

基于门架数据的高速公路运行状态评价研究

刘春生¹, 曹蓉², 王晓晗¹, 赵鹤然³, 贾健民^{1*}

(1. 山东建筑大学 交通工程学院, 山东 济南 250101; 2. 山东高速股份有限公司, 山东 济南 250101;
3. 北京交通大学 交通运输学院, 北京 100044)

摘要: 高速公路信息化建设日益完善, 为了能够更加详尽地掌握高速公路的交通状态, 以济青高速为研究对象, 通过挖掘门架数据, 设计了 k 均值聚类算法和基于密度的噪声应用空间聚类算法相结合的两阶段聚类方法来识别驶入服务区车辆和异常行驶的车辆, 再结合各车型流量占比加权的交通状态指数, 从时间和空间维度分析高速公路路段交通状态。研究表明, 两阶段聚类算法有着很好的识别效果, 通过交通状态指标发现在7:00—20:00时段高速公路存在3个交通状态较拥堵的时段, 精确地识别出高速公路中交通状态较拥堵的路段, 并且发现路段的大车混入率与交通拥堵程度呈现密切的正相关的趋势, 最后依据交通状态指数将济青高速路段交通状态划分为4个不同的等级, 为交通管理部门提供了路段交通状态评估的技术支持。

关键词: 门架数据; 交通状态; 聚类算法; 交通状态指数; 大车混入率

中图分类号: U491 文献标志码: A 文章编号: 1002-4026(2023)03-0100-08

开放科学(资源服务)标志码(OSID):



Exploring the traffic state identification of highway based on gantry data

LIU Chunsheng¹, CAO Rong², WANG Xiaohan¹, ZHAO Heran³, JIA Jianmin^{1*}

(1. School of Traffic Engineering, Shandong Jianzhu University, Jinan 250101, China;
2. Shandong Expressway Co., Ltd., Jinan 250101, China; 3. School of Traffic and
Transportation, Beijing Jiaotong University, Beijing 100044, China)

Abstract: To thoroughly investigate the traffic state of highways, Jiqing Highway was selected as the study case. By mining the gantry data, a two-stage clustering algorithm combining k -means and density-based special clustering of applications with noise (DBSCAN) algorithms were proposed. The method was used to identify vehicles entering the service area and driving abnormally. Subsequently, the filtered vehicle records were extracted to realize a traffic state index weighted by the vehicle type to analyze the traffic state of the highway in terms of spatiotemporal dimensions. Results indicate that the two-stage clustering algorithm performs very well in the identification. The traffic state index indicated three periods when the highway is defined as congested during 7:00—20:00. Furthermore, it accurately identifies the congested sections of the highway. Moreover, it shows out that the mixed rate of large vehicles and the degree of traffic congestion in

收稿日期: 2022-09-15

基金项目: 国家自然科学基金(41901396, 42001396); 山东省高等学校青创科技支持计划(2021KJ058)

作者简介: 刘春生(1999—), 男, 硕士研究生, 研究方向为智能交通。

* 通信作者, 贾健民, 男, 副教授, 博士, 研究生导师, 研究方向为交通大数据分析、智能交通。Tel: 15376181624, E-mail: jiajianmin18@sdjzu.edu.cn

a section have a close positive correlation. Finally, according to the evaluation index, the traffic state of the Jiqing Highway is divided into four levels, which provides technical support for the traffic authorities to evaluate and manage the highway sections.

Key words : gantry data; traffic condition; clustering algorithm; traffic state index; mixed rate of large vehicles

截至 2021 年,我国高速公路里程数已经达到 16.91 万公里,但是随着我国汽车保有量的增加,高速公路仍然会出现拥堵问题。为更好地评估高速公路运行状态,门架系统应运而生。目前,全国已建成了 2.66 万套门架系统,记录了海量的行车感知数据。随着数据挖掘技术逐渐成熟,通过智能化手段从海量的门架数据中挖掘出复杂的交通流变化态势是当前交通研究的热点与难点。在建成交通强国、智慧高速等背景下,研究高速公路路段时空维度的交通运行状态,对引导出行者出行、提高路段通行效率具有重要意义。

由于门架系统的建设周期较长,已有的高速公路交通管理研究对于大规模的门架数据仍少有涉及,刘群等^[1]利用门架数据进行了交通流预测的研究;邢丽峰等^[2]利用门架数据进行了高速公路各类车辆的交通量转换的研究;曹波^[3]在门架数据的基础上提出区间拥堵指数,用于估计路段交通状态。在道路运行状态估计方面,现阶段采用的数据大致分为固定式交通检测技术采集的数据(感应线圈、微波检测)^[4-5],移动式交通检测技术采集的数据(全球定位技术(global positioning system,GPS)^[6]、手机信令采集技术^[7]),以及收费站数据^[8],无人机^[9-10]、监控视频^[11]检测技术采集的数据。而门架数据相较这几类数据具有如下优势:固定式交通检测设备往往只在特定的路段布置,无法大规模地应用,而门架设备布置规模非常之广;移动式交通检测技术往往会受到周围环境的影响导致精确度大大下降,而门架数据精确度较高,几乎不存在数据大批量缺失的现象;利用收费站数据进行交通状态估计时,由于数据中仅仅记录车辆进入和离开高速路网的数据,所以数据处理过程相当复杂,而门架系统记录车辆经过每一个路网路段的数据,可以更好地检测车辆在每一个路段的行驶状态,计算过程相对容易;无人机、监控视频数据虽然精度高、灵活性强,但是成本高、不利于推广应用,而门架数据获取容易且维护成本也较低。综上所述,门架数据具有应用范围广、精度高、计算方便、成本低的特点。由于高速公路两两门架之间设有服务区,在计算驶入服务区的车辆平均行程速度时会偏低,根据梁奇^[12]统计,高速公路服务区客车、货车驶入率均超 20%,但现有研究方法对此问题通常不予考虑,造成结果可靠性下降。赖见辉等^[13]采用机器学习的方法对车辆是否驶入服务区进行预测,而该方法存在一定的误差,可解释性差,并且还需进行问卷调查来获取模型的训练数据,增加了工作量。直接利用原始门架数据,根据每个车辆的平均行程速度特性挖掘出驶入服务区的车辆的方法准确度高,可解释性强。

本研究通过两阶段的聚类方法准确地识别出路段中驶入服务区的车辆和行驶异常的车辆,再利用交通状态指数在时间和空间维度对济青高速的交通状态进行评价分析,挖掘其时空规律。最后根据交通状态指数将济青高速各个路段划分为 4 个拥堵等级。为高速公路管理部门准确了解高速公路运行状态,制定拥堵管理措施提供借鉴,进而提升高速公路的通行能力。

1 异常速度识别方法

1.1 k 均值聚类算法

k 均值聚类(k -means)算法^[14]作为一种经典的聚类算法,已经广泛地应用于各个领域。本研究将 k -means 算法用于识别驶入服务区的车辆,以时间维度和各个车辆平均行程速度 v_i 为聚类指标,利用是否驶入服务区的两类车辆在平均行程速度上存在较大差异的特点解决了 k -means 算法难以确定初始聚类中心的问题,识别过程如下:

(1)指定聚类数目 k 和聚类中心 $(c_1, c_2, \dots, c_i)$,由于本研究已经非常明确地将数据集划分为两个类别,并且未进入服务区车辆和进入服务区车辆在平均行程速度上存在明显的差异,所以将聚类数目 k 指定为 2,聚类中心 (c_1, c_2) 指定为 $([\bar{t}, \bar{v}], [\bar{t}, s])$ 。其中 $\bar{t}$ 为目标时段的中间时刻; $\bar{v}$ 为未驶入服务区车辆的平均行

程速度的经验值,本研究选取 90 km/h; s 为驶入服务区车辆的平均行程速度的经验值,本研究选取 20 km/h。

(2) 计算车辆 x_j 到各个初始聚类中心 c_i 的欧几里得距离:

$$D_i(x, c_i) = \sqrt{\sum_{t=1}^n (x_t - c_{it})^2}, \quad (1)$$

并将其划到欧几里得距离最小的聚类中心对应的簇上。 n 表示特征数,本文取值为 2。

(3) 根据各个类别中的车辆 x_j 来重新计算各个簇的质心,将重新计算的质心作为新的聚类中心:

$$c_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_j, \quad (2)$$

其中 n_i 为簇 i 的个数。

(4) 重复(2)和(3)步骤,直到各个聚类中心 c_i 不再发生变化,或者达到预设的迭代数。

1.2 基于密度的噪声应用空间聚类算法

基于密度的噪声应用空间聚类算法(density-based special clustering of applications with noise, DBSCAN)^[15]可以在具有噪声的数据集中发现任意形状的簇。该算法最重要的两个参数分别为邻域密度阈值 ρ_{MinPts} 以及邻域半径 R_{Eps} 。本研究利用该算法识别某一路段未进入服务区车辆的平均行程速度集合中的偏大或者偏小的离群数据,相对于传统的阈值筛选的方法更加灵活,识别过程如下:

(1) 定义某一路段未驶入服务区车辆的平均行程速度集合 D , 给定邻域密度阈值 ρ_{MinPts} 和邻域半径 R_{Eps} , 计算各个车辆平均行程速度 v_i 的邻域对象个数 $|N_{\text{Eps}}(v_i)|$, 若 $|N_{\text{Eps}}(v_i)| \geq \rho_{\text{MinPts}}$, 则 v_i 为核心对象。

(2) 判断各个核心对象是否在其他核心对象所构成的邻域里面,若在其中则该两个核心对象构成密度可达,密度可达的核心对象邻域构成的集合为一个簇,不在任何簇中的车辆平均行程速度则为异常数据。

2 交通状态评估指数

交通状态指数(I_{TS})^[16]可以用量化数值来表示道路交通状态,数值越大表示交通拥堵越严重,分为基本模型和加权模型。本研究使用各车型流量加权模型,计算公式如下:

$$T_{ij}^g = \frac{v_f^g - v_{ij}^g}{v_f^g} \times 100, \quad (3)$$

$$T_{ij} = \frac{\sum_{g=1}^n (f_{ij}^g \times T_{ij}^g)}{\sum_{g=1}^n f_{ij}^g}, \quad (4)$$

式中, T_{ij}^g 为路段 i 时间段 j 中 g 类车的 I_{TS} 值, v_f^g 为 g 类车的理想平均行程速度, v_{ij}^g 为路段 i 时间段 j 所有 g 类车的实际平均行程速度的平均值, T_{ij} 为路段 i 时间段 j 的 I_{TS} 值。

交通状态评估流程如图 1 所示,首先利用两阶段聚类算法对车辆平均行程速度进行筛选,避免驶入服务区车辆对路段平均行程速度计算的干扰,提高计算的可靠性,最后计算各路段 I_{TS} 来评估路段交通状态。

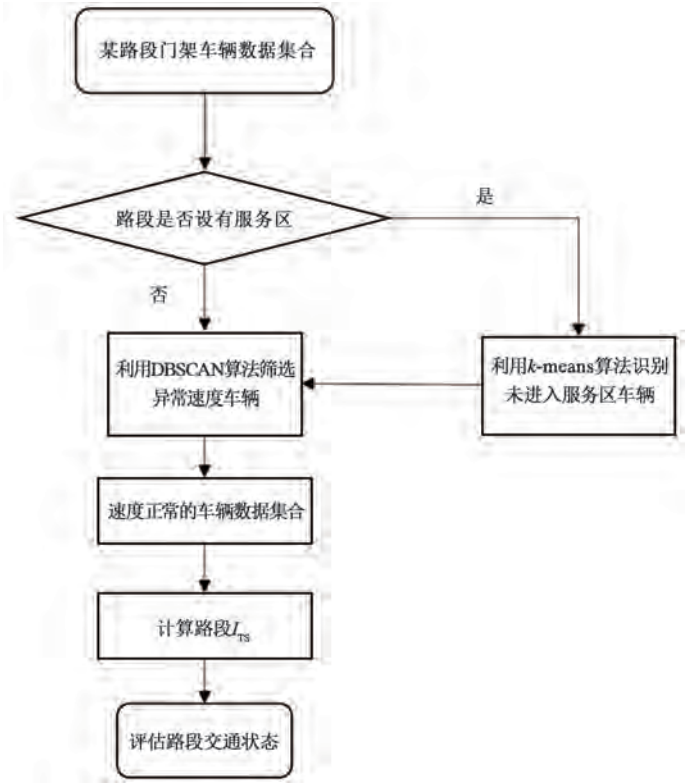


图 1 路段交通状态评估流程

Fig.1 Traffic condition assessment process of road sections

3 实例分析

3.1 数据概况与预处理

本研究所使用的数据来自山东省济青高速济南章丘—青岛机场北枢纽段共 26 个门架路段,路段全长 309.633 km,2021 年 9 月 6 日至 12 日共一周的门架数据 4 063 295 条,删除重复值和缺失值后还剩余 4 051 105条数据,字段信息如表 1 所示。

表 1 门架数据字段说明

Table 1 Description of the gantry data field

字段名称	说明
gantry_id	车辆通过当前门架的 ID
last_gantry_time	车辆通过上一个门架的时间
trans_time	车辆通过当前门架的时间
vehicle_plate	车辆的车牌号
vehicle_type	车辆类型

本研究以相邻两个门架组成的路段为研究路段,共 25 条,上游门架数据以 trans_time 和 vehicle_plate 字段为键,下游门架以 last_gantry_time 和 vehicle_plate 字段为键,进行数据拼接,拼接后的数据即为路段行车数据,包括车牌号、经过上下游路段时间、路段长度、车辆平均行程速度,其中某辆车经过该路段的平均行程速度 $v_i = \frac{l}{\Delta t}$, l 为路段长度, Δt 为经过上下门架的时间差。济南章丘—青岛蓝村的研究路段编号 0~24,每个路段的交通量周变化情况如图 2 所示,可以看出在这一周内高速公路周一交通量最大,周三交通量最小,并且路段越往青岛方向交通量越小。

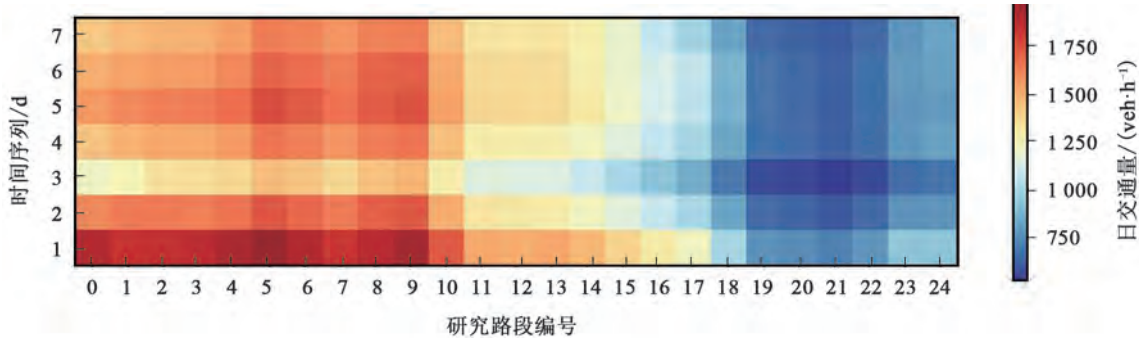


图 2 交通量周变化情况

Fig.2 Weekly changes in traffic volume

3.2 异常速度筛选

本研究以 1 h 为时间间隔,对设有服务区的路段采用 k -means 聚类算法将驶入服务区的车辆识别出来后,再将未进入服务区的数据利用 DBSCAN 聚类算法进行二次筛选,识别出速度偏大和偏小的异常数据,其中 k -means 聚类算法初始聚类中心 $([12,90],[12,20])$ 的最大迭代数设置为 300,DBSCAN 聚类算法的邻域半径 R_{Eps} 设置为 1 km/h,邻域密度阈值 ρ_{MinPts} 设置为 3,筛选结果如图 3 所示。从图 4 可以看出,经过筛选后的数据更趋近于正态分布。

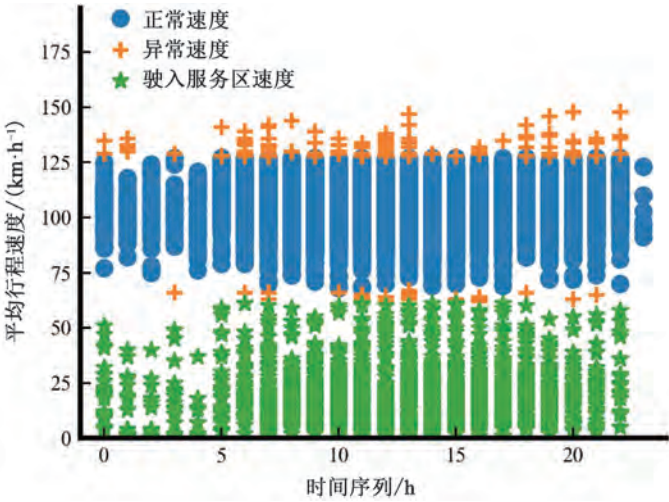


图 3 速度筛选结果

Fig.3 Speed screening results

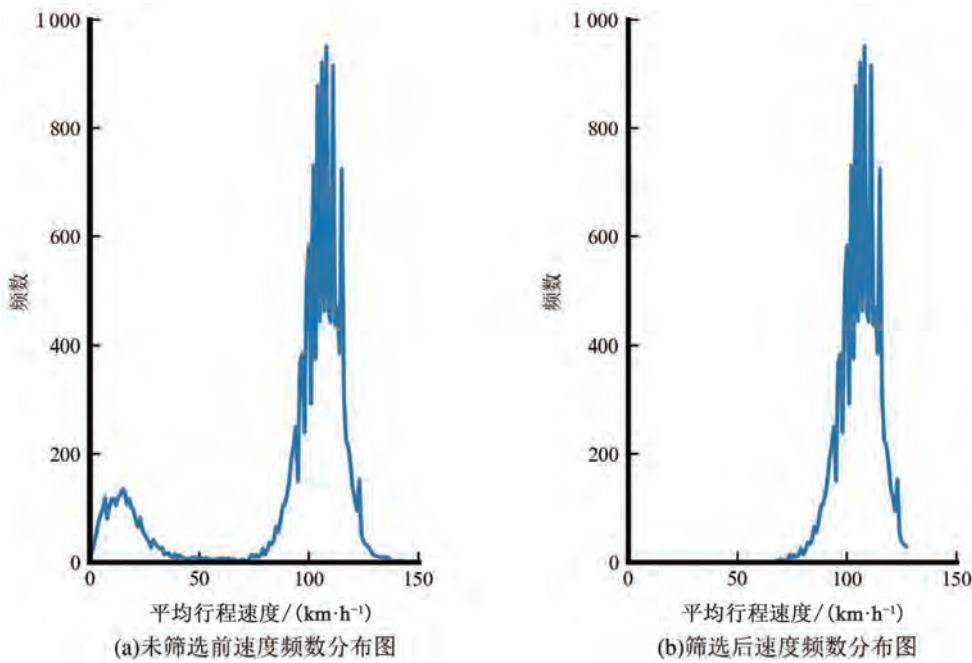


图 4 速度数据筛选前后对比图

Fig.4 Comparison before and after speed-data screening

3.3 交通状态评估

以行程速度为基础的交通拥堵评价指标是用于评估交通状态的常用手段,例如路段延误、TTI (travel time index) 等,但往往在理想行程速度的取值上选取道路的最高限速^[17-18],而不同车型之间的理想行程速度是存在差异的,所以本研究选取交通量占比较高的3种车型,如表2所示。计算3种车型流量加权求和后的 I_{TS} 指标来判断某一路段的交通状态,研究时段选取旅客出行活动较为频繁的上午7:00到晚上20:00,从时间和空间两个维度对济青高速进行交通状态的评估。在时间维度方面,计算济青高速研究时段的日平均 I_{TS} 指标变化情况,从图5可以看出,在研究时段中存在3个 I_{TS} 指标高峰时段,分别为9:00—10:00、14:00—15:00、19:00—20:00,其中19:00—20:00时段 I_{TS} 指标最高。

表2 各车型理想行程速度取值表

Table 2 Values of ideal travel speed for each model

车型	平均行程速度/(km·h ⁻¹)	理想行程速度/(km·h ⁻¹)	交通量占比/%
小型客车	107	120	72.5
小型货车	84	95	7.5
中大型货车	80	90	17.9
其他车型	—	—	2.1

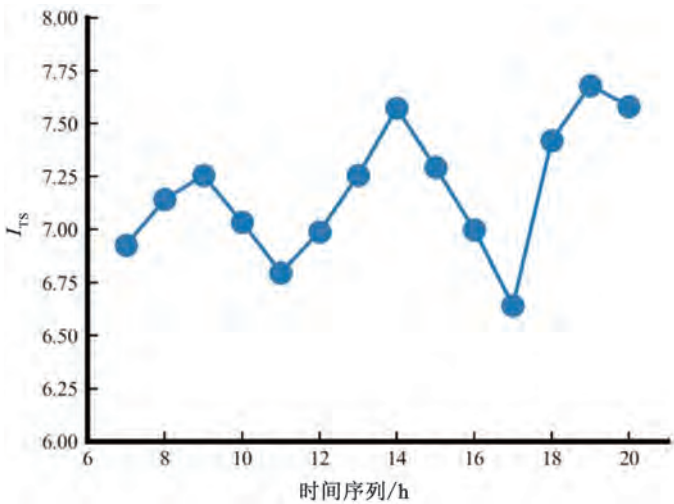


图5 济青高速 I_{TS} 指标日变化情况

Fig.5 Daily variation in the I_{TS} index of the Jiqing Highway

从空间维度分析各个研究路段7天中8:00—20:00的平均 I_{TS} 指标的日变化情况,从图6可以看出济青高速中交通状态最拥堵路段为9,从济南开始的0~6路段较为拥挤,较为流畅的路段为12~14路段以及18~22路段,其余路段交通拥堵状态较为一般。大车(中大型货车)在交通流中的混入率与交通拥挤程度有着重要的关系^[19],如图7所示。大车混入率与 I_{TS} 呈现正相关趋势,并且两者的皮尔逊系数为0.68,可见两者线性关系十分密切,其中路段9为临淄东—青州西路段,临淄区为山东省重要的工业园区,出入大车的数量会比较多,所以大车混入率较高,交通运行状态也较为拥堵。可以对大车实行夜间时段通行费用折扣的措施,错开小客车出行高峰时段,以此降低路段拥堵状态。

根据以上情况将济青高速线路的路段交通状态划分为4个等级,各个等级日平均 I_{TS} 指标范围如表3所示。

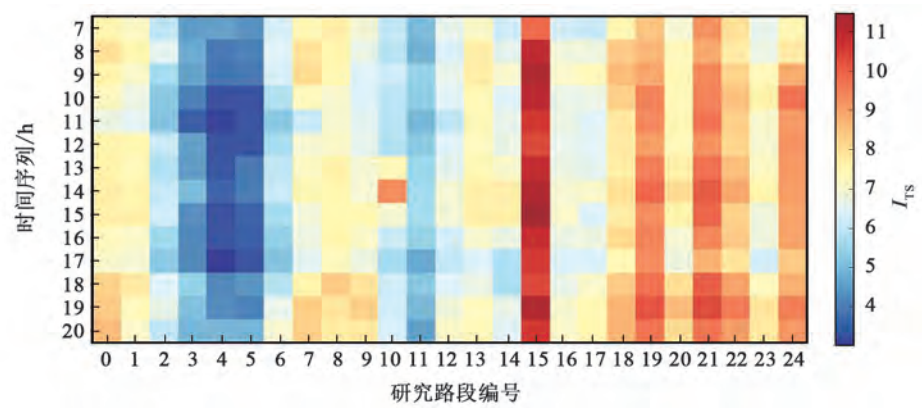


图 6 研究路段 I_{TS} 日变化情况

Fig.6 Daily variation in the I_{TS} of the tudy area

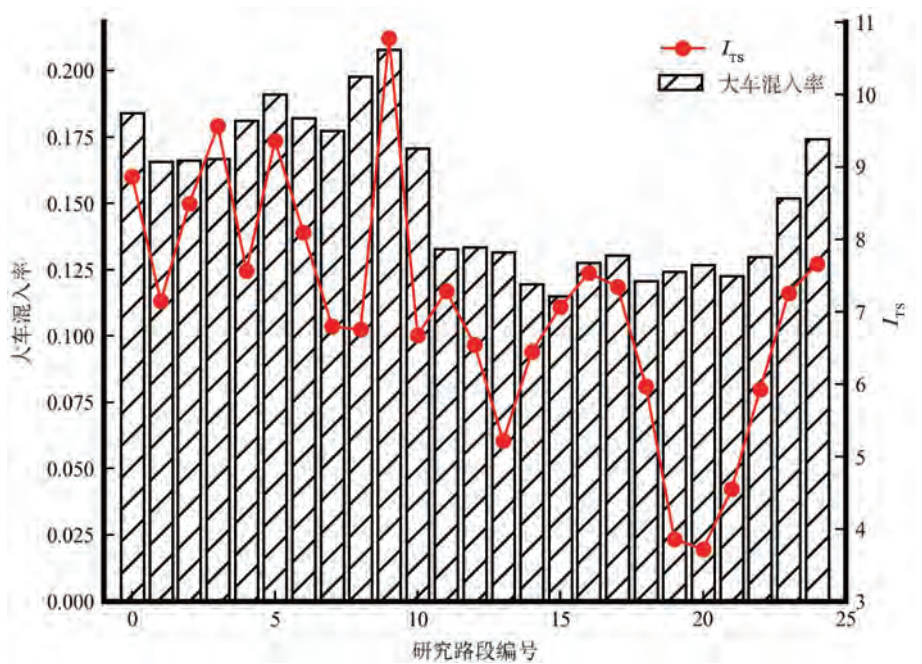


图 7 研究路段大巴混入率对比图

Fig.7 Comparison of the mixed rate of large vehicles in the research section

表 3 各路段交通状态等级划分

Table 3 Classification of traffic conditions in each road section

路段名称	交通状态等级	日平均 I_{TS} 指标
青州东—于河枢纽、潍坊东—胶州 章丘—邹平西、周村—淄博西	流畅	(0,6.54]
淄博高新区—临淄东、青州西—青州东、 潍坊西—潍坊东、胶州—机场北枢纽	一般	(6.54,7.66]
章丘西—章丘、邹平西—周村 淄博西枢纽—淄博高新区	拥挤	(7.66,9.57]
临淄东—青州西	拥堵	(9.57,∞)

4 结论

本研究提出了一种基于门架数据的两阶段聚类方法来识别高速公路驶入服务区的车辆和行驶速度异常

的车辆,再利用各车型流量占比加权的交通状态指数 I_{TS} ,有效地识别路段异常的行程速度数据,评估了济青高速25个路段的交通状态,发现大车混入率与 I_{TS} 呈现密切的正相关的关系,即大车混入率越高交通状态越拥堵,并且将济青高速路段交通状态划分为4个等级,本研究方法可以为交通管理部门发现高速公路瓶颈路段提供理论支持。但本研究所使用的的数据量较少,仅仅分析了交通状态的日变化和周变化的情况,提取的交通参数数量以及精度方面也有待提高,在以后的研究中应加大研究的数据量,分析更长周期的交通状态,利用更为先进的算法提取更多有意义的交通参数。

参考文献:

- [1]刘群,杨濯丞,蔡蕾.基于ETC门架数据的高速公路短时交通流预测[J].公路交通科技,2022,39(4):123-130. DOI: 10.3969/j.issn.1002-0268.2022.04.014.
- [2]邢丽峰,邱廷铨,邢宇鹏.基于ETC门架数据的高速公路交通量转换探究[J].中国交通信息化,2022(4):94-96. DOI: 10.13439/j.cnki.itsc.2022.04.004.
- [3]曹波.基于门架数据的道路运行指数应用研究[J].中国交通信息化,2022(1):110-113. DOI: 10.13439/j.cnki.itsc.2022.01.011.
- [4]王光辉.高速公路交通运行状态判别方法研究[D].长春:吉林大学,2015.
- [5]NANTHAWICHT C, NAKATSUJI T, SUZUKI H. Application of probe-vehicle data for real-time traffic-state estimation and short-term travel-time prediction on a freeway[J]. Transportation Research Record: Journal of the Transportation Research Board, 2003, 1855(1): 49-59. DOI: 10.3141/1855-06.
- [6]赵千里.基于车载GPS数据的城市交通状态估计和出租车需求预测研究[D].上海:上海交通大学,2016.
- [7]刘超彪.基于手机信令数据的交通路况预测的研究与实现[D].武汉:华中科技大学,2019.
- [8]YAO E J, WANG X W, YANG Y, et al. Traffic flow estimation based on toll ticket data considering multitype vehicle impact[J]. Journal of Transportation Engineering, Part A: Systems, 2021, 147(2): 04020158. DOI: 10.1061/jtepbs.0000488.
- [9]KE R M, FENG S, CUI Z Y, et al. Advanced framework for microscopic and lane-level macroscopic traffic parameters estimation from UAV video[J]. Intelligent Transport Systems, 2020, 14(7): 724-734. DOI: 10.1049/iet-its.2019.0463.
- [10]张新.基于无人机视频的交通参数提取方法及应用[D].上海:同济大学,2014.
- [11]LI J L, XU Z G, FU L, et al. Domain adaptation from daytime to nighttime: a situation-sensitive vehicle detection and traffic flow parameter estimation framework[J]. Transportation Research Part C: Emerging Technologies, 2021, 124: 102946. DOI: 10.1016/j.trc.2020.102946.
- [12]梁奇.高速公路服务区满意度与驶入选择行为研究[D].重庆:重庆交通大学,2019.
- [13]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. DOI: 10.19721/j.cnki.1001-7372.2022.03.017.
- [14]杨俊闯,赵超. k -means聚类算法研究综述[J].计算机工程与应用,2019,55(23):7-14. DOI: 10.3778/j.issn.1002-8331.1908-0347.
- [15]周玉,朱文豪,房倩,等.基于聚类的离群点检测方法研究综述[J].计算机工程与应用,2021,57(12):37-45. DOI: 10.3778/j.issn.1002-8331.2102-0167.
- [16]吉静,顾承华,翟希,等.基于交通状态指数的城市道路交通拥堵评价体系研究[C]//中国智能交通协会第十一届中国智能交通年会大会论文集.重庆:中国智能交通协会,2016:931-939.
- [17]张南,黄正国,叶彭姚,等.基于车牌数据的行程速度特性及交通状态评估[J].综合运输,2019,41(5):63-70.
- [18]WANG Z J, ZANG C, YANG C L. Multi-indicator Road traffic state assessment based on path state[EB/OL].[2022-08-20]. <http://dx.doi.org/10.12783/dtcse/icmsa2018/23259>.
- [19]胡建荣,何磊.基于尖点突变理论的高速公路交通流状态判别方法[J].中国公路学报,2017,30(10):137-144. DOI: 10.19721/j.cnki.1001-7372.2017.10.017.