2.0

$$1) y_t^{(w)} = \begin{cases} \frac{y_t^w - 1}{w}; & w \neq 0 \\ \log(y_t); & w = 0 \end{cases}$$

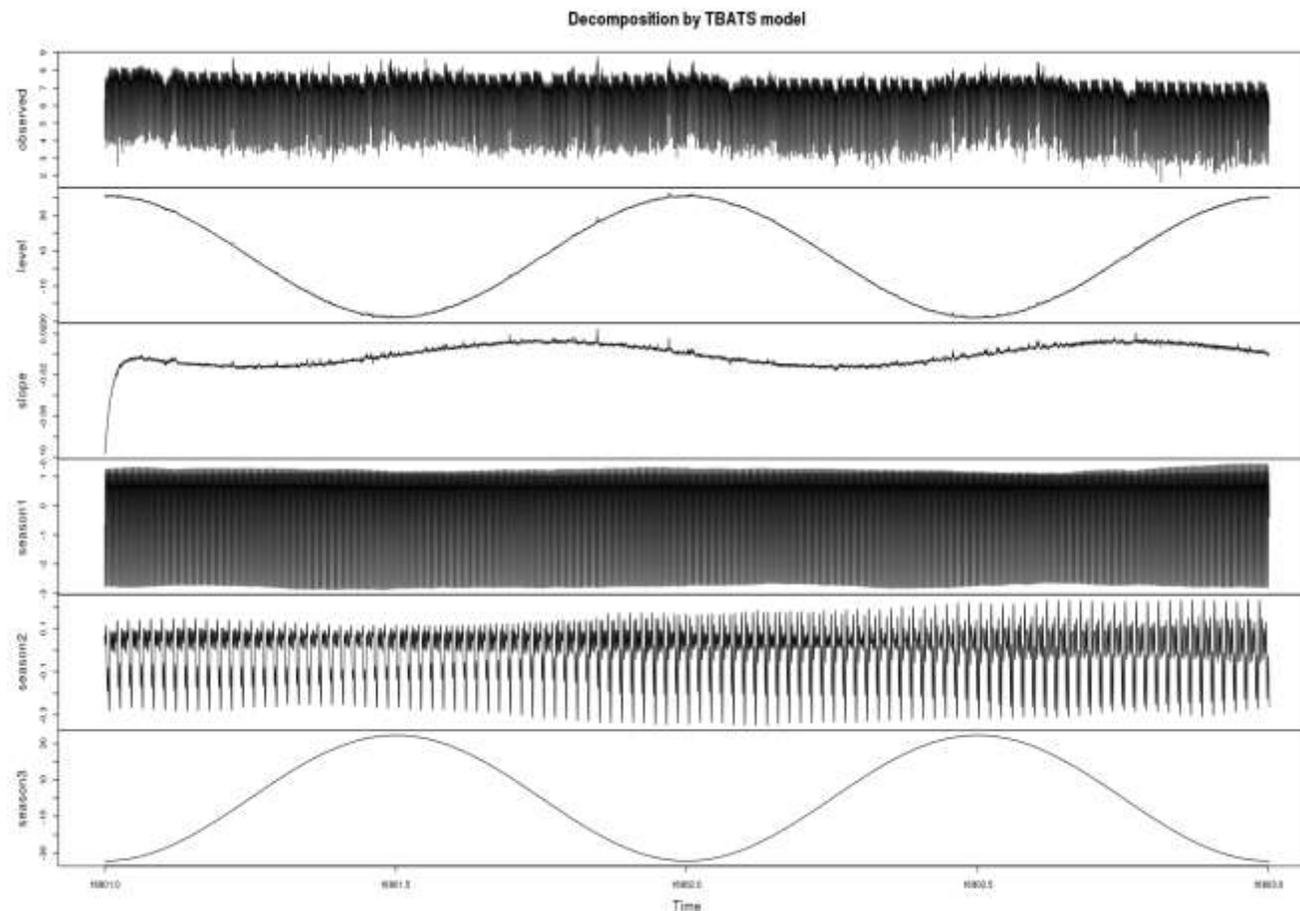
$$2) y_t^{(w)} = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^T s_{t-m_i}^i + d_t$$

$$3) l_t = l_{t-1} + \phi b_{t-1} + \alpha d_t$$

$$4) b_t = (1 - \phi)b + \phi b_{t-1} + \beta d_t$$

$$5) s_t^{(i)} = s_{t-m_i}^{(i)} + r_i d_t$$

$$6) d_t = \sum_{i=1}^p \varphi_i d_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t$$



◆ 项目中尚需改进的问题

- 节假日的预测；
- 重大事件导致的波动预测；

谢谢观赏！

