

文章编号:1001-7372(2023)11-0441-15

基于ETC数据的山区高速公路客货车 行驶速度特征研究

张 驰¹, 翟艺阳¹, 王 韩¹, 张 敏^{*2}, 袁华智³, 赵 霄⁴

(1. 长安大学 公路学院, 陕西 西安 710064; 2. 长安大学 运输工程学院, 陕西 西安 710064; 3. 兰州理工大学 土木工程学院, 甘肃 兰州 730050; 4. 四川智慧高速公路科技有限公司, 四川 成都 610000)

摘要:电子不停车收费(Electronic Toll Collection, ETC)系统作为智慧高速的重要组成部分,每天产生着海量实时的交通数据。面向ETC数据质量提升和行驶特征分析的需求,提出了一种基于ETC数据的高速公路客货车行驶速度特征分析方法,其包含数据优化、质量评估、行驶速度特征分析3个部分。在分析总结ETC门架及其数据特点的基础上,探讨了一种数据提取及优化方法。同时从数据规模、准确性、完整性3个方面,构建了用于ETC数据质量评估的量化指标体系;采用统计分析方法研究了不同路段客货车行驶速度特征的差异,进一步利用潜在类别分析(LCA)确定相似速度特性的车型类别。最后,广泛收集了西部某山区高速公路沿线ETC数据,实例验证了所提出的高速公路行驶特征分析的方法,并根据分析结果对高速公路限速方案提出建议。研究结果表明:提出的数据清理和质量评估方法能够较好地解决和评价当前ETC门架数据存在的缺陷;受路段的纵坡影响,客货车的行驶速度分布具有显著差异;基于行驶速度分布特征指标对各类车型行驶速度进行潜在类别分析,在平坡和下坡路段车辆行驶速度适用于4类别模型,在上坡路段车型行驶速度适用于5类别模型;对各路段限速方案提出根据各类别行驶速度特征调整限速值与细分车型限速的建议。研究结果有助于提升ETC数据的质量和 application 价值,同时为开展高速公路运行态势监测和分车型动态限速方案制定提供了理论和方法支持。

关键词:交通工程;ETC数据;数据挖掘;行驶速度;数据治理方法

中图分类号:U491

文献标志码:A

Research on Driving Speed Characteristics of Passenger and Freight Vehicles on Mountainous Highways Based on ETC Data

ZHANG Chi¹, ZHAI Yi-yang¹, WANG Han¹, ZHANG Min^{*2}, YUAN Hua-zhi³, ZHAO Xiao⁴

(1. School of Highway, Chang'an University, Xi'an 710064, Shaanxi, China; 2. School of Transportation Engineering, Chang'an University, Xi'an 710064, Shaanxi, China; 3. School of Civil Engineering, Lanzhou University of Technology, Lanzhou 730050, Gansu, China; 4. Sichuan Smart Expressway Technology Co. Ltd., Chengdu 610000, Sichuan, China)

Abstract: Amidst smart highways, the Electronic Toll Collection (ETC) system generates a large volume of real-time traffic data. To improve the ETC data quality and driving characteristic analysis, a method to analyze the driving speed characteristic of passenger and freight vehicles on

收稿日期:2023-05-17

基金项目:国家重点研发计划项目(2020YFC1512005);陕西省自然科学基金项目(2023-JC-YB-391);

四川省科技计划项目(2022YFG0048);四川省交通运输科技项目(2022-ZL-04);国家自然科学基金项目(52362050)

作者简介:张 驰(1981-),男,四川宜宾人,教授,博士研究生导师,工学博士,E-mail:zhangchi@chd.edu.cn.

* 通讯简介:张 敏(1981-),女,广西柳州人,副教授,工学博士,E-mail:minzhang@chd.edu.cn.

highways based on ETC data is proposed, which includes three steps: data optimization, quality assessment, and driving speed characteristic analysis. Based on the summary of ETC gantry and its data characteristics, data extraction and optimization methods were discussed. Furthermore, a quantitative index system to assess ETC data quality was constructed from three aspects, namely, data scale, accuracy, and completeness. Statistical analysis methods were used to study the driving speed characteristic differences of passenger and freight vehicles on different highway sections, and latent class analysis (LCA) was conducted to categorize vehicles with similar speed characteristics. Finally, extensive ETC data was collected along a mountainous highway in western China to empirically verify the highway driving characteristics analysis method proposed in this paper. The results show that the proposed data cleansing and quality assessment method effectively addresses and evaluates the existing flaws in the ETC data, and recommendations for highway speed limit adjustments were made based on the analysis results. Owing to the longitudinal slope of the road segment, the speed distribution of passenger and freight vehicles show significant differences. Using driving speed distribution feature indicators, LCA was performed on the driving speeds of different vehicle types. For flat and downhill road sections, vehicle speeds fit a four-class model, while uphill sections are consistent with a five-class model. Recommendations were made for speed limit adjustments in each road section, based on the speed characteristics of each class, and for subdividing speed limits by vehicle types. These results contribute to enhance the application and quality of ETC data and offer theoretical and methodological support for highway operational status monitoring and dynamic speed limit setting by vehicle types. The research findings help to enhance the application value and quality of ETC data. Additionally, they provide theoretical and methodological support for conducting highway operation status monitoring and formulating dynamic speed limit strategies based on vehicle types.

Keywords: traffic engineering; ETC data; data mining; travel speed; data governance

Received 17 May 2023

Funding: National Key R & D Program of China (2020YFC1512005); Natural Science Basic Research Program of Shaanxi Province (2023-JC-YB-391); Sichuan Science and Technology Program (2022YFG0048); Construction S & T Project of Department of Transportation of Sichuan Province (2022-ZL-04); National Natural Science Foundation of China (52362050)

* Corresponding author. *E-mail address:* minzhang@chd.edu.cn.

<https://doi.org/10.19721/j.cnki.1001-7372.2023.11.024>

0 引言

随着中国高速公路网络的快速发展和完善,电子不停车收费(ETC)系统实现全网运营^[1]。中国ETC系统作为全球最大的高速公路无线数据通信网络,全天候不间断服务,门架集成传感器每天产生近十亿条ETC交易大数据及大量路网感知数据。而传统的数据采集技术受限于设备特点、传输问题、交通流状态及天气状况^[2]等因素,获取数据内容不全且数据量有限。山区高速公路常受地形限制,且不同车型的速度特征存在差异,易存在交通安全隐患。因此,有效利用ETC系统获取的监测更广、覆盖

更全面的路网感知数据,分析山区高速公路路网分车型行驶速度特征,有利于进一步保障道路交通安全。

山区高速公路地质条件复杂,分析路段处的行驶速度特性能够直接反映出路段上车辆运行的动态规律^[3],研究多从线形设计^[3-5]、构造物设置^[6-7]、速度管理^[8-10]及不良天气^[11-12]层面挖掘山区高速公路行驶特征,为山区高速公路规划设计、运营管理提供理论依据。随着山区高速公路交通流中大型车的混入率逐渐增大,基于真实、可靠、准确的路网感知数据,进一步分析山区高速公路客货车行驶速度特征,成为数字化与智能化交通运营管理的必然要求。

要实现高速公路路网分车型的行驶速度精准管

理,需要获取实时、精确、足量的交通流数据,数据质量的好坏与时效性将直接影响路段交通特征分析的结果。现有交通流数据采集技术在数据利用驱动下已由人工采集转变为全自动设备采集,其主要包括固定式无标识数据采集、固定式有标识数据采集和移动式数据采集^[13]。固定式无标识数据采集设备^[14]包含线圈检测器、地磁检测器与微波检测器,能够获得路段速度、流量及占有率等交通参数^[15]。固定式有标识数据采集设备包含蓝牙识别检测器、射频识别检测器与视频识别检测器等。移动式数据采集方法是通过持续移动的设备如 GPS^[16-18]、移动通信设备^[9]等来获取交通数据。传统的固定式检测器数据由于设备安装时限较长且不易检修,导致数据获取质量低。因此,基于车检器数据的交通运行分析易受地点监测基本特性和布设地理位置的客观限制^[19]。基于移动式数据的交通特性分析方法是最近的热潮,但仅处于发展阶段,尚未成熟。^[2]

ETC 系统作为射频识别检测器,相较于其他数据采集系统,其数据样本量大,覆盖全面,记录连续且时效性强,在用于挖掘车辆行驶速度信息与分析路段交通组成方面是一种较好的手段。通过提取 ETC 路网感知数据,可以实现:了解交通流特征参数^[20],监测交通状态^[21];优化收费车道布设,提高收费效率^[22];获取实时交通信息,预测行程时间,选择最优出行方案^[23-25]。ETC 数据提取及利用能够为山区高速公路客货车行驶速度特征分析提供可靠来源。

综上,ETC 数据在行驶速度特征挖掘方面具有巨大的价值潜力,但是仍需要进一步完善数据质量和分析方法。因此,本研究利用 ETC 数据开展不同车型行驶速度特征研究,提出了 ETC 数据提取、处理、优化的分析方法,构建 ETC 数据量化指标体系验证数据质量。依托中国西部某高速公路近 6 个月的 ETC 数据,统计分析不同车型行驶速度差异,并采用潜在类别分析确定相似速度特性的车型类别。研究结果可以为分车型动态限速措施、公路交通态势监测提供依据,能够为高速公路不良天气、交通拥堵等特殊事件下的分级速度控制策略提供参考。

1 现行 ETC 门架系统数据分析

1.1 ETC 门架数据特点

(1)结构化数据

ETC 数据是具有高度组织性和整齐格式化的结构化数据,有利于数据存储、读取、处理和深入挖掘。以中国某高速 ETC 数据为例,其具有“中心名

称”、“门架名称”、“门架顺序编号”、“桩号”等固定的记录格式和存储类型,并允许以数据库表格、CSV 等格式进行输出。在研究车辆速度特征时,结构化数据方便快速定位所需数据门类。同时,基于数据输出格式,直接利用所需软件进行数据统计与数据参数化或非参数化关系构建,并将结果进行可视化展示。

(2)单条数据维度高

该高速公路单条 ETC 门架计费交易数据包含 161 条字段,ETC 门架牌识别数据包含 32 条字段。每一辆车通过 ETC 门架一次,产生的需要上传的数据就包含 193 条字段,其数据内容覆盖面广,数据维度高。在进行交通运行特征研究时,若所研究数据出现异常,研究结果则随之出现偏差。对于高维度的 ETC 数据,一方面,可利用与研究数据强相关的数据门类修正结果偏差;另一方面,若原始数据不存在异常,则还可以基于强相关的数据门类佐证研究结果。

(3)整体数据体量庞大

根据《收费公路联网收费技术标准》要求,ETC 系统需要满足 7 d×24 h 不间断运行的能力,加上 ETC 门架在空间范围布设甚广,因此,一条高速公路的 ETC 门架系统每天产生的数据是巨大、海量的。如此庞大的数据量,意味着从时间纵深和空间广度上,数据可被挖掘的深层次局部细节更加丰富,这也是明晰车辆速度特征的关键因素之一。但大多数数据都是通过 Oracle 数据库进行储存,当提取数据时,对于海量数据的遍历、清洗、整理具有一定的挑战性。

1.2 ETC 门架数据问题

1.2.1 数据缺陷总结

ETC 数据虽然具有数量基数大、覆盖内容广、可挖掘价值高的优势,但由于自身设备限制、邻道干扰、恶劣天气影响等,仍存在数据异常的现象。根据现有数据研究与高速实际情况,从 ETC 数据结构进行总结,目前常见的问题包括以下几点^[12,19,26-27]。

(1)数据失真:数据失真是指所获得 ETC 门架数据超出阈值、不满足公认的交通原理和规则或偏离历史整体趋势。例如,出现多个“默 A00000”牌号,这是因为违章车辆被 ETC 门架统一标记,还有应急救援车、警车等也是如此;ETC 门架记录的时间与行程差不匹配,这是由于车辆在通过第 1 个门架时,系统漏读数据,在经过第 2 个门架时,数据延迟读取补记入了系统;行驶方向与门架顺序编号不符,其原因为邻线门架错误读取数据;通过相邻门架的时间间隔过长,这是由于门架间存在服务区或发

生车辆故障。数据失真往往会影整体数据的品质,使得车辆区分、速度计算与分析难度增加。

(2)数据重复:如果 ETC 门架数据中存在完全相同的多行数据,则此 ETC 门架数据存在数据重复。例如,左右线门架的数据记录中同时有某车信息,这是由于左右线门架相邻近,导致 OBU 与左右线门架同时进行信息交换造成的;同一方向多条数据中门架名称相同,门架桩号相同,通过门架时间数据间隔非常短或者完全相同,其主要原因是车辆经过门架时,同一门架重复读取,导致 OBU 与门架信息交换重复 2 次。数据重复会导致研究对象的多次统计,从而影响不同车辆的权重占比,导致运行特征的统计分析不准确。

(3)数据缺失:如果 ETC 门架数据中存在一个或多个属性值为空值,则此 ETC 门架数据存在数据缺失。其主要表现形式为单点数据缺失与连续数据缺失,主要原因为车辆的卡片安装方式不规范、卡片性能存在问题、门架读卡器间歇性错误、同一时间门架通过的车辆数较多、重大节日小客车免费通行、部分车辆未安装 ETC 和恶劣天气影响等。数据不同程度的缺失都会导致 ETC 车辆流量统计数据不准确,从而影响最终的交通特征分析结果。

1.2.2 数据缺陷成因

(1)门架分布问题

高速公路 ETC 门架一般在出入口匝道、互通立交的端部 1.5~3.0 km 处,省界等地段均有设置,其在高速公路空间上分布广泛,整体数量可观。但是在一些特殊区域(如服务区前后)一般不设置门架系统,这会导致车辆运行特性统计与计算出现问题。

(2)门架设备异常

ETC 门架数据的获取主要通过 ETC 门架入口处的天线与车载单元之间的专用短程通讯(DSRC)实现,但 DSRC 技术存在重复、延迟、缺失读取等问题。同时,ETC 门架系统全天候不间断运行,随着使用年限增加,以及受到极端的天气影响,使得设备易出现损耗或故障,导致 ETC 数据可能出现问题。因此,对于设备的维护和设备系统的完整性与健康监测至关重要。

1.3 ETC 门架数据处理优化

随着 ETC 门架的普及,已有学者关注到其数据存在缺失、重复等问题^[28-29]。然而,目前 ETC 数据的处理优化研究是以统计遍历思想^[31-33]为主,整体遵循数据筛选重构、异常数据优化的思路,其中异常数据优化又按照剔除失真数据、删除重复数据、补全

缺失数据的步骤进行。

目前,学者们普遍认为利用 ETC 数据感知整体路网时,提出的数据处理优化方法在对 ETC 数据进行处理的过程中还存在一些问题^[30],主要有以下几方面:①数据采集异常,受限于设备和其他技术,海量 ETC 数据中常存在较多的数据重复、缺失、记录错位等问题^[31-33];②数据分析较难,ETC 数据的分析和挖掘需要大量的计算资源和专业技术,如最短路径算法、聚类分析算法^[34-35]等;③数据质量缺乏评估,ETC 数据分析已经得到广泛关注并取得阶段性成功,但在应用过程中缺乏 ETC 数据清洗质量评估,保证数据分析后的结论有效性。本文提出在利用 ETC 数据对高速公路车型车速分布的分析研究前,需要明确 ETC 系统设备布设及获取数据优缺点,对数据质量进行评测,确保数据的可靠性和准确性,以提高 ETC 数据的质量和值。

本文在数据提取和重构阶段,筛选出所需字段信息;以车牌号为基础,对车辆信息进行匹配,并计算速度;构建车辆 ID、数据字段、时间的三维数据集。同时,相较于目前应用广泛的方法,本文在优化步骤实施前,增加了数据准备阶段,有助于数据异常点识别、数据完整性检查和数据逻辑性梳理。数据优化遵从失真数据、重复数据的删除与缺失数据的补充。最终,基于“准确性”、“完整性”、“规模”3 个维度创新性地提出了一套适用于 ETC 数据的质量评估模型。

2 行驶速度分析方法

鉴于 ETC 门架数据因具有样本量大、覆盖全面、连续记录等优点,在分析车辆行驶速度上具有研究价值,但却容易受外界因素影响从而数据质量与完整性,本研究提出了一个基于 ETC 数据的高速公路货车速度特征提取方法,其能够实现 ETC 数据的提取、优化及质量评价,在此基础上实现行驶速度的统计分析与类别分析。方法框架如图 1 所示,共包含 3 个部分:①数据提取及优化,该部分目的是数据的清洗与重构;②数据质量评价,该部分旨在判断数据质量是否满足行驶速度特征分析的需求;③行驶速度分析,该部分主要实现不同纵断面条件下不同车型的行驶速度分布特征。

2.1 数据提取及优化

2.1.1 数据提取与重构

为了依托 ETC 数据有效开展交通特征分析,需要对其进行数据提取与重构。考虑 ETC 数据要素

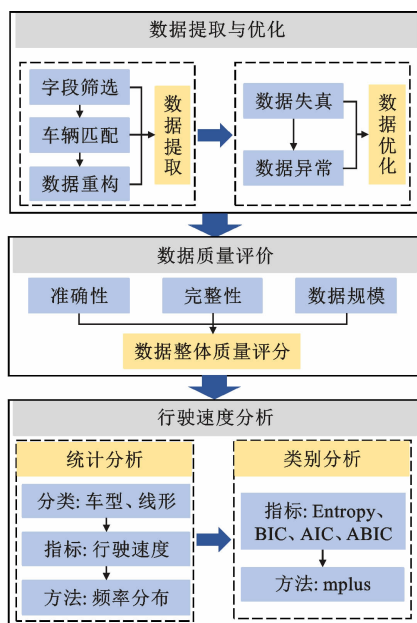


图 1 行驶速度分析方法框架

Fig. 1 Framework of Driving Speed Analysis Methods

多、数量大的特点,筛选提取有效信息能够有效减少数据量,有助于提高后期数据分析的效率^[21]。其次,按照交通特征分析的需求,重新构建数据集有助于进一步提升数据分析效率。具体处理步骤如下。

(1)字段筛选:考虑交通特征分析所需的车辆在空间和时间上的信息,本研究从 ETC 原始数据中提取的信息包括桩号、时间、车型、车牌等。

(2)车辆匹配:计算每一辆车的速度等信息,需要了解经过不同 ETC 门架所记录的桩号和时间信息。因此,本研究依托车辆号牌这一最准确的车辆识别信息,进行车辆信息的匹配。

(3)数据重构:考虑数据分析需求,本研究构建以车辆 ID、数据字段、时间 3 个信息为维度的 ETC 数据集,如图 2 所示。

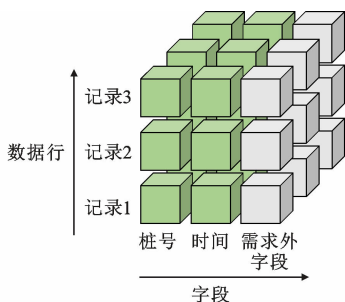


图 2 数据切片

Fig. 2 Data Slice

2.1.2 数据优化

根据章节 1 分析可知,ETC 数据存在记录异常、部分车辆行驶状况特殊等多方面的问题。因此,

在数据分析前,需要进一步对数据进行清洗和优化,以提高数据分析结果的准确性。本文针对 ETC 数据存在的具体问题,提出一套数据优化流程,如图 3 所示。数据优化流程的具体步骤如下:

(1)数据准备:当进行 ETC 数据提取与重构后,针对单车的全部数据进行初步分析很重要,因为这个阶段不仅有助于数据质量问题的检测、异常的发现以及数据可用性的提高,还为更深入的数据挖掘和分析奠定了基础。

数据准备阶段的初步分析可以帮助识别异常数据点,这些异常可能是由于设备缺陷所致;检查数据的完整性,确保所有必要字段都有正确的值,以避免后续分析中的歧义或错误。此外初步分析还可帮助梳理数据记录的逻辑性,理顺高速公路行驶车辆的 ETC 数据记录形式。

(2)数据优化:针对具体数据问题,采取一系列处理方法。处理数据失真问题,可利用车牌字段筛选特殊车辆并剔除,以单车数据行时间差为标准筛选异常数据^[33];处理数据异常问题,可直接筛选并删除完全重复数据行,对比行进方向与门架顺序,筛选左、右线中被重复记录的数据,保留其中正确数据^[28];也可以筛选存在数据空缺的数据行,利用车辆号牌匹配数据集,利用数据集中完整数据,补齐缺失数据。

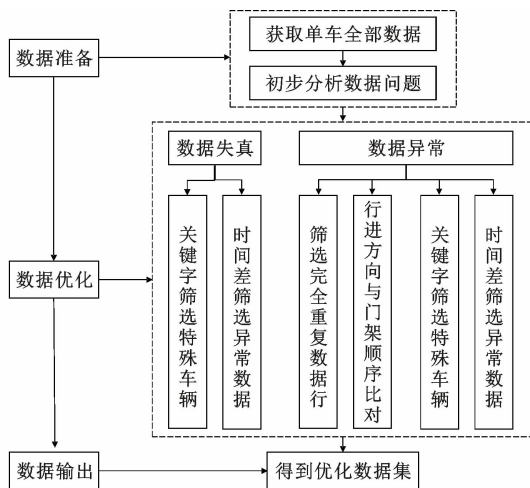


图 3 数据优化流程

Fig. 3 Data Optimization Process

ETC 数据经过上述优化处理后,数据质量提高,利用有效的 ETC 数据,可进一步提取行驶速度^[29]。由于 ETC 数据中包含桩号、时间、车辆号牌等信息,因此采用对单车数据行间作差提取行驶速度。通过比对通过相邻门架的通过时间、桩号,计算门架之间的里程,行驶里程与行驶时间之比即为车

辆的行驶速度。计算公式如下

$$V_{ij} = \frac{\Delta S_{ij}}{1\ 000} \cdot \left(\frac{\Delta T_{ij}}{3\ 600} \right)^{-1} = \frac{\Delta S_{ij}}{\Delta T_{ij}} \times 3.6 \quad (1)$$

式中: i, j 代表了相邻 2 个 ETC 门架编号; V_{ij} 为车辆在 i 门架和 j 门架之间的行程速度, 单位为 $\text{km} \cdot \text{h}^{-1}$; ΔS_{ij} 为 i 门架和 j 门架的桩号差, 单位为 m ; ΔT_{ij} 为 i 门架和 j 门架记录的车辆通过时间差, 单位为 s 。

单车数据行差值的行驶速度提取方法可以快速还原车辆在单个或多个时间段内的速度状态, 同时

适用于筛选某区段数据突变, 实现对数据优化结果的再次检验, 对 ETC 数据有较好的适用性。

2.2 数据质量评估

为了判断当前数据质量是否满足行驶速度特征分析的需求, 需要构建一套可行的数据质量评估方法, 以评估数据优化效果。因此, 本文从数据质量维度 (Data Quality Dimensions, DQD) 出发, 基于 ETC 数据对于“准确性”、“完整性”的需求以及 ETC 数据的信息特点, 分级划分数据维度, 提出了一套适用于 ETC 数据的质量评估指标。具体指标如表 1 所示。

表 1 数据质量评估指标

Table 1 Data Quality Evaluation Indicators

维度	子维度	评估目标	评估公式
准确性	数据异常情况	评估正确字段、记录、文件或数据集之外的无效数据情况, 如车辆类型记录为 0、门架顺序编号记录错误、某些字段为空白等数据情况。	$V_c = 100\% - C/M$ C 为有异常数据出现的数据行的数量; M 为被评价的数据行的数量。
	数据重复情况	评估特定字段、记录、文件或意外重复的数据情况, 如完全重复、同一门架左右线重复记录的数据情况。	$V_r = 1 - R/M$ R 为重复的数据行数量。
	数据内容正确性	评估所有字段、记录、文件或数据集中与实际不符的数据情况, 如车辆通过两门架之间时间间隔过长的数据情况。	$V_e = 1 - E/M$ E 为不满足数据正确性要求的数据行数量。
完整性	数据完整性	评估按照业务规则要求, 数据集中应被赋值的数据元素的赋值程度, 如单一数据行内某一信息缺失的数据情况。	$V_a = 1 - A/M$ A 为数据集中不满足业务要求赋值的数据行数量。
规模	数据量	当前被评价数据集的所有数据, 实际处理后的数据量与理论处理后的数据量相差情况。	$S = 1 - \left M_p - M \left(1 - \frac{C+R+E+A}{M} \right) \right \cdot \frac{1}{M}$ M_p 为处理后的数据行的数量。

由于特定 DQD 评分只能表示数据某一维度的质量, 在进行数据质量评估时, 需要对 ETC 数据进行整体质量的评估。因此, 本研究对不同 DQD 赋权, 得到了数据整体质量评价指标, 计算公式如式 (2) 所示

$$D = W_1 \frac{V_c + V_r + V_e}{3} + W_2 V_a + W_3 S \quad (2)$$

式中: D 为数据整体质量评分; V_c 为异常数据情况; V_r 为数据重复情况; V_e 为数据内容正确率; V_a 为数据完整性; S 为数据规模; W_1 为准确性维度权重; W_2 为完整性维度权重; W_3 为规模维度权重。

根据数据需求以及应用场景的不同, 特定 DQDS 的权重应当进行调整^[36]。本研究期望 ETC 数据能真实反应车辆行驶状况, ETC 数据的各类异常情况均会导致计算结果出现偏差, 影响分析结果的客观性。根据各维度中子维度的数量, 考虑对数据准确性和完整性相较规模有较高要求, 因此对不同 DQD 赋权 ($W_1 = 0.6, W_2 = 0.2, W_3 = 0.1$), 以满

足研究需求。

2.3 行驶速度特征分析

高速公路行驶速度是车辆行驶状态的一种体现, 与路段线形条件、驾驶环境、驾驶人、车辆类型等多种因素相关。在不同条件叠加影响下, 车辆行驶速度会呈现不一样的规律性。对于山区高速公路的海量 ETC 数据, 可采用频率统计分析、单因素方差分析以及潜在类别分析的方法, 总结在不同纵断面条件下不同车型的行驶速度分布特征。特征分析方法及流程如图 4 所示。

(1) 统计分析

在进行统计分析时, 首先对 ETC 数据进行基本频率统计分析, 将数据按数值范围或离散值分成不同的组或类别, 并计算每个组的频率或计数。这种分析的目的是为了更好地理解和可视化行驶速度的分布特征, 以便进行数据摘要、描述性统计和可视化呈现。

其次对于不同车型的行驶速度可以进行单因素

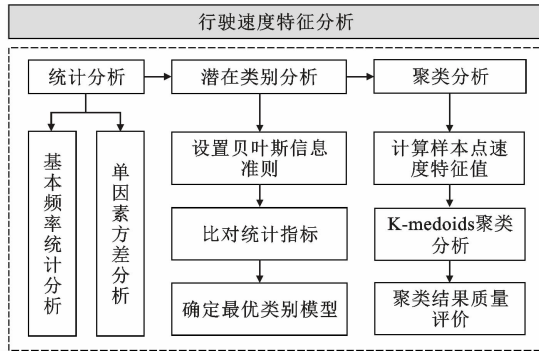


图 4 特征分析流程

Fig. 4 Feature Analysis Process

方差分析(One-way Analysis of Variance),这是一种用于研究一个自变量对一个连续的因变量的影响的方法。其主要目标是确定不同组之间的均值是否存在显著差异。用于比较多个(3个或更多)组之间的均值差异。在进行方差分析后,当需要确定哪些具体的组在均值上显著不同,而不仅仅是确定组之间是否存在差异时。可以进行 Tukey 事后检验进一步深入确定哪些组之间存在显著的均值差异^[37]。这是一种可以科学判断行驶速度是否存在显著性差异的方法。

(2) 潜在类别分析

为探究行驶速度特征采用潜在类别分析对各类车型的速度分布进一步研究,潜在类别分析(LCA)可以帮助识别观测数据中存在的潜在类别或群体,有助于研究者理解不同子群体的特征和行为。

在给定不同潜在类别 m 和每个维度 j 中的条件概率的情况下,计算随机变量 Y_i 取值为 j 的概率。这个概率是通过对不同潜在类别的加权求和,每个潜在类别的权重由 γ_m 给出,并且在每个维度 j 中,根据条件概率 $\rho_{k,r_j|m}$ 进行累乘得到。具体如下

$$P(Y_i=j) = \sum_{m=1}^M \gamma_m \prod_{j=1}^j \prod_{r_j=1}^{k_j} \rho_{k,r_j|m}^{I(y_j=r_j)} \quad (3)$$

潜在类别分析中没有单一的标准来确定类的最佳数量^[38]。通常采用 4 个标准准则来确定最优聚类数量,Entropy(熵)、BIC(贝叶斯信息准则)、AIC(赤池信息准则)、ABIC(调整的贝叶斯信息准则)。已有研究表明,在确定最佳类别数量时,BIC 比 AIC 更可靠^[39]。

上述 4 个指标是用于评估不同潜在类别模型的统计指标,帮助选择最合适的模型来解释数据。熵是用来评估模型分类的质量和清晰度的指标,越接近 1 表示模型的分类越清晰,模型对于个体的分类越准确。AIC、BIC、ABIC 用于衡量模型的拟合优

度和复杂性之间的权衡。三者的值越小表示模型越好。

通过数据优化提取行驶速度,通过抽样构造样本的速度均值、标准差、偏度和峰度等特征。以这些特征为输入变量,选择多组潜在类别的数量(k 值),根据拟合优度统计指标来确定最佳的 k 值。

通过软件如 Mplus 或 R 语言可以快速计算不同类别的相关统计指标 AIC、BIC、ABIC、熵。

$$V_{AIC} = -2\ln(\hat{L}) + 2k \quad (4)$$

$$V_{BIC} = -k\ln(n) + 2\ln(\hat{L}) \quad (5)$$

$$V_{ABIC} = V_{BIC} + \frac{2k(k+1)}{3(n-k-1)} \quad (6)$$

$$E_{Entropy}(X) = - \sum_{i=1}^n P(X=x_i) \cdot \log(P(X=x_i)) \quad (7)$$

式中: V_{AIC} 、 V_{BIC} 、 V_{ABIC} 分别为统计指标 AIC、BIC、ABIC 的值; $\hat{L}$ 为模型的最大似然估计对数似然函数; k 为模型参数的数量; n 为观测数据的数量; X 是离散的随机变量; $E_{Entropy}(X)$ 为随机变量 X 的离散概率分布的熵。

通过比对统计指标确定最优类别模型,LCA 模型可以估计每个潜在类别的条件概率分布,将车辆分为不同的类别,并获得每个类别各类车型的占比情况。根据最佳模型分类情况,结合不同车辆类型的速度特征值范围,可以尝试解释每个潜在类别的特征。

(3) 聚类分析

K-medoids 是一种聚类分析方法,与 LCA 在某种程度上有些相似,都是用于探索数据的内在结构或模式的方法。区别在于 K-medoids 使用不同的距离度量来定义簇内的相似性,而 LCA 可以使用似然度量来描述类别内的数据分布。受空间维度限制,聚类分析往往只能从 2 到 3 个维度进行分类,而聚类分析相较于潜在类别分析过程简单,结果呈现更为直观,各有优劣。

K-medoids 算法相比于 K-means 聚类算法对噪音和孤立点具有更强的鲁棒性,更适合应对数据中的异常值。K-medoids 通常使用不同的距离度量来衡量相似性,包括欧式、曼哈顿、切比雪夫以及明可夫斯基距离等多种方法。这些距离度量可用于确定数据点之间的距离和簇类相似性。通常可以采用明可夫斯基距离公式对上述距离概括表达,从而实现多维参数进行距离衡量,其计算公式如下

$$d(A,B) = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{1/p} \quad (8)$$

式中: x_i 、 y_i 分别为点 A 和点 B 在第 i 个维度上的坐标; p 为明可夫斯基距离的参数, 用来控制距离的计算方式, 根据不同的 p 值, 可以获得不同的距离度量 (通常 p 取 1 或 2)。

具体分析时的步骤为通过随机抽样选取主要研究的速度特征, 通过随机抽样构造样本点并计算出相应的样本点速度特征, 并将其作为点的坐标值。例如将行驶速度进行抽样, 计算每个样本的速度均值、标准差、偏度, 并以此作为 xyz 坐标, 构造三维点群, 通过 K-medoids 算法进行聚类分析。

对于聚类结果质量评价可以采用轮廓系数, 是一种用于度量聚类质量的指标, 它能够衡量聚类的紧密度和分离度, 通常用于评估聚类算法的聚类效果。轮廓系数的取值范围为 $[-1, 1]$, 越接近 1 表示聚类内部的样本距离相比于其他聚类之间的样本距离更加紧密, 聚类效果越好; 越接近 -1 表示聚类内部的样本距离相比于其他聚类之间的样本距离更加分散, 聚类效果越差, 根据最终聚类的轮廓系数确定最优的车型分类结果。

最后, 根据车辆类型的分类结果, 可制定相应的政策, 如改进道路安全措施、优化交通流程或者速度控制策略。

3 实例分析

3.1 数据来源

数据来源项目为西部山区某高速公路, 数据时

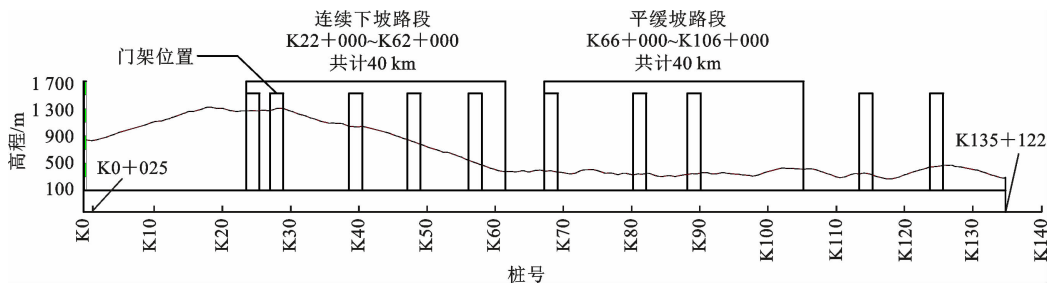


图 5 高速公路纵断面

Fig. 5 Vertical Section of Expressway

3.2 数据优化及质量评估

3.2.1 数据优化

针对上述 ETC 数据存在的问题, 根据数据实例介绍了一种数据清洗思路。数据清洗前, 需先将所有 ETC 门架数据按照车牌号码进行分组, 并对排序的单辆车数据表进行计算, 通过 Δs 、 Δt 、 v 三项计算结果进行判别和剔除, 处理数据重复、异常、缺失问题。

各类值的处理方式如表 3 列出的单辆车门架数

间跨度上, ETC 门架系统采集数据涵盖的时间范围为 2020 年 9 月 24~30 日, 共计 6 日。

该高速公路采用双向四车道设计, 设计时速为 $80 \text{ km} \cdot \text{h}^{-1}$ 。路线全长 136.9 km, 主要研究对象大致位于 K22~K106, 将路段划分出 2 段各 40 km 长的路段, 并分别将其定义为连续纵坡路段和平缓路段, 项目路段的基本情况如表 2 所示, 路段在纵断面的分布情况与门架的分布位置如图 5 所示。该高速公路路线平面基本为南北走向, 个别圆曲线半径较小的路段, 其最小圆曲线半径采用设计速度下的一般值, 平面线形整体较为平直; 其构造物情况在 K22~K62 间包含 4 条隧道, 在 K66~K106 间包含 1 条隧道, 构造物数量较少。道路平面线形与构造物对于行驶速度具有影响, 但在该实例中选取的 2 个研究路段可认为纵坡坡度是影响行驶速度的主要因素。

表 2 项目路段概况

Table 2 Overview of the Project Section

桩号	左右线	路段	起终点高差/m	平均纵坡坡度/%	门架设置情况
K22~K62	右线	连续下坡路段	887.87	2.22	共计 5 处断面设置 10 个门架, 左右线各 5 个门架
K22~K62	左线	连续上坡路段			
K66~K106	右线	平缓坡路段	25	0.06	共计 3 处断面设置 6 个门架, 左右线各 3 个门架
K66~K106	左线	平缓坡路段			

据, 可以通过计算每辆车车型的众数, 来补全车型信息, 原始数据左线共 879 817 条, 右线共 1 011 833 条。对于缺失值的处理, 共计补全车型信息左线 194 439 条, 右线 172 011 条, 占比分别为 22.1%、16.9%。

对于重复值的处理主要分为 2 类, 一类情况是对完全相同的数据行去重, 二类情况是筛除过小 Δt 的数据行, 2 类共计去除重复数据左线 9 766 条, 右线 12 648 条, 分别占比 1.1%、1.3%。

表 3 单辆车门架数据示例

Table 3 Example of Single Vehicle Gantry Data

优化项	桩号	门架标识	通过时间	车辆类型	$\Delta s/m$	$\Delta t/s$	$v/(km \cdot h^{-1})$
数据缺失	K83+200	201	2020/9/28 16:31:15	0	-	-	-
	K66+510	201	2020/9/28 16:40:45	0	-16 690	568.00	-105.78
数据重复	1 类	K104+600	201	2020/9/28 16:07:24	1	-	-
		K104+600	201	2020/9/28 16:07:24	1	0	0
	2 类	K93+430	201	2020/9/28 16:13:55	1	-	-
		K93+430	201	2020/9/28 16:13:56	1	0	1.00
数据异常	1 类	K93+430	201	2020/9/30 10:45:13	1	-	-
		K104+600	201	2020/9/30 10:51:22	1	11 170	369.00
	2 类	K66+510	201	2020/9/28 16:40:46	1	0	1.00
		K66+420	201	2020/9/28 16:44:14	1	-90	208.00

注:车辆类型“1”代表 1 类客车;“0”代表车辆类型数据缺失。

异常值的处理亦分 2 类,一类是删除速度方向与门架标识不符的数据行,二类是筛除不符合高速公路正常运行速度区间的数据行,2 类共计去除异常数据左线 44 078 条,右线 60 204 条,分别占比 5%、6%。

最终,对示例数据中列出数据行删除完成数据清洗。上述过程可以通过计算机编程语言按步骤编写,可解决问题数据行,得到单辆车的桩号、通过门架时间、行驶速度对应关系。

3.2.2 数据质量评估

根据 2.2 节提出的数据质量分析指标,结合已有的门架数据,从“准确性”“完整性”“数据量”出发对数据处理结果进行评价。

由图 6 可知原始数据经数据优化后,经处理后的数据规模大幅缩减,数据内容的准确性、完整性有所提高。具体表现为:删除了左右线的异常、重复数据使得 V_c 、 V_r 有所提升,补充了缺失数据使得左右线数据完整性(V_a)从 78.9、83.39 提升至 100,数据内容正确性(V_e)经优化后分别达到 99.96 和 99.97。

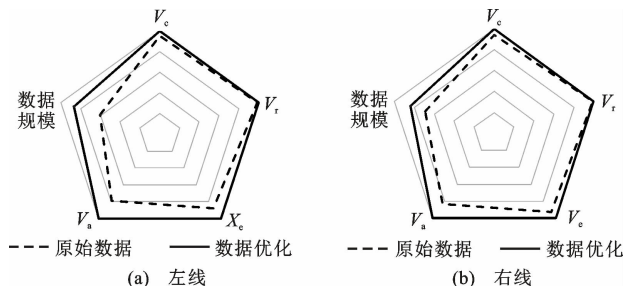


图 6 左右线数据评价雷达

Fig. 6 Left and Right Line Data Evaluation Radar Chart

根据本文第 2.2 节数据质量评分公式,计算左线和右线数据相比原始数据质量评分提升 15.7% 和 12.0%。从数据逻辑性来看处理得到的单辆车

的数据能够反映出车辆在一段时间内连续、单次通过门架并被门架记录下连续的行驶轨迹。避免了由于相邻方向上的门架错误记录,计算出速度大小值正常但速度方向错误的问题。

3.3 行驶速度特征分析

通过数据清洗后虽然行驶速度的准确性和完整性有所提升,但数据规模缩减较大。为了得出可靠的结论需要庞大的数据作为支撑,因此将数据样本的时间跨度扩充为 2022 年 1~6 月份全天 ETC 数据,并按 1 个月 ETC 门架产生的数据为单位,分析 2022 年 1 月至 6 月的数据。

高速公路行驶速度存在不确定性和随机性,路段的行驶速度受到很多复杂因素的影响,其中道路纵断面线形、车辆性能对于车辆的行驶速度有较大影响。统计学原理表明,统计分析大量数据时,随机现象会呈现出一定的规律性,对行驶速度作统计分析可以总结行驶速度的分布特征。

3.3.1 统计分析

将 2022 年 1 月该高速公路按照客货车型和路段分类统计行驶速度的频率分布,将速度从低到高排序按照 $5 km \cdot h^{-1}$ 的间隔进行频率统计,如图 7 所示,得到客车、货车在不同路段的频率直方图。

根据图 7 可以发现客、货车的速度分布呈现明显的特征,具体如下。

(1) 客车和货车的速度分布呈偏态分布,在不同路段受车辆特性影响速度分布的均值、方差和偏度不同,总体上客车速度均值高于货车。

(2) 在平坡、上坡路段客车与货车的速度均值相差较大而在下坡路段客车和货车的速度均值相差较小,客、货车在上坡路段速度分布较为分散而在下坡路段速度分布更为集中。

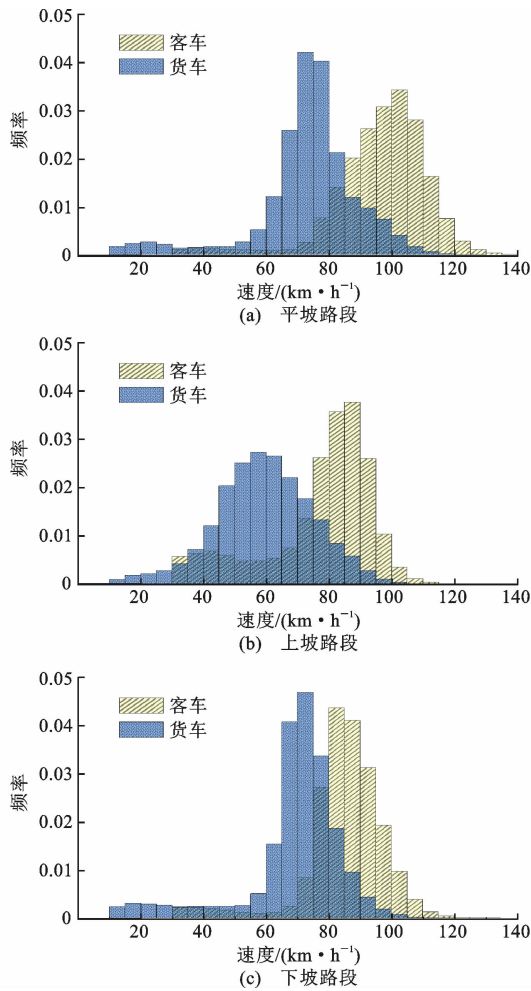


图 7 客货车不同路段频率直方图

Fig. 7 Frequency Histogram of Different Sections of Passenger and Freight Cars

(3)通过单因素方差分析和 Tukey 事后检验分析客车、货车在不同路段上行驶时其行驶速度平均值是否存在显著差异,在 Tukey 事后检验中,通常使用 P 值来判定是否存在显著差异。如果 P 值很小,通常小于设定的显著性水平(如 0.05),则会拒绝无差异的假设,认为差异是显著的。表 4 分析结果表明,客车和货车在不同路段的速度均值在显著性水平 0.05 下存在显著差异。

表 4 单因素方差分析

Table 4 One-way Analysis of Variance

车型	差异源	自由度	MS	F	P 值	F 临界值
客车	组间	2	46 561 129.00	182 228.5	0	2.995 737
	组内	1 826 203	255.51			
货车	组间	2	6 089 761.00	23 747.3	0	2.995 759
	组内	332 034	256.44			

3.3.2 潜在类别分析

由统计分析可知,客、货车在不同路段的行驶速

度平均值具有显著差异。其中不同类型车辆速度分布表现出来的参数各不相同,如:速度分布曲线的均值、标准差、偏度、峰度为全面探究这些特征潜在结构关系,选用潜在类别分析来进一步研究各类车型的速度分布。

根据上述提到的速度均值、标准差、速度分布曲线偏度、峰度和车型,利用 Mplus 进行潜在类别分析。以 2022 年 1 月份的 ETC 数据为研究对象,在上坡、下坡、平坡路段对各车型速度数据随机抽取 1 000 个样本,计算样本的速度均值、标准差、速度分布曲线的偏度、峰度,从 4 个维度对上坡、下坡、平坡路段行驶速度进行研究,对车辆类型分类进行探讨。

通过分析不同潜在类别数的统计指标,包括 Entropy、BIC、AIC 和 ABIC,帮助选择最合适的模型来解释数据,不同潜在类别数下的统计指标情况如图 8(a)、(c)、(e)中所示。

在图 8(a)、(c)、(e)中通过熵作为主要分类模型的指标,在熵相同情况下选择 AIC、BIC、ABIC 指标较小的分类模型。在平坡、下坡路段[图 8(a)、(e)],当使用 4 类别模型进行分类时,熵值达到 1,类别数再增加时 AIC 等统计指标数值下降放缓,熵值偏离了 1,因此 4 类别模型的分类效果较为理想;上坡路段[图 8(c)],统计指标 BIC 值在 5 类别模型时最小,AIC 和 ABIC 指标减小趋势也在 5 类别模型后放缓,此时熵值为 1,5 类别模型分类效果较为理想。

在确定了平坡、上坡和下坡路段行驶速度的潜在类别数后,对各类别中不同车型所占比例进行统计,结果如图 8(b)、(d)、(f)所示,每一条折线代表了不同的潜在类别,竖轴则表明了该类别中每种类型车辆的占比情况。

对于平坡、下坡路段分为 4 类别,上坡路段分为 5 类别这一现象,结合图 8(b)、(d)、(f)各类别中的车型占比情况发现平坡、下坡路段各类别中车型占比情况十分相似,而在上坡路段 2~6 类货车进一步分化为 2 个类别。产生这一现象的原因可能与客货车的动力性能与荷载质量有关。货车之间的荷载量往往差距很大,在平、下坡路段货车选取较为安全的速度区间驾驶,行驶速度区分不明显。在上坡路段,每辆车的爬坡功率、荷载都不相同,对于荷载量较大的货车,受限于车辆的功率质量比造成了各类货车的频率分布图的“错峰”。进而导致行驶速度在平坡、下坡路段分为 4 类,而在上坡路段分为 5 类。

为验证规律一致性,对 2~6 月份该高速公路

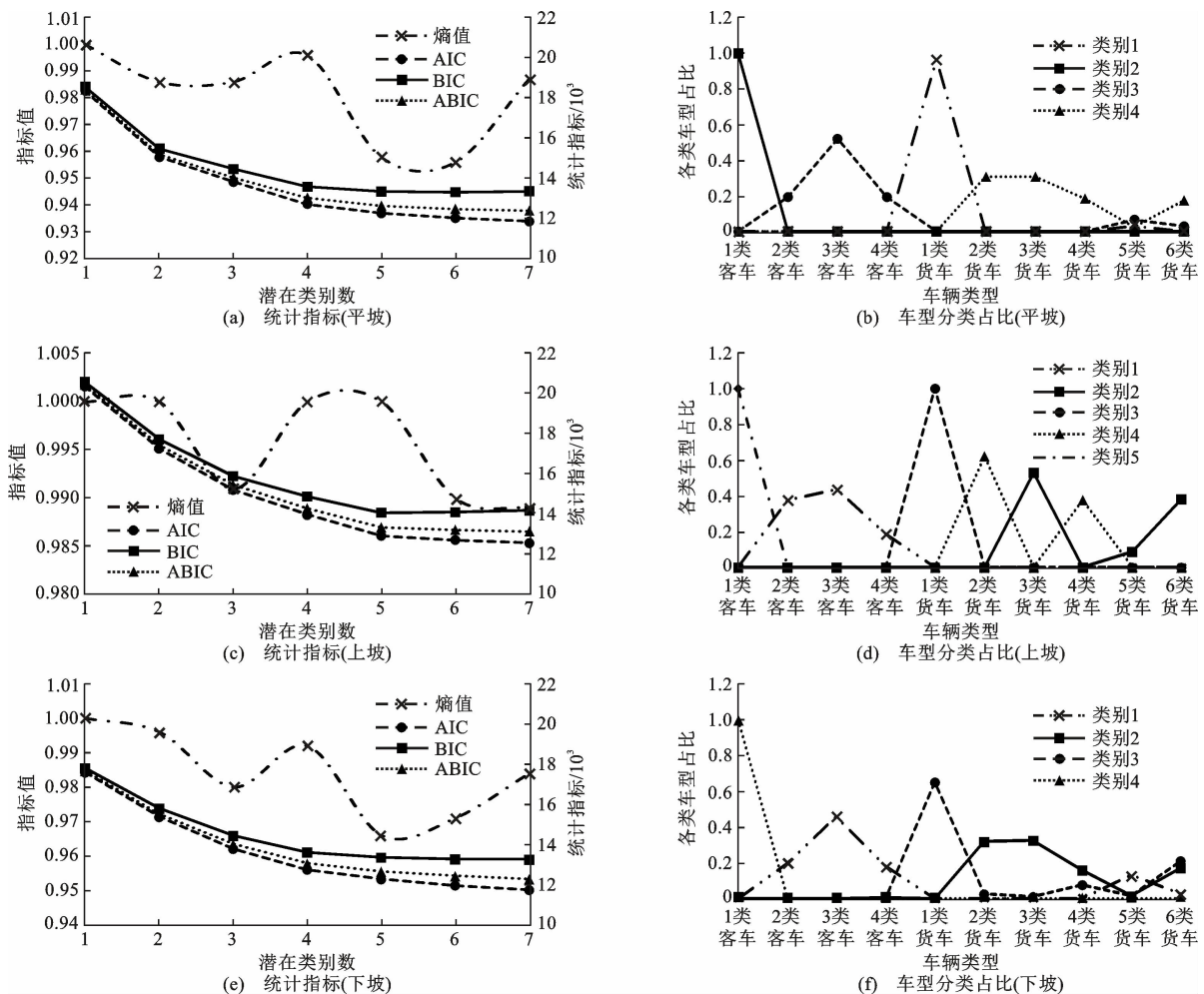


图 8 各个坡段统计指标与车型潜在分类

Fig. 8 Statistical Indicators for Each Section and Latent Classification of Vehicle Models

ETC 数据作重复性验证,根据速度均值、标准差和速度分布的偏度绘制三维散点图,分类结果如图 9 所示。图 9 中各类别在三维散点图中区分明显,分类结果与 1 月份潜在类别分析结果基本一致。

说明在平、上、下坡路段不同车型的行驶速度具有明显特征性,且在较长的一段时间内均保持相同的规律性,也一定程度上验证了前述 ETC 数据优化、提取和分析方法的有效性。

图 9 可以直观的看到 1 类客车作为单独的一个潜在类别,其簇类具有较强的独立性,在各个路段均和其他车型的行驶速度特征存在明显差异。该类别的特点在于速度均值大,速度标准差较大,说明速度较为离散;偏度呈负值,说明数据的主体集中在右侧,即行驶速度偏高。这一现象的原因可能与 1 类客车(通常为私家车、SUV)动力性能强劲有关,平、上、下坡路段对其速度的影响并不显著。

表 5 为 2022 年 1~6 月的 ETC 数据分析得到的各个路段的车型分类结果,由表 5 可知在该高速

公路的 3 种路段中,平坡和下坡的车型分类结果基本一致,上坡路段与其他路段的分类结果存在差异,差异的车型主要在于对 2~6 类货车的分类;上坡路段对于轴数较多的货车速度影响较大,使得 2~6 类货车进一步分化成 2 个子类,说明 5、6 类货车对于坡率的敏感性更高;下坡路段中,6 类货车的类别 2 和 3 中均存在,表明下坡时 6 类货车之间的速度特征具有较大差异。

由此可见不同车型在不同路段的行驶速度分布特征的差异性,启示管理者在不同路段作不同限速要求,例如:在上坡路段可能需要更具差异化的限速政策,以适应不同类别货车的速度特征;制定限速政策时应考虑货车的轴数和坡度;根据不同路段和时间的交通状况,可以考虑实施动态限速策略。

上述结论同时对公路安全管理具有重要参考价值,不同车型在不同路段的速度分布特征不同,这也意味着不同车型在不同路段的交通事故风险不同。因此,公路管理部门可以根据速度统计分析,提供有

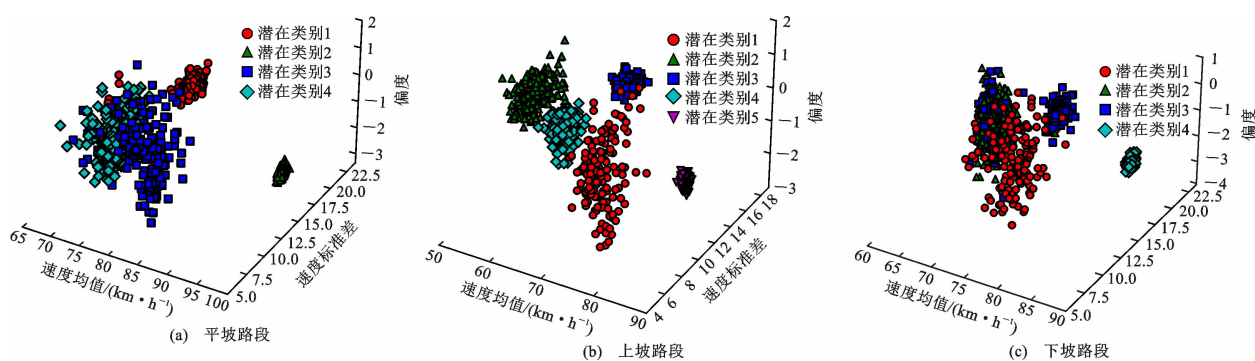


图9 各个路段潜在类别分析

Fig. 9 Latent Class Analysis of Different Sections

表5 路段分类结果统计

Table 5 Statistics of Road Segment Clustering Results

路段	类别	主要车型	占该类别的比例	车速/($\text{km} \cdot \text{h}^{-1}$)
平坡路段	类别1	1类货车	96.68%	73~79
	类别2	1类客车	100%	95~97
	类别3	2、3、4类客车、5类货车	客车:分别为19.3%、52%、19.3%;货车:6.6%	79~82
	类别4	2、3、4、6类货车	分别为30.7%、30.7%、18.4%、17.5%	66~74
上坡路段	类别1	2、3、4类客车	分别为37.9%、43.5%、18.6%	74~77
	类别2	3、5、6类货车	分别为52.9%、8.8%、38.2%	52~54
	类别3	1类货车	100%	64~66
	类别4	2、4类货车	分别为62.2%、37.8%	58~62
下坡路段	类别1	2、3、4类客车、5类货车	客车:分别为19.8%、46.2%、18.1%;货车:13.3%	76~79
	类别2	2、3、4、6类货车	分别为32%、32.7%、16%、17.3%	64~67
	类别3	1、6类货车	分别为65.3%、21.3%	73~76
	类别4	1类客车	100%	86~88

针对性的路面维护 and 安全管理策略,对公路的交通安全管理进行有及时的加强和调整。

3.4 限速方案分析

3.4.1 限速情况

高速公路设计时速为 $80 \text{ km} \cdot \text{h}^{-1}$,通过梳理研究路段限速方案,可以归纳出该高速主要的限速类型为分车型与分车道组合限速,各路段的限速方案具体见表6。

表6 各路段限速方案

Table 6 Speed Limit for Each Section

路段	限速方案	限速情况
平坡	单一限速	内外车道均限速相同值,取 $60 \sim 100 \text{ km} \cdot \text{h}^{-1}$ 不等
上坡	主要为单一限速、部分为分车道限速	单一限速:内外车道限速 $80 \sim 100 \text{ km} \cdot \text{h}^{-1}$ 。分车道限速:内车道 $100 \sim 120 \text{ km} \cdot \text{h}^{-1}$,外车道 $60 \sim 100 \text{ km} \cdot \text{h}^{-1}$
下坡	主要为分车道限速、部分为单一限速	分车道限速:内车道 $80 \text{ km} \cdot \text{h}^{-1}$,外车道 $60 \text{ km} \cdot \text{h}^{-1}$ 。单一限速:内外车道限速 $80 \sim 100 \text{ km} \cdot \text{h}^{-1}$

注:部分路段存在安检站,对应主线限速为 $60 \text{ km} \cdot \text{h}^{-1}$ 。

3.4.2 限速建议

对比表5中各个类别的车速情况和表6的限速情况,可以对当前各路段限速方案的合理性进行分析,存在的问题如下。

(1)平坡路段:路段采用单一限速方案,该路段内使内外车道限速值相同。对比表5各个类别的车型行驶速度均值范围,可以发现类别2(主要为1类客车)与其他类别的行驶速度相差较大(速差可达 $20 \text{ km} \cdot \text{h}^{-1}$),因此在该路段作单一限速不合理。

(2)上坡路段:路段采用单一限速、分车道限速相结合,现有的限速值在内车道上最高采用 $120 \text{ km} \cdot \text{h}^{-1}$ 进行限制。根据提取的行驶速度最高在 $83 \sim 85 \text{ km} \cdot \text{h}^{-1}$ 左右,结合高速公路设计时速 $80 \text{ km} \cdot \text{h}^{-1}$,认为路段限制速度与行驶、设计速度均不协调。存在多处路段最高限速 $120 \text{ km} \cdot \text{h}^{-1}$,与行驶速度 $83 \sim 85 \text{ km} \cdot \text{h}^{-1}$ 相比,其差值大于 $20 \text{ km} \cdot \text{h}^{-1}$ 。

(3)下坡路段:路段内5类货车的行驶速度偏高,其行驶速度与其他货车相比明显偏高,可能由于

空载的 5 类货车较多,其载荷量小,行驶速度与 2~4 类客车相当。因此按传统的客货分道,会使得 5 类货车在外车道行驶速度被迫降低,增加了 5 类货车变道超车的可能性,进而导致了增加高速公路追尾碰撞的风险。

综上,针对平坡路段,在限速值上,取消单一限速以满足较大的行驶速度差异。针对下坡、平坡路段,进行细分车型限速:1~4 类客车和 5 类货车速度较高,在内车道行驶;2、3、4、6 类货车速度较低,在外车道行驶。针对上坡路段,建议调整单一限速路段以满足客货车速差(隧道等特殊路段除外),对于限速值进行调整,当前多处路段实际行驶速度与限速值不符,建议在进行分车道限速时,调整限速值为内车道 $80\sim 100\text{ km}\cdot\text{h}^{-1}$,外车道 $60\sim 80\text{ km}\cdot\text{h}^{-1}$ 。

4 结 语

针对 ETC 数据的特点,本文以西部某高速公路 ETC 数据为实例,结合 ETC 门架系统特点总结了 ETC 数据中存在的问题,提出了交通特征分析方法,对高速公路行驶速度提取和优化进行总结,分析了高速公路不同路段速度分布特征并对潜在类别进行判别,确定了基于 ETC 数据的高速公路行驶速度提取、处理和分析的方法体系,得出如下结论。

(1)门架设置特点和由门架系统产生的数据决定了 ETC 数据是一个海量、时效性强、获取简便、准确性高的数据库。目前数据采集的过程中仍会导致 ETC 数据存在一些问题。对 ETC 存在的问题进行总结发现,主要有数据失真、数据重复、数据缺失以及一些特殊情况数据,分析了数据问题背后的原因。

(2)针对现有 ETC 数据存在的问题提出行驶速度提取方法和数据优化思路,构建了数据质量评估模型。对实例数据进行数据优化和数据提取,通过数据质量评估发现提出的方法能够较好地解决 ETC 数据存在的问题。经对比,处理后的左右线数据在数据完整性、数据内容正确性方面均有大幅提升。左线和右线数据质量评分同比提升 15.7% 和 12.0%

(3)从数据提取与优化、数据质量评估、行驶速度特征分析三部分提出 ETC 数据提取、优化和分析方法体系。从微观层面对高速公路的行驶速度进行分析,为高速公路速度安全问题提供了针对性的分类参考,为 ETC 数据挖掘提供借鉴价值。

(4)以西部某高速公路 2022 年 1~6 月产生的 ETC 数据为例,对行驶速度进行统计分析与潜在类

别分析。结果表明,通过数据挖掘的行驶速度具有较强规律性:客车和货车的速度分布呈偏态分布,在不同路段受车辆特性影响速度分布的均值、方差和偏度不同,客货车在不同路段的速度均值在显著性差异;平坡、下坡路段更适合 4 类别模型,上坡更适合 5 类别模型;分类结果中,轴数较多的货车速度对于坡率的敏感性更高,提示了管理者分车型制定限速措施时需要考虑车辆的荷载和动力性能等因素。

分析结果为实际高速公路速度管理措施的制定提供了具体的参考和指导,为针对不同车型采取不同的速度控制措施提供了依据,为不同路段制定不同的限速措施提供了理论支持。行驶速度分析结论是基于西部某高速公路得出,其结论对于其他高速公路不具备普适性和代表性,而数据提取、优化和分析方法则具有较好的适用性,有助于挖掘 ETC 数据的价值。

参考文献:

References:

- [1] 中华人民共和国中央人民政府. 加快推进高速公路电子不停车快捷收费应用服务实施方案. [EB/OL]. (2019-05-28) [2023-09-29]. https://www.gov.cn/zhengce/zhengceku/2019-09/27/content_5434216.htm?ivk_sa=1023197a.
The Central People's Government of the People's Republic of China. Implementation plan for accelerating the application and service of electronic non stop fast toll collection on free-ways [EB/OL]. (2019-05-28) [2023-09-29]. https://www.gov.cn/zhengce/zhengceku/2019-09/27/content_5434216.htm?ivk_sa=1023197a.
- [2] 杜 昭,张文榕,朱兴一. 基于网联车辆数据融合的路面平整度评估方法[J/OL]. 中国公路学报, (2023-06-27) [2023-09-30]. <http://kns.cnki.net/kcms/detail/61.1313.U.20230627.0943.004.html>.
DU Zhao, ZHANG Wen-rong, ZHU Xing-yi. Road roughness assessment based on fusion of connected-vehicles data [J/OL]. China Journal of Highway and Transport, (2023-06-27) [2023-09-30]. <http://kns.cnki.net/kcms/detail/61.1313.U.20230627.0943.004.html>.
- [3] 张 驰,任士鹏,王 博,等. 长大下坡路段货车运行速度特性及预测[J]. 华南理工大学学报(自然科学版), 2022, 50(3): 38-49.
ZHANG Chi, REN Shi-peng, WANG Bo, et al. Speed characteristics and prediction of trucks on long and steep downgrade sections [J]. Journal of South China University of Technology (Natural Science Edition), 2022, 50(3): 38-49.
- [4] 吴少峰,陈智威,张高峰,等. 基于实车数据的高速公路行驶轨迹偏移和车道侧向余宽[J]. 中国公路学报, 2023, 36(5): 197-209.
WU Shao-feng, CHEN Zhi-wei, ZHANG Gao-feng, et al. Investigation on trajectory deviation and lane lateral margin of

- passenger cars on freeways based on field driving data [J]. China Journal of Highway and Transport, 2023, 36 (5): 197-209.
- [5] 郭启明,王雪松,陈志贵. 基于驾驶模拟实验的山区高速公路运行速度建模[J]. 同济大学学报(自然科学版), 2019, 47(7): 1004-1010.
- GUO Qi-ming, WANG Xue-song, CHEN Zhi-gui. Modeling operation speed on mountainous freeways: A driving simulator study [J]. Journal of Tongji University (Natural Science), 2019, 47 (7): 1004-1010.
- [6] 徐进,崔强,林伟,等. 螺旋匝道和螺旋桥的小客车行驶速度特征[J]. 中国公路学报, 2019, 32(7): 158-171.
- XU Jin, CUI Qiang, LIN Wei, et al. Speed behavior of passenger cars on helical ramps and helical bridges [J]. China Journal of Highway and Transport, 2019, 32 (7): 158-171.
- [7] 吴玲,刘建蓓,单东辉,等. 高速公路特长隧道环境车道行驶速度特性及短时预测模型研究[J]. 隧道建设, 2023, 43(6): 1035-1044.
- WU Ling, LIU Jian-bei, SHAN Dong-hui, et al. Characteristics and short-term prediction model for detecting vehicle driving speed in extra-long freeway tunnels [J]. Tunnel Construction, 2023, 43 (6): 1035-1044.
- [8] KAN Y H, QIU X, SHEN X A, et al. Research on speed limit of two-lane highway of mountain area [J]. Applied Mechanics and Materials, 2012, 178-181: 1904-1907.
- [9] 李传华,吴昊,孙一帆. 基于交通冲突的山区高速公路可变限速控制方法[J]. 长安大学学报(自然科学版), 2018, 38(5): 154-161.
- LI Chuan-hua, WU Hao, SUN Yi-fan. Variable speed limit method of mountainous expressway based on traffic conflict Index [J]. Journal of Chang'an University (Natural Science Edition), 2018, 38 (5): 154-161.
- [10] COSTA A T, FIGUEIRA A C, LAROCCA A P C. An eye-tracking study of the effects of dimensions of speed limit traffic signs on a mountain highway on driver's perception [J]. Transportation Research Part F: Traffic Psychology and Behaviour, 2022, 87: 42-53.
- [11] 戴晓峰,张琪,覃文文,等. 恶劣天气对高原山区高速公路交通流特征的影响分析[J]. 交通信息与安全, 2020, 38(4): 10-16.
- JI Xiao-feng, ZHANG Qi, QIN Wen-wen, et al. An impact analysis of bad weather on traffic flow characteristics of plateau mountain expressway [J]. Journal of Transport Information and Safety, 2020, 38 (4): 10-16.
- [12] ATTALLAH M, KIANFAR J, WANG Y D. Impact of high resolution radar-obtained weather data on spatio-temporal prediction of freeway speed [J]. Sustainability, 2022, 14 (22): 14932.
- [13] 孙超,张红军,陈小鸿. 基于多源浮动车数据融合的道路交通运行评估[J]. 同济大学学报(自然科学版), 2018, 46(1): 46-52.
- SUN Chao, ZHANG Hong-jun, CHEN Xiao-hong. Road traffic operation assessment based on multi-source floating car data fusion [J]. Journal of Tongji University (Natural Science), 2018, 46 (1): 46-52.
- [14] WANG L, ABDEL-ATY M, SHI Q, et al. Real-time crash prediction for expressway weaving segments [J]. Transportation Research Part C: Emerging Technologies, 2015, 61: 1-10.
- [15] 曲栩,甘锐,安博成,等. 基于广义时空图卷积网络的交通群体运动态势预测[J]. 交通运输工程学报, 2022, 22(3): 79-88.
- QU Xu, GAN Rui, AN Bo-cheng, et al. Prediction of traffic swarm movement situation based on generalized spatio-temporal graph convolution network [J]. Journal of Traffic and Transportation Engineering, 2022, 22 (3): 79-88.
- [16] 郭凤香,黄金涛,陈显光,等. 多种缺失模式下交通数据组合近似填补方法[J/OL]. 交通运输系统工程与信息, (2023-09-21) [2023-09-30]. <http://kns.cnki.net/kcms/detail/11.4520.U.20230920.1914.004.html>.
- GUO Feng-xiang, HUANG Jin-tao, CHEN Yu-guang, et al. Combined approximate imputation method for traffic data under multiple missing modes [J/OL]. Journal of Transportation Systems Engineering and Information Technology, (2023-09-21) [2023-09-30]. <http://kns.cnki.net/kcms/detail/11.4520.U.20230920.1914.004.html>.
- [17] 李颖,赵莉,赵祥模,等. 基于大货车GPS数据的轨迹相似性度量有效性研究[J]. 中国公路学报, 2020, 33(2): 146-157.
- LI Ying, ZHAO Li, ZHAO Xiang-mo, et al. Effectiveness of trajectory similarity measures based on truck GPS data [J]. China Journal of Highway and Transport, 2020, 33 (2): 146-157.
- [18] 李清泉,胡波,乐阳. 一种基于约束的最短路径低频浮动车数据地图匹配算法[J]. 武汉大学学报(信息科学版), 2013, 38(7): 805-808, 885.
- LI Qing-quan, HU Bo, LE Yang. Flowing car data map-matching based on constrained shortest path algorithm [J]. Geomatics and Information Science of Wuhan University, 2013, 38 (7): 805-808, 885.
- [19] 李自圆,孙昊,李林波. 基于手机信令数据的长三角全域城际出行网络分析[J]. 清华大学学报(自然科学版), 2022, 62(7): 1203-1211.
- LI Zi-yuan, SUN Hao, LI Lin-bo. Analysis of intercity travel in the Yangtze River Delta based on mobile signaling data [J]. Journal of Tsinghua University (Science and Technology), 2022, 62 (7): 1203-1211.
- [20] ZHAO N L, QI T Y, YU L, et al. A practical method for estimating traffic flow characteristic parameters of tolled expressway using toll data [J]. Procedia - Social and Behavioral Sciences, 2014, 138: 632-640.
- [21] 赵怀鑫,邓然然,张英杰,等. 一种用于高速公路通行情况分析的收费数据挖掘方法[J]. 中国公路学报, 2018, 31(8): 155-164.
- ZHAO Huai-xin, DENG Ran-ran, ZHANG Ying-jie, et al. Method of mining fee data for expressway traffic analysis [J]. China Journal of Highway and Transport, 2018, 31 (8): 155-164.
- [22] 刘成龙,陶莎,赵聪,等. 高速路网不停车收费车道优化布设方法[J]. 中国公路学报, 2022, 35(5): 179-188.
- LIU Cheng-long, TAO Sha, ZHAO Cong, et al. Optimal de-

- ployment of electronic toll collection lanes for freeway network [J]. China Journal of Highway and Transport, 2022, 35 (5): 179-188.
- [23] CHEN L W, CHEN D E. Exploring spatiotemporal mobilities of highway traffic flows for precise travel time estimation and prediction based on electronic toll collection data [J]. Vehicular Communications, 2021, 30: 100356.
- [24] MYUNG J, KIM D K, KHO S Y, et al. Travel time prediction using k nearest neighbor method with combined data from vehicle detector system and automatic toll collection system [J]. Transportation Research Record: Journal of the Transportation Research Board, 2011 (2256): 51-59.
- [25] 赖见辉,齐悦,王扬,等. 基于收费数据的交通运行参数估计方法[J]. 中国公路学报, 2022, 35(3): 205-215.
- LAI Jian-hui, QI Yue, WANG Yang, et al. Estimation method of traffic state parameters based on toll data [J]. China Journal of Highway and Transport, 2022, 35 (3): 205-215.
- [26] 裴莉莉,孙朝云,韩雨希,等. 基于 SSC 与 XGBoost 的高速公路异常收费数据修复算法[J]. 吉林大学学报(工学版), 2022, 52(10): 2325-2332.
- PEI Li-li, SUN Zhao-yun, HAN Yu-xi, et al. Algorithm for repairing abnormal toll data of expressway based on SSC and XG Boost [J]. Journal of Jilin University (Engineering and Technology Edition), 2022, 52 (10): 2325-2332.
- [27] 魏广奇,苏跃江,吴德馨,等. 基于高速公路流水数据的通勤车辆特征研究[J]. 交通运输系统工程与信息, 2019, 19(3): 237-244.
- WEI Guang-qi, SU Yue-jiang, WU De-xin, et al. Trip characteristics of vehicle with commuting property based on highway ticket data [J]. Journal of Transportation Systems Engineering and Information Technology, 2019, 19 (3): 237-244.
- [28] 罗永煜,王阳生,赖树坤,等. ETC 门架交易系统的异常交易数据修复方法[J]. 福建交通科技, 2023(5): 79-83.
- LUO Yong-yu, WANG Yang-sheng, LAI Shu-kun, et al. Repair method of abnormal transaction data in ETC portal trading system [J]. Fujian Transportation Science and Technology, 2023 (5): 79-83.
- [29] 翁剑成,刘力力,杜博. 基于 ETC 电子收费数据的信息提取技术研究[J]. 交通运输系统工程与信息, 2010, 10(2): 57-63.
- WENG Jian-cheng, LIU Li-li, DU Bo. ETC data based traffic information mining techniques [J]. Journal of Transportation Systems Engineering and Information Technology, 2010, 10 (2): 57-63.
- [30] 《中国公路学报》编辑部. 中国交通工程学术研究综述·2016 [J]. 中国公路学报, 2016, 29(6): 1-161.
- Editorial Department of China Journal of Highway and Transport. Review on China's traffic engineering research progress: 2016 [J]. China Journal of Highway and Transport, 2016, 29 (6): 1-161.
- [31] 翁剑成,王媛,袁荣亮,等. 基于 ETC 交易数据的高速公路行程速度提取模型[J]. 交通信息与安全, 2017, 35(5): 76-82.
- WENG Jian-cheng, WANG Yuan, YUAN Rong-liang, et al. A model of travel speed calculation for freeway road segments with ETC transaction data [J]. Journal of Transport Information and Safety, 2017, 35 (5): 76-82.
- [32] 徐进,丁瑞,李诗佳,等. 山区高速公路车辆速度特性与区间限速方法[J]. 东南大学学报(自然科学版), 2022, 52(2): 352-361.
- XU Jin, DING Rui, LI Shi-jia, et al. Speed behavior of mountain expressway and determination method of interval speed limit [J]. Journal of Southeast University (Natural Science Edition), 2022, 52 (2): 352-361.
- [33] 戴晓峰,孔晓丽,陈方,等. 基于 ETC 数据和 A-BiLSTM 神经网络的高速公路节假日短时交通流预测模型[J]. 交通信息与安全, 2023, 41(3): 166-174.
- JI Xiao-feng, KONG Xiao-li, CHEN Fang, et al. A forecasting model of short-term traffic flow on expressways during holidays based on ETC data and A-BiLSTM neural network models [J]. Journal of Transport Information and Safety, 2023, 41 (3): 166-174.
- [34] 杜瑾,郝珺,樊海玮. 基于簇形均衡评估的高速公路收费数据聚类[J]. 交通运输系统工程与信息, 2016, 16(2): 98-103.
- DU Jin, HAO Jun, FAN Hai-wei. Expressway toll data clustering based on evaluation with balance of clusters' shapes [J]. Journal of Transportation Systems Engineering and Information Technology, 2016, 16 (2): 98-103.
- [35] 徐进,杨子邈,陈钦,等. 基于电子不停车收费数据的山区高速公路车速分布与车型分类研究[J]. 交通运输系统工程与信息, 2022, 22(5): 75-84, 116.
- XU Jin, YANG Zi-miao, CHEN Qin, et al. Speed distribution and vehicle type classification of mountain expressway based on electronic toll collection data [J]. Journal of Transportation Systems Engineering and Information Technology, 2022, 22 (5): 75-84, 116.
- [36] LI W Q, XU S P, PENG X D. Research on comprehensive evaluation of data source quality in big data environment [J]. International Journal of Computational Intelligence Systems, 2021, 14 (1): 1831.
- [37] 郭羽熙,袁小慧,高在峰,等. 超速行为多通道告警及差异化矫正研究[J]. 中国公路学报, 2022, 35(9): 26-37.
- GUO Yu-xi, YUAN Xiao-hui, GAO Zai-feng, et al. Multi-channel alarm and differential correction of overspeed behavior [J]. China Journal of Highway and Transport, 2022, 35 (9): 26-37.
- [38] LIU S J, LIN Z J, FAN W D. Investigating contributing factors to injury severity levels in crashes involving pedestrians and cyclists using latent class clustering analysis and mixed logit models [J]. Journal of Transportation Safety & Security, 2022, 14 (10): 1674-1701.
- [39] NYLUND K L, ASPAROUHOV T, MUTHÉN B O. Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study [J]. Structural Equation Modeling: A Multidisciplinary Journal, 2007, 14 (4): 535-569.