

交通检测器数据质量控制及预测

任务描述：

任务一：

描述：为了掌握交通运行状态，在公路及城市道路中大量应用交通检测器（如环形线圈检测器、微波检测器、超声波检测器等）用于采集断面交通数据（如车流量、速度等）。但因种种原因（如供电系统不稳、信号传输故障等等），采集得到的数据质量存在着诸如缺失、异常等诸多问题。其中，常见的缺失数据有两种情况，一类是数据部分信息缺失，即一条数据中字段不完整或个别字段出现空缺，或是某字段下出现了多个不同类型的数据，另一类是数据未上传。缺失数据的处理方法一般包括删除数据、数据插补等。若缺失数据未处理，会导致之后的数据应用存在诸多方面的问题。一条道路的通行能力有限，且行驶车速也有限，因此交通设备检测的流量、车速值、时间占有率在一个阈值范围内，超出这个范围则认为其异常，并将其归为异常数据。对异常数据进行在线识别，首先要检查数据的完整性，包括某时刻是否有数据、字段是否完整以及时间是否正确。接下来的方差检验要检验速度、流量、时间占有率三参数的方差是否为 0（即识别出恒定不变的数据），对于有参数方差为 0 的数据进行阈值检验，参数不为 0 的数据先进行组合检验，然后进行阈值检验。最后检验数据的 0 值情况。无论是缺失数据，还是异常数据等，在进行原始数据处理时，都需要先进行数据清洗，清洗之后的数据才能用于交通预测和应用。检测器采集的原始数据存在的这些问题，对交通流数据分析以及交通管控精准性造成影响。因此，提高检测器数据质量是实现交通运行状态监测的基础，是交通状态辨识、交通管理及控制等工作正常进行的必然要求，也是本竞赛需要完成的第一项任务。

要求：对于给定数据 (data 1)，需要建立识别问题数据（包括异常数据、重复记录等）的方法和流程，并完成对问题数据的修正及缺失数据的填补。

任务二：

描述：作为一个具有惯性的系统，交通系统在时间上表现出较为强烈的依赖性、拟周期性、相似性等显著特征。在时间尺度上的建模，对于分析交通的演变规律具有十分重要的意义和作用。不仅对于交通系统分析及改进具有积极的促进作用和指导意义，而且对于完后本竞赛后续任务也起到一定的支撑作用。

要求：因此任务二中，首先需要在时间维度上对交通流参数（流量、速度和时间占有率）进行建模，用以刻画交通系统在时间上的演化过程。其次，需要对这三个参数之间的关系进行建模。对于给定的数据 (data 2)，将缺失的数据 (NA) 补全，即用数值代替 NA。

任务三：

描述：研究交通量的变化规律，并对未来时刻交通量或发展趋势进行实时、准确、科学合理地预测，对于进行交通规划、交通诱导、交通管理、交通控制与安全等都具有十分重要的意义。在交通规划领域中，交通需求预测四阶段法包括交通生成预测、交通分布预测、交通方式划分和交通分配。交通需求预测是进行交通规划乃至城市规划、地区规划必不可少的基础，是确保交通规划符合未来发展状况的重要条件。在交通管理与控制领域中，需要对现状交通进行调研，并定量分析各种土地使用及开发强度与交通发生与吸引量的关系，预测目标年产生的交通量及其发展趋势，并考虑交通需求与路网容纳能力是否匹配，为改善现有交通管理与控制设施提供参考。如：交通信号灯的设置需要根据道路车流量的信息来实现，通过对现有车流量的监测和对未来车流量的预测，来改善交通信号灯配时与布设。通过交通流量预测，可以优化道路网性能使得城市道路更加高效、安全，还可以平衡交通流量、缓解交通拥堵、提高道路交通网络的整体效率。在交通诱导方面，交通诱导是将时变的交通出行需

求合理分配到不同路径上,以降低个人的出行费用,优化交通网络资源。而交通流量的预测是交通诱导中动态路径诱导算法的实现依据,交通诱导行为往往发生在比较大的空间范围上,需要跨越多个预测区段。交通流量预测在交通的各个领域均有不可忽视的应用,是各个应用开始的基础,也是改善现有道路交通设施的关键前提。

要求: 由于不同的应用场景对预测有不同要求,因此本任务要求对给定的数据(data 3),建立模型并对于给定的数据(data 3),将缺失的数据(NA)补全,即用数值代替 NA。