

交通信息检测与处理

实验 1：基于线圈数据的交通流密速分析（地点：交通学院 A206）

- 注：
- 1) 参考思路以及相应的参考函数仅为其中一种实现；
 - 2) 提前了解 tidyverse 系列包很有帮助，如 tidyr, dplyr, 可在邮箱中下载资料；
 - 3) 有问题及时与助教联系。

1 线圈数据样例

FSTR_LOOPGROUPID	FDT_TIME	FINT_VOLUME	FINT_SPEED	FINT_OCCUPY
检测器编号	日期 时间	流量	速度	占有率
NHNX40(1)	2010-04-24 00:00:00	2	72	2
NHNX40(1)	2010-04-24 00:00:20	7	73	8

2 数据读取

给出的数据有 tmp001.xls~tmp005.xls 共 5 份，每个.xls 文件中有多个 sheet，记录了从 2010-04-18 到 2010-04-24 的多个检测器的数据。

- **提取全部 NHNX40(1)检测器的数据。**（可以先确定该检测器在哪些.xls 文件中，只读取这些文件，也可以读取 5 个文件之后再进行筛选）
 - 参考包和函数： readxl::read_excel, readxl::excel_sheet, dplyr::bind_rows

3 冗余和缺失处理

- **统计冗余的记录数：**冗余是指相同时刻有不只一条记录。
- **冗余数据处理：**如有冗余数据，则流量、占有率和速度采用冗余记录的平均值作为该时刻的记录值。
- **统计每天的记录数，得到每天的记录缺失个数：**缺失是指按照固定间隔采集的记录中，某时刻的记录不存在。
 - 参考包和函数：as.Date, dplyr::group_by, dplyr::summarise, dplyr::left_join, dplyr::arrange, strptime
 - 参考思路：构建完整的时刻序列作为一个数据框，与原数据进行 join 操作，缺失位置即可得到
- **对于缺失数据，采用前三个周期的平均值代替。**（数据中可能有多个连续时刻记录缺失的情况，对于本次实验来说，仍依次按照前三周期均值修复）

4 数据聚集

- **编写数据聚集函数，实现返回不同时间间隔下的流量、速度和占有率。**（函数要求：给定初始 20s 间隔无缺失的检测数据和聚集采用的时间间隔，返回聚集后的数据；注意进行聚集时，各变量的聚集结果不一定是取均值得到；有的变量聚集计算之后会存在 NaN，应予以剔除；明确各变量单位，流量单位统一为 veh/h）
 - 参考思路：确定时间间隔之后，即可知道每几个记录为一组，给它们设定相同的分组记号，然后采用函数 dplyr::group_by 和 dplyr::summarise 进行后续操作

5 可视化与分析

□ 分别基于 20s, 5min, 15min 的所有数据, 绘制 V-O, V-Q, O-Q 散点

图: 同一时间间隔的三张图可以作为子图包含在一张图中; 说明不同时间分辨率时三种图的异同点 (V: FINT_SPEED, Q: FINT_VOLUME, O: FINT_OCCUPY, V-O 图是指: V 为纵坐标, O 为横坐标)

□ 基于 5min 数据的图, 估计通行能力、自由流车速、最佳占有率、最佳车

速、阻塞占有率 (参考《交通数据采集与分析》112 页内容, 各指数看图取估计值即可)

□ 基于 5min 的 O-V 的数据, 移除明显的异常数据, 建立占有率 O 和速度

V 的线性回归模型, 给出关系式, 并对模型参数进行统计检验。 (异常数据可以根据散点图的结果, 对明显偏离可能的回归直线的数据点设置条件进行删除)

■ 参考函数: lm, summary

□ 选择第一天的数据

■ **分别绘制流量、速度、占有率时序图**, 在同一张图中包含某变量的 20s、5min 和 15min 间隔下的变化曲线, 对图表进行讨论。(讨论内容参考: 高峰时段分布, 不同时间分辨率时三组曲线之间的异同点等)

■ **基于 15min 数据, 绘制流量-速度时变图**, 说明二者的变化关系。

(即在同一张图中包含流量随时间变化的曲线和速度随时间变化的曲线, 两条曲线分别采用左右两个纵坐标)

■ **基于 15min 的数据**

- 估计车辆平均有效长度
- 绘制车长的时序图
- 绘制车长的分布