**数据修复步骤**

针对杭州试点区91个路口进行数据修复，最终构建道路交通信息模板，并进一步对实施数据进行修复。

**1. 异常数据的自动判别**

针对异常数据进行自动判别（以单线圈一天的数据为对象），判别规则如下：

1）判断当前日期线圈数据为0的个数，如果为0 的个数小于10，则判断当前日期该线圈的数据为正常；

2）如果为0 的个数大于90，则判断当前日期该线圈的数据完全不可用，该线圈为数据完全不可用线圈；

3）如果为0 的个数在10~90，则判断当前日期该线圈的数据部分可用，该线圈为数据部分可用线圈。

通过上述步骤，可以统计出当前线圈数据质量情况。（上述10和90两个数值的设定，根据具体情况可变，代码：20170801.m）

**2.异常数据的自动修复**

对于上述存在质量问题的数据进行修复（以单线圈一天的数据为对象），修复规则如下：

1）对于当前日期数据存在异常，结合周期的相似性，利用相邻（相同）星期的数据进行修复；如果相邻（相同）星期的数据也存在问题，判断相同行车方向线圈个数是否大于1，是，转到2；否，当前日期数据无法修复；

2）如果相同行车方向不止一个线圈，可以用其他线圈数据进行修复；如果相同行车方向线圈数据质量也存在问题，当前日期数据无法修复。

3）如果上述方法仍有线圈无法修复，则可以利用空间关联关系表进行修复。首先判断当前线圈所属交叉口的上游交叉口，然后判断上游交叉口相同断面所有的车道，选择其中直行的车道，将流量、饱和度分别加和除以需要修复的断面所有的车道个数，即为需要修复的车道的流量和饱和度。（20170811.m）

针对道路交通长时段修复的结果进一步进行单点修复。（170912.m）

**通过上述步骤实现道路交通信息模板的构建。**

**3. 实时数据的自动判别**

实时数据的**自动判别**主要是针对7:00-22:00时段的流量数据进行的，通过设定阈值判断流量是否异常。判别的步骤如下：

1) 如果在该时段的最大流量小于等于30，则认为该天该线圈车道流量较小，认为数据正确，并将其日期、车道、线圈编号信息写入到non\_need\_repair.txt，否则转到2；

2) 如果在该时段的最小流量大于等于15，也认为该天该线圈车道流量正常，并将其日期、车道、线圈编号信息写入到non\_need\_repair.txt，否则转到3；

3) 如果1)和2)都不满足的话，则认为该天该线圈车道流量异常，并将其日期、车道、线圈编号信息写入到need\_repair.txt. （170918.m）

**4. 实时数据的自动修复**

实时数据的**自动修复**主要是利用自动判别方法产生的non\_need\_repair.txt对need\_repair.txt中记录的异常数据进行修复，修复的步骤如下：

1) **首先利用时间相关性进行修复**。 对need\_repair.txt中的第m行，判断在non\_need\_repair.txt中是否存在相同的路口、线圈、相应日期（相同星期）的记录行，如果有，设为第n行，然后对txt2中第m行对应记录的数据在7:00-22:00时段进行修复，如果在i时刻的流量小于15，则用non\_need\_repair.txt中第n行对应记录的数据在i时刻的流量和饱和度进行替换。如果没有，则转到2)

2) **道路交通信息模板进行修复**。 如果1)和2)中均无法修复，则利用道路交通信息模板中的数据进行修复。对txt2中的第m（1<m<txt2行数）行记录，如果在i时刻的流量小于15，找到信息模板中相同路口、相同线圈、相应日期（最近日期的相同星期）的数据，用该数据在i时刻的流量和饱和度替换修复。

**5. 道路交通信息模板的更新**

道路交通信息模板的更新，主要是利用2修复好的数据，添加到对应路口、相同线圈的道路交通信息模板当中，同时将日期最旧的数据删除，完成道路交通信息模板的更新，模板随着不断更新，可靠性不断增大。