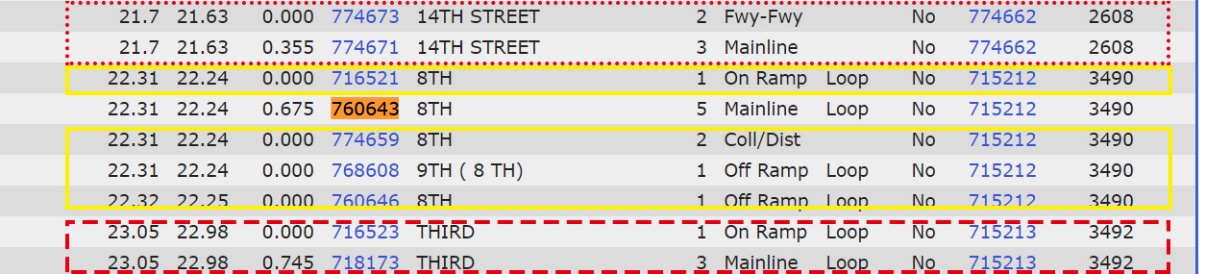
1. 数据集说明



* 1. 上图解释：
     1. 图中橘黄色标注的760643为目标传感器位于主干道(Mainline)上数据，即我们要在该点的历史数据进行建模预测。
     2. 黄色虚线框标注了两部分数据，其中On ramp 代表上游匝道，即表示其他道路的车辆要驶入本条道理，并在若干时间后经过本传感器。Off Ramp 表示下游匝道要驶离该道路。由于这些匝道数据缺失较多，不予考虑
     3. 红色虚线框标注的数据表示距离该节点最近的两个上下游传感器节点，且都位于主干道上。上游传感器节点为774671，下游传感器节点为718173。
     4. 在这里提取了目标传感器节点和两个上下游传感器节点的数据作为训练数据。并整理在ending\_dataset文件中。

1. 算法描述
   1. 如上图所示，橘黄色标记部分是目标道路，在我们的研究中只使用该点历史上2.5天的历史数据作为训练数据对未来五分钟进行预测(确定数据量为2.5天的原因是因为高斯过程回归模型对于数据规模和特征维度有一定限制性)。在上述历史数据中其中2天数据假设是存在于云端的具有高噪声的历史数据，0.5天是存在于边缘节点的无噪声历史数据。这样构建一个融合模型对未来5分钟内的交通流量进行有效预测。再经过五分钟后，我们可以获取了道路最近五分钟的真实历史数据后，该模型可对下一个五分钟进行准确预测。该模型由于使用了较少的历史数据，模型泛化性能不太好，但为了追求模型小型化因此本模型需要两小时更新一次，来保证预测时间段内模型预测效果的准确性。
   2. 对上述论述进行举例
      1. 例如当前时刻为周三中午12时整，此时边缘节点可从本地缓存池中获取今天0时-12时的无噪声历史数据和来自云端周一至周二区间内48小时的有噪声历史数据作为训练数据进行数据建模。该模型首先对12时5分进行交通流量预测。当时间经过12时5分后，在边缘节点缓冲池中已获得12时5分的真实数据，因此我们可以利用这个真实数据对12时10分时的交通流量进行准确预测。在这个过程中，我们的算法模型是不再进利用最新获得历史数据进行实时更新的。
      2. 经实验得知，该局部模型的有效预测性能基本可持续两个小时，也就是说在下午14时，我们会对模型进行更新。因此我们可以这样设定，边缘节点由于自身存储能力有限需要每隔两个小时将数据上传至云数据中心，这样保证每隔两个小时，我们可以从云数据中心获取最新的有噪声的实时更新历史数据和边缘端无噪声的实时更新历史数据。
      3. 总结，这样类似于一个间隔为5分钟的时间窗，但模型不需要每五分钟更新一次。