

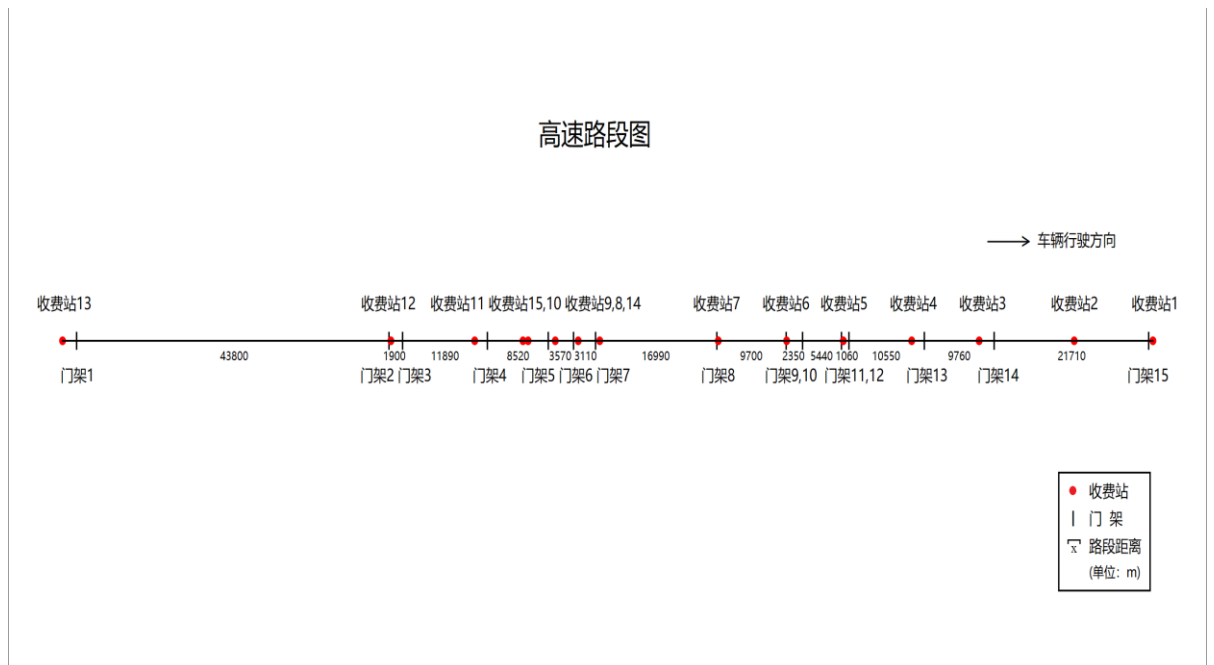
高速公路车辆轨迹数据的分析应用

近年来，随着我国经济社会发展和人民群众生活水平提高，高速公路交通流量大幅增加，高速公路用户对高速公路服务水平的要求也不断提高，因停车收费造成交通拥堵的现象时有发生。因此为了克服传统的全人工及半自动收费中存在的种种弊端，以不断提升高速公路通行效率和服务水平，电子收费系统(ETC)应运而生。该系统最主要的组成部分是建造在高速公路主线横断面上的门架，主要包含车辆视觉识别、天线、边缘计算服务单元等关键设备。ETC 依托高速公路一定规模区域内密集、连续布设的门架系统，结合高速公路收费站等其他必要的数据资源，实现车辆行驶轨迹检测。

一、分析问题

为了分析某高速路区间的交通状况，对经过该高速路区间车辆的信息进行采集。该区间以收费站 13 为起点，依次包含收费站 12、11、15、10、9、8、14、7、6、5、4、3、2，以收费站 1 为终点，共 15 个收费站。区间内同时设有用于车辆识别的 ETC 门架共 15 个，依次从门架 1 至门架 15。收费站、门架分布以及以相继两个门架为起、终点进行分段的各路段间距见附件 1 “高速路段图”。附件 2 记录了由收费站和门架采集的自 2022 年 2 月 22 日至 2 月 27 日的六天时间内经过该高速路区间的全部车辆的轨迹信息。

如图：



二、假设化简

轨迹表字段释意：

列序号	字段	备
A	车牌编号	编号从“车 000001”到“车 203467”，一个车牌编号对应一辆车
B	轨迹编号	车辆轨迹序号，一辆车的轨迹 i 表示该车在所考查的时间段内第
C	信息类型	包括基本信息、ETC 进站、ETC 出站、MTC 进站、MTC 出站、补
D	收费站/门架编号	收费站包括收费站 1-收费站 15，区间外收费站为其他；门架包
E	记录时间	车辆通过对应检测系统的时间，格式：日/月/年 时:分:秒
D	车型	包括客车、货车及其他车型
E	轨迹数	车辆轨迹总数

轨迹表见于附件 2

备注：附件 2 将共计 203467 辆车的轨迹记录在 30 张轨迹表中。例如“轨迹表 1”包含从“车 000001”到“车 006094”的全部轨迹信息，“轨迹表 2”则记录了“车 006095”到“车 012358”的轨迹信息，依此类推。下面以“轨迹表 1”中“车 000001”、“车 000025”为例进行说明。

“轨迹表 1”的第 2-8 行记录了“车 000001”的信息，它表示该车是一辆客车，在 2022 年 2 月 22 日至 2 月 27 日的六天时间内经过该高速路区间共 1 次。这次旅程的入站点和出站点都不在所列的 15 个收费站里，知道该车这次旅程进入高速公路（即经过入站点）的时间是 23/2/2022 07:56:40，不知道这次旅程离开高速公路（即经过出站点）的时间。区间内门架记录的这次旅程的信息是该车在 23/2/2022 08:37:17 经过门架 7，在 23/2/2022 08:46:19 经过门架 8，在 23/2/2022 08:56:21 经过门架 9，在 23/2/2022 08:56:22 经过门架 10。

“轨迹表 1”的第 213-233 行记录了“车 000025”的信息，它表示该车是一辆货车，在 2022 年 2 月 22 日至 2 月 27 日的六天时间内经过该高速路区间共 4 次。第 1 次旅程开始进入高速公路的时间是 22/2/2022 14:08:45，入站点不在所列的 15 个收费站里，这次旅程在 22/2/2022 15:13:04 以 MTC 收费的方式经过收费站 6 离开高速公路。这次旅程被区间内的门架 7、8、9 检测到，记录信息是该车在 22/2/2022 14:46:36 经过门架 7，在 22/2/2022 15:02:16 经过门架 8，在 22/2/2022 15:10:06 经过门架 9。第 2、3、4 次旅程类似。

三：建立模型

一、一日内小时客流分布特征

轨道交通一日内小时客流随人们的生活节奏和出行特点而变化。其通常是夜

间少，早晨渐增，上班和上学时达到高峰，午间稍减，傍晚下班和放学又是高峰，此后又逐渐减少，午夜最少。因此，轨道交通一日内小时客流通常是双峰型，这种规律在国内外的轨道交通线路上几乎都是一样的，只是程度不同而已。反映轨道交通线路分时客流不均衡程度的系数可按下式计算。

$$a_1 = P_{\max} E H_t = 1 P_t / H$$

式中， a_1 为单向分时客流不均衡系数； P_{\max} 为单向高峰小时最大断面客流量

(人)； P_t 为单向分时最大断面客流量(人)； H 为全日营业小时数(个)。单向分时客流不均衡系数值恒大于 1。 a_1 趋向于 1 表明分时客流分布比较均衡， a_1 越大表明分时客流分布越不均衡。当 $a_1 \geq 2$ 时，表明分时客流的不均衡程度比较大。位于市区范围内的地铁、轻轨线路的 a_1 值通常为 2 左右，而通往远郊区市域轨道交通线路的 a_1 值通常大于 3。

在一日内小时客流不均衡程度较大的情况下，为实现运营组织的经济合理性，可考虑采用小编组、高密度列车开行方案。小编组、高密度与大编组、低密度两种列车开行方案的分时列车运能不变，但在客流低谷时段，小编组、高密度列车开行方案具有既能提高客车满载率，又不降低乘客服务水平等优点。

应该指出，小编组、高密度列车开行方案只是在一定的客流条件下才是可行的。分时客流不均衡程度比较大是一个条件，线路的客流量较小、尚未达到设计客流量是另一个条件。在线路客流量较小的情况下，由于在客流低谷时段列车开行数较少，会使乘客候车时间延长，降低乘客服务水平；而如果为保持乘客服务水平，在客流低谷时段增加列车开行数，则又会使车辆满载率降低，产生运营不经济的情形。小编组、高密度方案的优点是既不增加列车运能，又能提高列车密度，从而解决了上述两个问题。但如果线路客流量已经较大，甚至接近设计客流量，采用小编组、高密度方案，在低谷时段增开列车问题不大，但在高

开列车则会受到线 通过能力的限制。

二、一周内全日客流分布特征

由于人们的工作与休息是以周为循环周期进行的，所以这种活动规律性必然要反映到一周内全日客流的变化上来。在以通勤、通学客流为主的轨道交通线路上，双休日的客流会有所减少；而在连接商业网点、旅游景点的轨道交通线路上，双休日的客流又往往会有所增加。与工作日的早、晚高峰出现时间比较，双休日的早高峰出现时间往往推迟，而晚高峰的出现时间又往往提前。另外，星期一与节假日后的早高峰小时客流和星期五与节假日前的晚高峰小时客流都会比其他工作日的早、晚高峰小时客流要大。

根据全日客流在一周内分布的不均衡和有规律的变化，轨道交通运营部门常在一周内实行不同的全日行车计划和列车运行图，以适应不同的客运需求和提高运营的经济性。

三、季节性或短期性客流分布特征

在一年内，客流还存在季节性的变化，如由于梅雨季节和学生复习迎考等原因，6 月份的客流通常是全年的低谷。另外，在旅游旺季，流动人口的增加也会使轨道交通线路的客流增加。短期性客流激增通常发生在举办重大活动或遇到天气骤然变化的时候。对季节性的客流变化，可采用实行分号列车运行图的措施来缓和运输能力紧张的状况。当客流在短期内增加幅度较大时，运营部门应针对某

些作业组织环节、某些设备的运用方案采取应急调整措施，以适应客运需求。

四、车站高峰小时客流分布特征

车站高峰小时客流是确定车站设备容量或能力的基本依据。车站高峰小时客流分析，首先应确定进出站高峰小时的出现时间，其次才是分析客流量的大小。此外，还应分析客流的发展趋势。随着轨道交通新线的投入运营和既有轨道交通线路的延伸，高峰小时进、出站客流会发生较大的变化。而车站吸引区内住宅、商业和娱乐等方面的发展也会使高峰小时进、出站客流发生较大的变化。研究表明轨道交通车站高峰小时客流具有以下特征：

1、车站客流的进、出站高峰小时出现时间与断面客流的高峰小时出现时间通常不相同。

2、各个车站客流的进、出站高峰小时出现时间通常不相同。

3、同一车站客流的进、出站高峰小时出现时间通常不相同。

4、同一车站工作日客流与双休日客流的进、出站高峰小时出现时间通常不相同。

五、车站超高峰期客流分布特征

为了避免因超高峰期内特别集中的客流影响乘客顺畅地进出车站，甚至影响列车的正常运行秩序，在确定车站设备容量或能力时，有必要适当考虑车站客流在高峰小时内分布的不均衡性。车站超高峰期的客流强度可用超高峰系数来反映，它是单位时间内的超高峰期平均客流量与高峰小时平均客流量的比值。超高峰系数一般可取值为 1.1~1.4。对终点站、换乘站和客流较大的中间站通常取高限值，而其余车站则可取低限值。

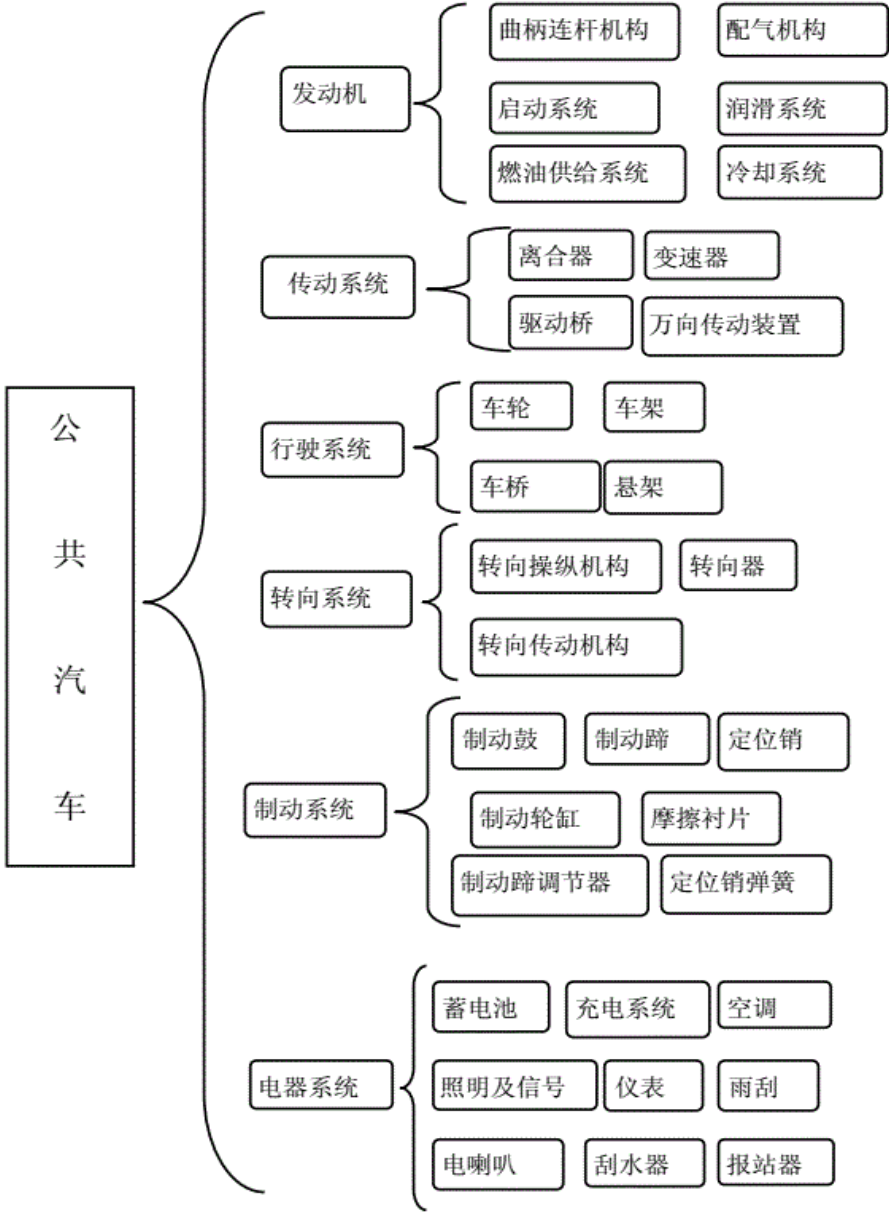
随着我国城镇化速度的加快，公共汽车对人们的日常出行的地位也越来越重要。公共汽车的运营程度直接关系到城市的发展程度。公共汽车作为城镇市民出行的主要交通方式之一，它的运行可靠性和运行效率受到相关部门的高度关注。由于公共汽车的结构与功能同趋复杂，性能日臻完善，价格更加昂贵，使之成为生产过程中需要付出巨大的经济代价才能维持使用的一种设备。为了保持和恢复公共汽车的性能而需要进行维护和修理所花费的人力、物力正以惊人的速度增加。

本文通过对某型公共汽车的日常维修活动的统计分析，主要是针对日常维修活动中的维修系统、维修频度、修理时间等方面的数据分析，从而找出各系统中的哪些部件是易损件即可靠性较低，就要缩短它的预防维修周期，增加他们的保养次数；找出哪些部件是不易损坏的部件即可靠性高，就不需要频繁的对其进行预防维修活动，减少它们的保养次数。通过统计分析，我们最根本的目的还是要提高车辆的使用可靠性能，使用经济性性能，避免维修过度、不合理维修所造成的经济损失，为公共汽车运营公司的预防维修制度的建立指明方向。

关键字：公共汽车；维修系统；维修频度；修理时间；可靠性

2.3 公共汽车主要系统的分类

公共汽车是一个复杂的机械系统，它有很多的系统组成，失效模式也多种多样，本文只分析公共汽车的发动机、传动系统、行驶系统、制动系统和电器系统。将其做成结果简图，如下图 2.1 所示：



ETC 门架系统日常检查主要包括 ETC 门架系统设备外观及功能测试及成功率统计测试。

一、ETC 门架系统设备外观及功能测试

1、检查范围

主要对车道控制器、RSU、车牌图像识别设备（含补光设备）、高清摄像机、固定枪机、供电设备、机柜等设备的外观及功能进行测试。。

2、检查内容

（1）设备安装及外观

ETC 门架系统的车道控制器、RSU、车牌图像识别设备（含补光设备）、高清摄像机、固定枪机、ETC 门架服务器、供电设备、工业交换机等设备的安装符合要求。设备及配电箱内布线整齐、固定可靠、余留规整、标识清楚。

（2）设备状态监测

连接测试用计算机与门架交换机，在 Linux 系统下运行部提供设备状态监测软件，在监测软件界面逐项查看车道控制器、RSU、车牌图像识别等设备的在线状态及工作状态。断开车道控制器或 RSU 电源，查看监测软件是否提示故障，并及时发出报警信息。

（3）高清视频监控设备功能

- a. 查看断面高清摄像机的视频图像，图像中应能有效辨别车牌，视频监控可覆盖门架所处断面。
- b. 查看视频中的字符叠加信息，应包括时间、摄像机编号、门架编号等信息。
- c. 对图像进行评分，不低于 4 分（采用五级损伤制评分）。

（4）高清车牌图像识别设备功能

- a. 在收费分中心查看过车产生的图像流水记录，图像流水记录应包含车牌颜色、车牌号码和时间、门架信息、车道号、方向及车辆图像信息。
- b. 查看 300 万像素与 900 万像素抓拍单元拍摄的通行车辆全景图片，应包含机动车前部（或后部）全貌、号牌、颜色等信息。

（5）机柜

对机柜的门禁控制功能、柜内照明、空调、火灾报警及防盗和防破坏设施进行功能检查。

(6) 输入输出电压

用万用表测量，供电系统进线端电压范围与设备供电端电源电压范围是否满足要求。

(7) 后备电源

断开市电，查看设备是否不间断工作，后备电源应无切换时延；

图一

表 1 高速公路收费数据属性信息列表

属性标识	标识含义	属性标识	标识含义
WasteID	流水号	EnStation	入口站号
ExTime	出口时间	EnLane	入口车道编号
ExStation	收费站编号	EnTime	入口时间
ExLane	车道编号	EnVC	入口车型
LaneType	车道类型	RealFlagStationInfo	实际经过标识站信息
VC	最终车型	FlagStationInfo	应经过标识站信息

对于任意路段 D 在推断交通量之前，还要求经过路段 D 的经由路径 L 中每辆车通过路段 D 的时间。本研究使用匀速平均法来确定每辆车所经过路径中每一个路段的时间。匀速平均法是假设车辆进入高速路网后以匀速行驶。来计算该车辆通过各个路段的时间。因为车辆入站时间、出站时间、以及所经路线的行驶总距离为已知，所以可以容易计算出匀速行驶的速度等参数再通过各个路段的长度和与入口之间的距离推出车辆经过每个路段的时间。设 AB 是经过路段 D 的最短路径集合 S 中的任意一个进出站组合。车辆 i 从 A 站进入高速的时间为 TA 车辆从 B 站驶出高速的时间为 TB 车辆从站点 A 到路段 D 的路程为 LND 车辆由进入站 A 到驶出站 B 的最短路径的路程为 LN 则根据匀速假设可以求出车辆 i 通过路段 D 的时间为

$$TAB(i)=T/+(TA-TB)*LNA/LN \quad (1)$$

其中 LN 为由 A 到 B 的最短路径经过的所有路段集合 SAB 中的路段距离之和。 $LN=\sum LN(k \in SA)$;LN 为由 A 到 B 的最短路径经过路段 D 之前的所有路段集合 SAD 中的路段距离之和 $LN=\sum LN(k \in S)$ 。若以路段 D 的中点的断面交通量代表通过路段 D 的交通量，则在计算 LN 时需要多考虑路段 D 本身距离长度的一半。设路段的长度为 LND，则 $LN=\sum LN_s \times LN(k \in SA)$ 。

计算任意路段 D 在任意时段 T(TT) 内的交通量 V(T) 本质就是计算经由路段 D 的最短路径的集合 S 中对于每一组进出高速的 AB 组合，在时段 T 内经过路段 D 的车

辆数量的累加。用 N 代表累计个数增量 N 数值取 1 ΣN 代表满足条件的车辆 i 的累计总数。设时间段 T 的起始时间为 T 终止时间为 $TVAL(T)$ 为由 A 进入高速、由 B 驶出高速、且在时间段 T 内经过路段 D 的车辆数, 则有

$$VAB(T)=\sum N(i \text{ew 满足 } T < TDAB(w) < T2) \quad (2)$$

$VAB(T)$ 代表进出组合高速站点组合 AB 在时间段 T 经过路段 D 的车辆数。对于单一路段 D 可以根据 S 中的每一组合 AB 的所有车辆进出高速时间和距离, 计算其经过 D 的时间, 从而求得 $VAB(T)$ 。将集合 S 中的所有组合的车辆数累加, 从而得到路段 D 在时间段 T 中的交通量。

$$VD(T) \Sigma VAB(T) (ABSD) \quad (3)$$

若考虑路段经过性判断, 可将上式写为

$$V, (T) \Sigma (VDB(T) \times JB) \quad (4)$$

[摘要]高速公路联网收费信息系统中包含了海量的公路交通信息[11, 通过收费数据挖掘得到长时间断面交通量的算法已经实现, 但是针对短时间粒度的断面交通量算法尚属空白。而短时间粒度、分车型交通量是公路交通情况调查系统采集数据的主要指标。本研究松建了针对短时、分车型数据的挖掘分析方法, 充分利用高速公路收费数据, 采用车辆动态模拟的分配方式, 计算全路网各路段短时间粒度的动态交通量。将收费数据推算结果与自动化交通情况调查站监测结果对比, 高速公路网日交通量平均误差为 8.87%, 小时交通量误差为 9.02%。本文提出的方法实现了高速公路网全域、高效的自动化交通量推算, 对于推动公路交通基础信息采集技术发展及提升多源大数据分析能力建设具有重要意义。

[关键词] 交通量 • 收费数据 • 交通调查 • 车型 • 推算方法

在进行分析建模前, 首先要对收费数据进行数据清洗。一方面删除不符合常规的、或明显错误的信息, 另一方面, 将与推算断面交通量相关的信息进行读取与分析, 包括进出高速公路的收费站编号、出入时间, 标识站、车型信息等, 而忽略其他数十种无用属性信息。用于数据处理的收费数据属性信息如表 1 所示。

3 推算方法

推算方法包括路网拓扑构建和交通量推算两部分。

3.1 路网拓扑构建

路网拓扑构建是将收费数据与实际路网相结合、进行动态模拟分配、推算路段断面交通量的基础。路网拓扑建构需要考虑高速公路网路段间的拓扑关系, 收费站和路网的关系, 以及收费站之间的关系、标识站位置等。在进行路网拓扑构建时, 还要考虑必要的路段属性信息, 如每条路段的长度距离, 用于进行下一步建模。路网采用近邻表的方法进行构建。路网拓扑节点的建立采用虚实结合的方法。四川路网中的收费站、标识站作为路网中的实节点, 而路段间相交的、但没有布设收费站或标识站的节点作为虚节点。近邻表实节点包括已建设 529 个收费站,

129 个标识站，结合路段相交点的虚节点，共有 830 各路网拓扑节点。网络拓扑节点编号的建立需要与收费数据中的节点编号相对应。在建立路网拓扑过程中，发现有些收费站中同一个收费站却配有不同的编号，需用一张单独表将这些编号不同而实体相同的节点一一对应，以便在处理收费数据中不产生歧义。

3.2 交通量推算

交通量推算分为三个主要步骤，分别为路段经过性判断、车辆通过时间判断和集成交通量。根据处理数据的时间，可以分为离线处理和在线处理两种情况。其中，路段经过性判断属于离线处理，即在分析收费数据前就进行的步骤车辆通过时间判断、集成交通量两个！属于在线处理，即与处理收费数据同步边步骤。路段经过性判断的方法是最短路径法。对于路网上任意两个收费站节点(标识站看作特殊的收费站，可以将车辆的实际路径描述为最短路径的集合)，用迪杰斯特拉(Dijkstra)算法计算两个收费站的最短路径，作为进出这两个收费站点的车辆的行驶路径。四川路网收费站 529 个，标识站 129 个，即共有 658 个网络实节点。

对于任意路段 D，在推断交通量之前，还要求经过路段 D 的经由路径 LAB 中每辆车通过路段 D 的时间。本研究使用匀速平均法来确定每辆车所经过路径中每一个路段的时间。匀速平均法是假设车辆进入高速路网后以匀速行驶。来计算该车辆通过各个路段的时间。因为车辆入站时间、出站时间、以及所经路线的行驶总距离为已知，所以可以容易计算出匀速行驶的速度等参数，再通过各个路段的长度和与入口之间的距离，推出车辆经过每个路段的时间。设 AB 是经过路段 D 的最短路径集合 SD 中的任意一个进出站组合。车辆 i 从 A 站进入高速的时间为 T_{iA} ，车辆从 B 站驶出高速的时间为 T_{iB} 车辆从站点 A 到路段 D 的路程为 LN_{AD} ，车辆由进入站 A 到驶出站 B 的最短路径的路程为 LN_{AB} ，则根据匀速假设，可以求出车辆 i 通过路段 D 的时间为： $TDAB(i) = T_{iA} + (T_{iA} - T_{iB}) * LN_{AD} / LN_{AB}$ (1)

其中， LN_{AB} 为由 A 到 B 的最短路径经过的所有路段集合 $SLAB$ 中的路段距离之和。 $LN_{AB} = \sum_{k \in SLAB} LN_k$ ； LN_{AD} 为由 A 到 B 的最短路径经过路段 D 之前的所有路段集合 $SLAD$ 中的路段距离之和， $LN_{AD} = \sum_{k \in SLAD} LN_k$ 。若以路段 D 的中点的断面交通量代表通过路段 D 的交通量，则在计算 LN_{AD} 时需要多考虑路段 D 本身距离长度的一半。设路段的长度为 LN_D ，则 $LN_{AD} = \sum_{k \in SLAD} LN_k + 0.5 * LN_D$ (kESLAD)。

计算任意路段 D 在任意时段 T(T_1, T_2) 内的交通量 $VD(T)$ ，本质就是计算经由路段 D 的最短路径的集合 SD 中，对于每一组进出高速的 AB 组合，在时段 T 内经过路段 D 的车辆数量的累加。用 N_i 代表累计个数增量， N_i 数值取 1。 $\sum N_i$ 代表满足条件的车辆 i 的累计总数。设时间段 T 的起始时间为 T_1 ，终止时间为 T_2 ， $VDAB(T)$ 为由 A 进入高速、由 B 驶出高速、且在时间段 T 内经过路段 D 的车辆数，则有

$VDAB(T) = \sum N_i$ ($i \in E_w$ 满足 $T_1 \leq TDAB(i) \leq T_2$) $VDAB(T)$ 代表进出组合高速站点组合 AB 在时间段 T 经过路段 D 的车辆数。对于单一路段 D，可以根据 SD 中的每一组合 AB 中的所有车辆进出高速时间和距离，计算其经过 D 的时间，从而求得 $VDAB(T)$ 。将集合 SD 中的所有组合的车辆数累加，从而得到路段 D 在时间段 T 中的交通量。

$$VD(T) = \sum VDAB(T) (AB \in SD)$$

(3)

若考虑路段经过性判断, 可将上式写为

以上是以路段为对象考虑的交通量推算方法。在四川路网中根据节点划分了几百个路段。在计算不同路段的交通量时, 会涉及同一辆车在不同时间行驶到不同路段的情况。为了减少计算量, 在计算不同路段时不重复计算每辆车的基本信息, 本研究采用以每辆车为对象的动态模拟的方法标定其行驶轨迹, 从而计算路网路段的交通量。通过收费数据进出收费站的时间, 推算车辆在高速公路网中经过每一个路段中点的时间, 进一步推断该路段的交通量。车辆 i 从 A 站进入高速的时间为 T_{iA} 。

推断结果

采用上述算法对四川省高速公路 2018 年 1-6 月的收费数据进行分析, 并与已建设的 70 个自动化交调站采集的数据进行校验。图 1 为可视化的四川工作日晚高峰流量空间分布图, 可以看出, 成都周边路网车流量较大, 其他路段流量相对较小。

图 2 为可视化的路网各路段车速分布。可以看出, 四川省高速路网中, 成都环线的路网较为拥堵, 速度较为缓慢, 而从成都环线分向各个放射线高速的车速较为通畅。对于行驶路径依次经过放射线和成都环线的车辆, 实际车速并非理想的匀速行驶, 从而按照匀速平均法计算的路段交通量会有微小偏差。但根据数据分析, 在计算交通流量时, 匀速行驶假设基本可以满足以小时为单位的短时间粒度交通量计算要求。

将由收费数据转化的交通量和交调站点直接采集的断面交通量在同一空间、时间范围下进行比对。图 3 为基于全网具备交调站点的路段的平均流量与相对应的收费数据转化的平均流量的比较。从图中可以看出, 二者变化趋势一致, 高速公路网日交通量平均误差为 8.87%, 小时交通量误差为 9.02%。对于单一路段, 高速公路日交通量误差为 13.80%, 小时交通量误差为 17.20%。分车型看, 单个站点的中小客车小时交通量误差在 12% 左右。车型校验结果如图 4 所示。

5 结论

高速公路收费数据中蕴含了三富的动态路网交通信息。本文研究了收费数据推算交通量的数据挖掘方法, 首次采用动态模拟分配的方法得出了短时间粒度的全网各路段流量; 且经过车型匹配算法, 得出基于公路交调系统的分车型交通量数据。通过收费数据转换计算短时的路段交通量数据, 能够掌握全路网各路段交通情况, 可以作为公路交通情况调查数据的有益补充, 可为后续相关研究与应用奠定基础。

主要参考文献

[1] 付鑫, 王建伟, 袁长伟. 基于高速公路收费数据库的断面交通量计算方法[J]. 交通运输研究, 2006(6): 150-153.

[2]交通运输部规划研究院. 2017 年国家干线公路交通情况分析报告[R]. 2017.

[3]胡闰秀, 李梦雪. 基于高速公路收费数据的交通量调查[J]山东交通学院学报, 2016, 24(1): 21-26.

[4]孙忠辉, 从联网收费数据中提取管理信息方法探讨[J]中国交通信息化, 2009(8): 70-72.

[5]钟足峰, 刘伟铭, 基于联网收费数据预测交通流量的实现[J]中国管理信息化, 2009, 12(2): 59-61.

短期交通流量预测摘要交通流量是一种对于一段时间内在某个路口内通过的交通实体量, 在现在的社会中, 智能运输系统等交通理论的研究已经渐渐成为发达国家的研究对象, 而交通流量预测分析是其中的核心研究之一。所以, 对于交通流量的预测成为叩开智能交通系统大门的最有力的那一把钥匙。

在前面, 我们首先面临的一个问题是对于数据的处理。题目以 15 分钟为一个时间段来测量交通流量, 一共有三天的数据, 应该有 288 个数据, 但是题目只给出了 276 个。另外, 在数据中还有两个为负的数据。面对缺失数据和异常数据, 我们分别使用了热卡插补法和平均值填补法来解决。

然后在进行预测时, 我们分别使用了不同的软件来建立不同的预测模型。首先我们使用了灰色预测 GM 软件来进行灰色模型的预测, 在预测前, 我们先用模型和前两天的交通流量来预测第三天的交通流量, 然后将第三天的真实交通流量与预测交通流量进行相关性检验, 检验通过后, 再用于预测第四天的交通流量, 最后评价模型的好坏。

接着, 我们使用了 spss 软件来进行回归分析模型的预测。在预测之前, 我们需要先对数据进行相关性检验, 若没有相关性, 则回归方程会没有意义。接下来, 通过对回归方法的决定性系数检验和方差分析检验, 得到最合适方法。之后再再进行第四天的预测及预测结果的评价。

然后, 我们使用了 metlab 软件来实现 BP 神经网络模型的预测。BP 神经网络的实质是用已给出的数据来推出需要的数据, 并将新预测出的数据重新返回输入中, 得到误差, 一直重复, 直到误差到达合理的范围内。在预测之前, 我们先得出了误差在合理范围, 并且看到已给出数据的真实值与预测值得对比。在确保模型是可用的之后, 在进行预测与预测结果的评价。

最后, 我们使用了 eview 软件来进行时间序列的预测。时间序列预测要求数据必须是平稳的, 所以在预测前, 先要对数据进行 ADF 检验, 在检验通过后, 才能进行预测, 得到预测后的表达式和残差。在最后, 还必须对残差进行分析估计。这样之后, 对模型进行评价。

在本文的最后, 我们进行了进一步的讨论和改进, 对四种预测方法进行了一个比较, 判断出那个模型是最适合这个题目的。并且对文章中所涉及的模型进行推广, 使其更便于运用于生活实际中。

1 问题重述

一 1.1 问题背景

交通流量指的是在一段时间内通过道路某一地点、某一断面或某一车道的交通实体数。随着交通基础设施建设和智能运输系统的发展, 交通规划和交通诱导已

成为交通领域研究的热点。对于交通规划和交通诱导来说，准确的交通流量预测是其实现的前提和关键。

交通流量预测根据时间跨度可分为长期交通流量预测和短期交通流量预测，长期交通流量预测以小时、天、月甚至年为时间单位，是宏观意义上的预测；短期交通流量预测一般的时间跨度不超过 15 分钟，是微观意义上的预测。短期交通流量预测是智能运输系统的核心内容和实现其智能化功能的基础平台。短期交通流量预测具有高度非线性和不确定性等特点，并且同时间相关性较强，研究表明，城市交通路网中交通路段上某时刻的交通流量与本路段前几个时段的交通流量有关，并且交通流量具有 24 小时内准周期的特征。

1.2 问题研究

现有 3 天的交通流量数据，假设从第 1 天 0 时 15 分开始，每隔 15 分钟记录一次该段时间内的交通流量，预测出第 4 天的交通流量，并指出模型的优缺点。

2 我、二问题分析

题目要求我们根据已给出的三天的数据来预测第四天的交通流量，并且评价判断模型的优缺点。

首先，题目给出的是三天的数据，以 15 分钟为一个截点，应该有 $3 \times 24 \times 4$ 个数据，但实际只有 266 个数据。另外，在数据中有负数的情况，按照该题的实际情况而言，不可能出现为负的情况，交通流量不可能为负。所以，首先要对缺失数据进行处理，我们使用了单一插值法中的热卡插补法来补齐缺失数据和异常数据。

然后我们需要进行预测第四天的数据，在这里我们只需要进行短期的交通流量预测。在短期预测中，我们以原始的 15 分钟为一个时段，预测未来一天的交通流量。在这里我们一共运用了四种预测方法，分别是灰色预测模型，回归分析预测方法，时间序列和神经网络。

最后，通过对每种方法预测结果的分析与判断，总结出每种方法的优点和缺点。

三、符号说明

3 符号说明

符号	含义说明
$x(l)CD, X_{(1)}(2), X_{(1)}(3)$	第一二三天中，不同时间点的交通流量
i	一天中，以 15 分钟为时间段的时间序
	列编号
a	权值
X_j	时间数列
Y_i	交通流量数列
\bar{x}	时间的平均值
\bar{y}	交通流量的平均值

四 模型假设

1. 假设题目所给的所有数据都是真实有效的。

2. 假设在这四天中并没有特殊的会聚集人群的事件发生。
3. 假设测量的误差小，对结论的影响程度低，甚至没有影响。
4. 假设灰色模型、MATLABSPSS 软件、EViews 软件选取的预测模型都是最好的。
5. 假设测量交通流量时，把测量路段的所有车辆看作一个点。
6. 假设随机选取检验的数据，对结论没有影响。

五 模型建立与求解

这道题目是一道关于交通流量的研究的问题，我们以某三天内的每 15 分钟一个时段的交通流量作为基础，在以 15 分钟时间段为时间窗宽的情况下，对第四天做出了预测，验证了模型的可行性，最后通过对模型结果的分析，评价模型的好与坏。

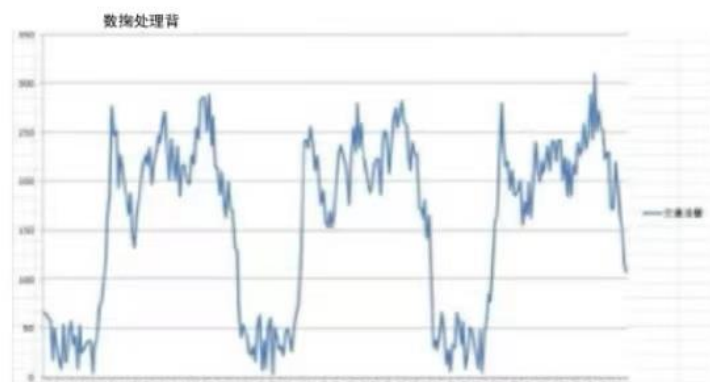
模型数据的处理

问题中要求我们根据前三天的数据预测出第四天的数据，前三天的数据应该有 288 个，但是实际给出的数据只有 276 个，且给出的数据中还存在有两个为负数的数据，在题目的现实意义中，交通流量根本不可能为负数。所以据判断，这组数据中存在着缺失值和异常数据，如何处理数据是一个重要的点。

Step1 缺失数据的处理

首先对于缺失数据来说，我们并不知道缺失的数据是随机缺失数据还是非随机缺失数

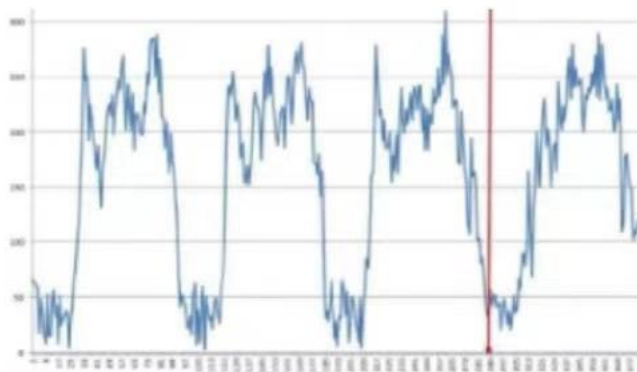
据，所以我们将数据点作图，观察曲线的趋势，折线图一如下：



从图中可以明显看出，这组数据有明显的周期性，所以可以大胆判断缺失的 12 个数据为第三天的 21 点到 24 点的数据。对此，我们选择了热卡插补法进行缺失值的填补，所谓热卡插补法 (Hotdeckimputation)，即对于一个包含缺失值的对象，在完整数据中找到一个与它最相似的对象然后用这个相似对象的值来进行填补。在这里，我们用与缺失值最相似的前两天同一时段的数据的平均值来代替缺失值。用替换掉缺失值的数据画折线图得到

	真实值	预测值
真实值 Pearson 相关性	1	.785 [*]
显著性(双侧)		.000
N	20	20
预测值 Pearson 相关性	.885 ^{**}	1
显著性(双侧)	.000	
N	20	20

运用给出的方程式以及对于残差的分析估计，得到第四天中每个时间段的交通流量的预测值，依旧将它与其它第三天的数据一起，画出它的折线图，如下：



这个题目要求我们根据前三天的数据来预测第四天的交通流量，没有规定使用何种预测方法。而我们为了保证预测结果的准确性，选择了不同的预测方法来进行预测，下面是4种不同的预测方法的结果：

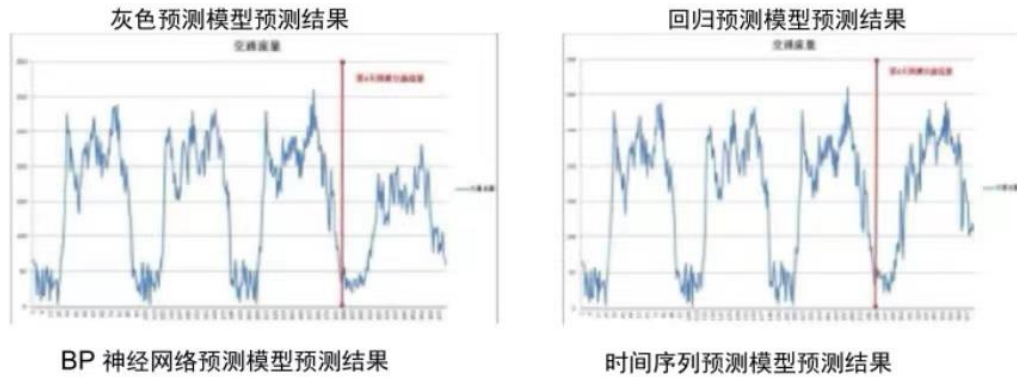
模型汇总

R	R 方	调整 R 方	估计值的标准误
	.894	.799	.792

自变量为 VAR00002 ANOVA

	平方和	df	均方	F	Sig.
回归		3			.000
残差		92			
总计		95			

自变量为 VAR00002



从四幅图中可以看出，第一个灰色预测模型的预测结果太过高于其他三天的值，第三个神经网络模型的预测结果又过于太低于其他三天的值，第二个回归模型的预测结果虽然很符合总体趋势，但在实际中，它的细节处还不够，而第四个时间序列模型的预测结果与前面几天的最相似。再结合在预测过程中，几个模型的操作难易程度，参数检验结果和结果的显示，我们认为，时间序列预测模型是最适合这道题的模型，可以用它来进行接下来的其他时候的预测。

模型的推广

1. 以上预测模型可以运用到很多预测和控制领域，尤其是时间序列、回归拟合模型。
- 2 . 以上模型还可以用于客流量预测， 降水量预测等更多地同类型预测问题。

8 参考文献

1. 王沁，时间序列分析及其应用，四川西南交通大学出版社，2008；
- 2 . 初连禹，杨兆升，基于神经网络的实时交通流量模糊预测；
3. 宫晓燕，汤淑明，基于非参数回归的短视交通流量预测与事件检测综合算法；
4. 邓聚龙，灰色系统理论教程，武汉：华中理工大学出版社，1990；
5. 韩超，基于时间序列分析的短期交通流量实时自适应预测，北京：北京工业大学，2004；
6. 邓志龙，李全，陈茜，基于灰色系统理论的短视交通流量预测，公路交通技术，2006, 2, 第一期：117——119 ；

8 附录