

# 基于门架数据的高速公路拥堵识别研究

余 畅<sup>1</sup>, 苏跃江<sup>1,2</sup>, 韦清波<sup>3</sup>

(1. 广州市交通运输研究院有限公司, 广州 510635; 2. 华南理工大学 土木与交通学院, 广州 510614;  
3. 广州市公共交通数据管理中心有限公司, 广州 510620)

**摘要:**有效识别高速公路拥堵情况, 能为高速公路的交通管控及建设规划提供有效依据。基于高速公路门架数据, 提出了一套基于行程时间阈值的拥堵识别分析方法, 提取车辆的行程时间集合, 利用聚类算法对路段的不同拥堵等级的行程时间阈值进行标定, 从而对高速公路拥堵情况进行判别。在实例分析中, 以广州市高速公路网为研究对象, 开展了拥堵识别分析。结果表明, 广州市高速公路总体运行状态良好, 80%的时段路段处于顺畅状态。

**关键词:**智能交通; 门架数据; 聚类算法; 行程时间阈值; 拥堵识别

**中图分类号:** U491 **文献标志码:** A **文章编号:** 1671-1807(2023)22-0113-06

近年来, 随着机动车保有量的迅猛增长, 高速公路拥堵愈发频繁发生, 已成为影响交通安全和运行效率的主要问题。有效识别高速公路拥堵情况, 准确掌握高速公路交通运行状态, 是对高速公路进行科学的交通管理控制以及建设规划的关键前提, 对于改善高速公路拥堵状况、提升运行效率和交通安全具有重要意义。

关于高速公路的交通拥堵识别分析, 国内外学者对于拥堵识别算法进行了大量研究, 主要可以归纳为直接检测算法和间接识别算法两大类。直接检测算法主要是基于各类视频图像识别技术对交通拥堵事件进行检测<sup>[1-3]</sup>, 该类算法可以从微观角度开展交通拥堵识别分析, 但在实际应用中成本较高。间接识别算法主要根据拥堵对交通流的影响来识别拥堵, 主要包括基于交通流模型的算法以及基于智能算法的算法<sup>[4]</sup>, 其中交通流模型中常用于拥堵识别的度量参数有道路交通流车辆速度、车辆密度、占有率、车流量和行程时间等<sup>[5-7]</sup>, 交通流模型能够定量分析交通运行情况, 但部分交通流模型参数较多导致复杂度较高; 智能算法主要包括深度神经网络<sup>[8]</sup>、支持向量机<sup>[9]</sup>、梯度提升<sup>[10]</sup>、聚类分析<sup>[11]</sup>等, 智能算法能够从数据层面准确识别交通运行状态, 但无法探究交通内部的运行机理。

数据源方面, 目前被应用于高速公路拥堵识别的

数据来源多样, 主要包括车检器数据<sup>[12]</sup>、视频数据<sup>[13]</sup>、收费站流水数据<sup>[14]</sup>、移动通信数据<sup>[15]</sup>及多源数据融合等类型<sup>[16]</sup>。张腾月<sup>[17]</sup>基于高速收费流水数据, 分匝道和主线两部分对行程时间进行估算并求取交通流参数, 从而进行交通状态识别, 但由于收费站流水数据颗粒度较粗, 对于细分路段的分析精确度不高; 韩坤林<sup>[18]</sup>提出一种决策级融合算法对车检器数据和收费数据进行融合, 以判别高速公路的交通异常状态, 但国内高速公路现有检测设施布设稀疏, 数据检测率有待提高; 陈子瑜<sup>[19]</sup>提出了基于电子收费(electronic toll collection, ETC)交易数据的区段交通拥堵检测方法, 从区段交通流量和行程时间速度两个指标进行检测分析, 但其仅根据相关技术标准来设定指标阈值, 设定方法主观性较强, 不够精细。

随着ETC系统技术的发展成熟, ETC门架系统已广泛应用于高速公路, 其生成的门架流水数据体量大、覆盖范围广、颗粒度细、采集处理方便, 为高速公路的拥堵识别分析提供了新的路径。本文基于海量高速公路门架数据, 全面获取所有车辆出行信息, 结合交通流参数和智能算法的优势, 提出通过聚类算法获取各路段拥堵行程时间阈值的拥堵识别方法, 能够快速实现对高速公路全网全路段全时段的拥堵识别分析, 具有较强的普适性以及应用可行性。

**收稿日期:** 2023-08-11

**基金项目:** 广州市科技计划项目(202206010056)。

**作者简介:** 余畅(1992—), 男, 江西抚州人, 广州市交通运输研究院有限公司, 工程师, 硕士, 研究方向为交通大数据挖掘分析; 通信作者 苏跃江(1983—), 男, 贵州六盘水人, 广州市交通运输研究院有限公司, 华南理工大学土木与交通学院, 高级工程师, 硕士, 研究方向为交通大数据和公共交通; 韦清波(1984—), 男, 广东高州人, 广州市公共交通数据管理中心有限公司, 高级工程师, 硕士, 研究方向为交通大数据和公共交通。

## 1 研究方法

本文提出了一套基于门架数据的拥堵识别分析方法,研究框架如图1所示。第一,通过数据清洗及预处理流程,过滤错误及异常数据,并对缺漏数据进行补全;第二,以门架对应路段为单位组织数据,提取所有车辆在各门架路段上的行程时间集合。第三,基于目标路段所有的行程时间样本统计来定义合适的参数特征,利用 DBSCAN (density-based spatial clustering of applications with noise) 聚类算法,对目标门架路段的行程时间集合进行聚类分析,根据聚类结果的类别特征对路段的拥堵行程时间阈值进行标定;第四,对于各个门架路段进行以上聚类分析流程,形成各门架路段的拥堵行程时间阈值经验库;第五,对于目标时段路段,将路段对应行程时间阈值与目标行程时间样本集的统计特征进行对比分析,实现对任意时段的路段拥堵情况进行判别,从而掌握高速公路的拥堵时空分布特征。

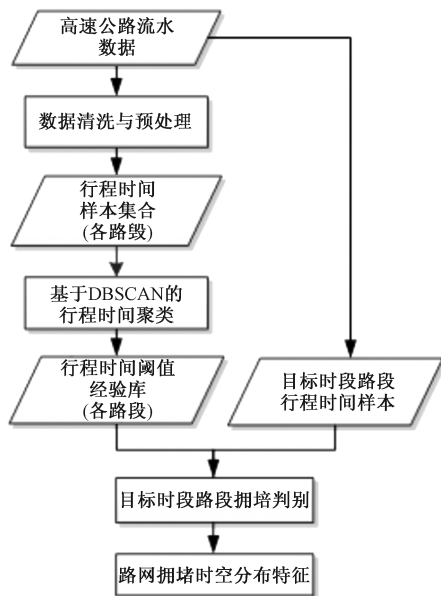


图1 高速公路拥堵识别分析框架

### 1.1 数据清洗与预处理

为实现路段级别的精细化拥堵识别,采用门架数据作为基础数据,门架数据主要包括车辆信息(车牌号、车辆类型、门架路段、所在高速、通过时间等)和交易信息(通行介质、计费里程等),如表1所示。由于高速电子不停车收费(electronic toll collection, ETC)门架系统存在的不完善,采集到的门架数据会出现出行漏检(过车时无识别)、错检(过车时被对向门架检测到)的情况。为了提高原始数据的质量,提升车辆出行信息获取的准确度,需要对原始数据进行以下数据清洗及预处理流程。

表1 门架数据示例

| 门架路段      | 通过时间                   | 车牌号      | 车辆类型 | 通行介质 | 计费里程/km |
|-----------|------------------------|----------|------|------|---------|
| 三元里-上高速匝道 | 2021-05-12<br>09:38:06 | 粤 AXXXXX | 1型客车 | OBU  | 0       |
| 三元里-白云新城  | 2021-05-12<br>09:39:52 | 粤 AXXXXX | 1型客车 | OBU  | 2.2     |
| 白云新城-新市   | 2021-05-12<br>09:40:45 | 粤 AXXXXX | 1型客车 | OBU  | 0.7     |

第一步:异常及冗余数据剔除。针对车牌号识别失败、计费里程为0 km等异常数据,以及同一车辆被同一门架检测多次产生的冗余数据,直接筛选出来进行剔除。

第二步:错检数据修正。针对通过车辆被对向门架检测而产生的错检数据,将门架路段信息融入高速路网模型,根据路网模型中的门架路段拓扑关系梳理出对向门架对应关系以及门架路段序列,判断异常过车逻辑;对于被对向门架检测的错误数据,将其替换为正确的门架路段。

第三步:漏检数据补充。针对通过车辆未被门架有效识别或数据异常导致的门架数据缺失,利用均值插补法进行补全。根据高速路网模型中的相邻门架对应逻辑关系及门架路段序列,补全缺失的门架路段,对于相邻门架之间的通行时间,则统计同时段在该路段(相邻门架之间)所有同车型通行车辆的平均通行时间,依次倒推各个缺失门架的过车时间,从而对缺失门架数据进行补全。

### 1.2 路段行程时间聚类

路段行程时间是一个能反映路段交通运行状态的重要指标。基于门架数据统计分析所有相邻门架对之间所有小客车(货车运行速度差异较大,本文研究暂不考虑)的行程时间,得到相邻门架路段在全天不同时段行程时间集合;该集合中包含多个不同车辆样本在全天不同时段(包含高峰时段和空闲时段)的行程时间,基本能反映路段在不同交通状态下对应的行程时间情况。为科学准确地从中提取出用于判别交通拥堵的行程时间阈值,利用 DBSCAN 聚类算法<sup>[20]</sup>对行程时间集合进行聚类,提取其中大部分小客车在不同交通状况下行驶所需的行程时间,从而实现对路段的拥堵行程时间阈值的标定。

DBSCAN 算法优点在于不需要定义聚类个数,对于不同特征的路段可以适应性划分交通状况类别数量,并且可以有效剔除数据集集中的异常点。DBSCAN 算法根据区域点密度的不同来定义聚类

集群,该算法包含最大密度可达距离  $\epsilon$  和最小样本点数  $\minPts$  两个全局参数。如果在一个样本点周围半径为  $\epsilon$  的范围内,有至少  $\minPts$  个样本点(密度足够高),那么这个点可以被标定为一个聚类中心点  $i_c$ ,表达为

$$|N_{\epsilon(i_c)}| \geq \minPts \quad (1)$$

式中:  $N_{\epsilon(i_c)}$  为到  $i_c$  的距离  $d(i_c, i)$  小于  $\epsilon$  的样本点的数量。

在利用 DBSCAN 算法对行程时间集合进行聚类流程前,首先需要对聚类的两个关键参数进行定义。最大密度可达距离  $\epsilon$  为相同交通状态下允许不同车辆样本的行程时间差值,表示合理控制聚类个数,将该参数定义为

$$\epsilon = (\max T_i - \min T_i) / \left( \frac{\sigma_{T_i}}{\text{avg} T_i} n \alpha \right) \quad (2)$$

式中:  $T_i$  为门架路段  $i$  对应的行程时间集合;  $\sigma_{T_i}$  为行程时间的方差;  $\frac{\sigma_{T_i}}{\text{avg} T_i}$  为行程时间样本的变异系数,反映数据的离散程度;  $n$  为划分拥堵等级的最多个数;  $\alpha$  为可达距离的修正系数。

最小样本点数  $\minPts$  用于判别聚类类别的有效性,排除个别噪声数据干扰,该参数定义为

$$\minPts = \mu_i / \beta n \quad (3)$$

式中:  $\mu_i$  为门架路段  $i$  对应的过车样本数量;  $n$  为划分拥堵等级的最多个数;  $\beta$  为允许样本量偏差的修正系数。

为了科学地设定参数中的修正系数取值,采用不同的修正系数取值对聚类结果进行了敏感性测试,最终将可达距离的修正系数  $\alpha$  设为 4,样本量偏差的修正系数  $\beta$  设为 10。

利用 DBSCAN 算法进行路段行程时间聚类的具体流程如下:

1)按门架对应路段提取相应的行程时间样本集合。

2)对目标路段的行程时间集合中随机提取一条未访问样本,形成一个簇并将该样本标记为已访问。

3)以递归的方式,遍历获取下一条未访问样本与当前各个簇的中心的是否密度可达(距离小于  $\epsilon$ )。

4)若存在密度可达的类则将该样本扩展这个簇的密度区域,若不存在与该样本密度可达的簇,则将这条样本形成一个新的簇。

5)重复 3)步直到集合中所有样本都访问过,对

于每个簇,如果总样本数少于  $\minPts$ ,则该簇中的样本标记为噪声,否则确认为一个有效聚类。

6)将各个聚类中心点的样本平均值按大小排序,得到目标路段的不同拥堵等级对应的行程时间。

7)按 1)~6)步处理所有门架路段的行程时间集合,形成全路段的不同等级拥堵的行程时间阈值经验库。

### 1.3 路段拥堵判别

对所有路段进行以上行程时间聚类流程,形成各路段的拥堵行程时间阈值经验库。对于需要分析的目标时段及路段,首先获取该路段的行程时间阈值集合  $\{P_1, P_2, \dots, P_n\}$ ,对于某行程时间值  $i$  对应的拥堵等级  $C_i$  划分标准如下:

$$C_i = \begin{cases} C_1, i \leq P_1 \\ C_2, P_1 < i \leq P_2 \\ \vdots \\ C_n, P_{n-1} < i \leq P_n \end{cases} \quad (4)$$

提取其中所有小客车的行程时间样本,将该行程时间样本集合  $\{i_1, i_2, \dots, i_n\}$  的统计特征与目标路段对应的拥堵行程时间阈值进行对比:统计所有目标样本的行程时间均值  $\bar{i}$ ,根据路段相应的拥堵行程时间阈值经验库,找出行程时间均值所对应的拥堵等级;按照行程时间阈值,对所有目标样本分别进行划分,找出划分样本量最多(众数)的拥堵等级;综合对比两次判定结果,将两次结果的平均值四舍五入取整作为分析目标对应拥堵等级。利用以上方法,实现对任意时段任意路段的拥堵等级判别,从而掌握高速公路的拥堵情况。

## 2 实例分析

### 2.1 路段拥堵识别分析

采用 2021 年 5 月 10 日至 5 月 16 日一周的广州市高速公路门架数据作为研究基础数据集,数据集规模达到 2.84 亿条,平均每日 4 059 万条。按照前文提出的研究方法对数据集进行处理,展开对广州市高速公路的拥堵识别及分析。以广州北环高速 S81 上的相邻门架对“沙贝立交虚拟拆分点-广清”和“广清-广花”之间对应路段为例,在一周时段内共提取 27.8 万条小客车过车的行程时间样本。对样本集进行了 DBSCAN 聚类流程,最终的聚类分类结果如图 2 所示。

根据聚类结果将该路段划分为 4 个拥堵等级,分别为:“顺畅”等级,对应行程时间阈值为 254 s,在该阈值以内路段处于顺畅的自由流状态下,行程时间样本呈现近似正态分布特征;“轻度拥堵”等级,

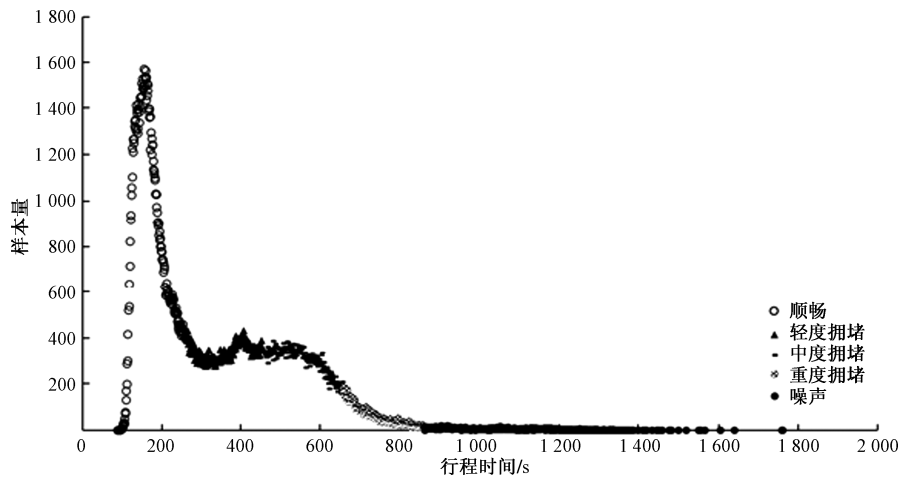


图2 路段行程时间样本聚类结果

对应行程时间阈值为 453 s;“中度拥堵”等级,对应行程时间阈值为 655 s;“重度拥堵”等级,对应行程时间阈值为 858 s;拥堵等级越高,行程时间偏离越大,样本量呈现逐渐减少的趋势;而对于行程时间大于 858 s 的等级,由于样本量未达到聚类参数的最小样本点数,被判定为噪声。

基于路段各拥堵等级对应的行程时间阈值,可对该路段的不同时段的拥堵情况进行判别。利用门架数据,对该路段在不同出行日不同时段平均行程时间进行统计,如图 3 所示。

经过路段拥堵判别流程,将各时段的行程时间统计特征与对应的阈值进行对比,最终对该路段的拥堵等级判别情况如图 4 所示。从拥堵时段分布情况来看,该路段工作日的早高峰时段拥堵情况较为严重,08:00—11:00 都处于中度及以上拥堵的状态,拥堵持续时间较长;周末的上午时段基本处于顺畅状态,拥堵主要出现在下午时段以及夜间 21:00—22:00 点时段。

## 2.2 路网拥堵特征分析

利用以上方法,对广州市所有高速公路门架路段进行拥堵判别分析。根据相邻门架对划分路段,共计划分出 1 307 个有效路段,对全网所有路段的全天各个时段进行拥堵判别,即每天对应 31 368 个路段样本。统计一周不同出行日内的路段拥堵等级分布情况,如表 2 所示。

从分布情况来看,全网在周五出现拥堵的路段时段最多,其次为周六,而周日出现拥堵的路段时段最少;总体而言,一周内有 80% 的路段时段处于顺畅状态,20% 的路段时段处于拥堵状态,其中大部分为轻度拥堵,整体运行状态良好。

从拥堵路段分布来看,全周累计发生拥堵的路

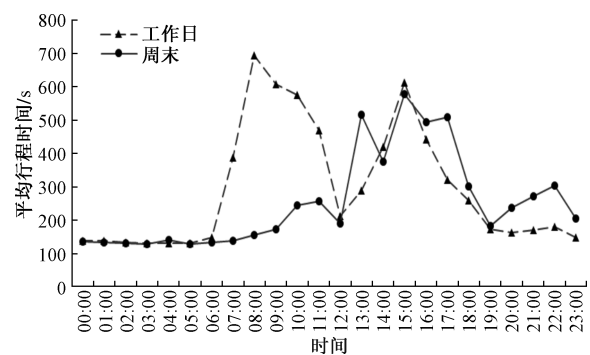


图3 各时段平均行程时间分布情况

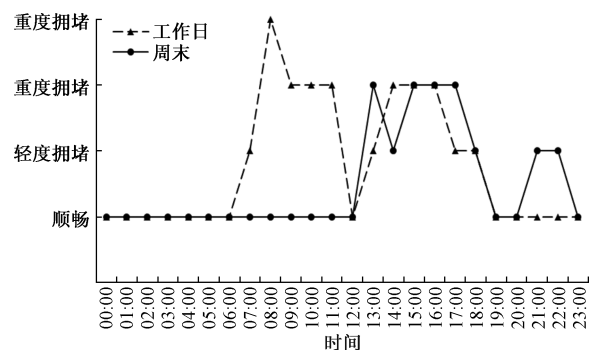


图4 各时段拥堵等级判别情况

表2 全路段时段拥堵等级日分布情况

| 日期   | 顺畅/个    | 轻度拥堵/个 | 中度拥堵/个 | 重度拥堵/个 | 拥堵合计/个 |
|------|---------|--------|--------|--------|--------|
| 周一   | 25 187  | 5 109  | 512    | 560    | 6 181  |
| 周二   | 25 544  | 4 962  | 439    | 423    | 5 824  |
| 周三   | 25 556  | 4 968  | 434    | 410    | 5 812  |
| 周四   | 25 268  | 5 135  | 508    | 457    | 6 100  |
| 周五   | 23 947  | 5 974  | 693    | 754    | 7 421  |
| 周六   | 24 995  | 5 254  | 568    | 551    | 6 373  |
| 周日   | 25 986  | 4 619  | 380    | 383    | 5 382  |
| 合计   | 176 483 | 36 021 | 3 534  | 3 538  | 43 093 |
| 占比/% | 80.4    | 16.4   | 1.6    | 1.6    | 19.6   |

段时段达到 43 093 段次,其中累计拥堵段次超过 1 000次的高速公路达到 17 条,如表 3 示。发生路段拥堵最多次的 5 条高速分别是 S4 华南快速、S41 机场高速、S81 北环高速、G1508 北二环高速、S73 南沙港快速。

结合门架流水数据以及高速收费站流水数据,根据车辆在各门架路段的通过时间以及进出站时间,可对各收费站匝道进行拥堵判别分析。从收费站匝道拥堵情况来看,进站上高速匝道全周累计发生拥堵 8 876 段次,表 4 所示为上高速匝道常发拥堵的收费站;出站下高速匝道全周累计发生拥堵 22 595段次,表 5 所示为下高速匝道常发拥堵的收费站。对于常发拥堵的收费站匝道,后期可考虑通过改扩建工程以及交通引导等措施进行优化治理。

表 3 常发拥堵的主要高速公路

| 编号 | 高速公路名称 | 拥堵段次  | 编号 | 高速公路名称 | 拥堵段次  |
|----|--------|-------|----|--------|-------|
| 1  | 华南快速   | 4 111 | 10 | 广惠高速   | 1 662 |
| 2  | 机场高速   | 3 546 | 11 | 广深高速   | 1 654 |
| 3  | 北环高速   | 2 815 | 12 | 沿江高速   | 1 550 |
| 4  | 北二环高速  | 2 707 | 13 | 广清高速   | 1 285 |
| 5  | 南沙港快速  | 2 246 | 14 | 增从高速   | 1 259 |
| 6  | 华快三期   | 1 949 | 15 | 肇花高速   | 1 234 |
| 7  | 广河高速   | 1 891 | 16 | 大广高速   | 1 178 |
| 8  | 东新高速   | 1 819 | 17 | 广珠东线   | 1 021 |
| 9  | 花莞高速   | 1 798 |    |        |       |

表 4 上高速匝道常发拥堵的收费站

| 编号 | 高速公路名称 | 收费站名称 | 拥堵段次 |
|----|--------|-------|------|
| 1  | 南沙港快速  | 亚运城站  | 306  |
| 2  | 华南快速   | 花城大道站 | 302  |
| 3  | 广深高速   | 科韵路站  | 283  |
| 4  | 机场高速   | 机场站   | 264  |
| 5  | 机场高速   | 人和站   | 232  |
| 6  | 北环高速   | 沙贝站   | 216  |
| 7  | 华南快速   | 中山大道站 | 208  |
| 8  | 北环高速   | 广汕站   | 168  |
| 9  | 华南快速   | 华快广汕站 | 168  |
| 10 | 花莞高速   | 知识城站  | 167  |

表 5 下高速匝道常发拥堵的收费站

| 编号 | 高速公路名称 | 收费站名称 | 拥堵段次 |
|----|--------|-------|------|
| 1  | 机场高速   | 机场站   | 553  |
| 2  | 华南快速   | 凤凰站   | 479  |
| 3  | 华南快速   | 中山大道站 | 347  |
| 4  | 机场高速   | 太成站   | 346  |
| 5  | 华南快速   | 黄埔大道站 | 336  |
| 6  | 机场高速   | 三元里站  | 334  |
| 7  | 广深高速   | 大观路站  | 328  |
| 8  | 北二环高速  | 香雪站   | 320  |
| 9  | 华南快速   | 土华站   | 298  |
| 10 | 广深高速   | 黄村站   | 290  |

3 结论

本文提出了利用 DBSCAN 聚类算法确定行程时间阈值来划分路段拥堵等级的拥堵识别方法,并基于广州市高速公路门架数据,对广州市高速公路进行了全路网的拥堵识别和特征分析。

1)通过对行程时间样本聚类进行拥堵等级划分的结果呈现以下特征:顺畅状态下路段处于自由流状态,行程时间样本呈现近似正态分布特征;拥堵状态下拥堵等级越高,行程时间偏离越大,样本量呈现逐渐减少的趋势。

2)从拥堵时段分布来看,工作日的早高峰拥堵情况较为严重,且持续时间较长;周末的上午基本处于顺畅状态,拥堵主要出现在下午时段。

3)从全路网拥堵的出行日分布来看,全网在周五的拥堵情况最为严重,其次为周六;总体而言,广州市高速路网一周内 80%的路段时段处于顺畅状态,出现拥堵的大多数为轻度拥堵,整体运行状态良好。

4)本文的拥堵识别方法能够有效识别全网范围内常发拥堵的路段以及收费站匝道,基于各路段的拥堵行程时间阈值经验库来判别拥堵,易于实际应用。

本文的研究能帮助全面掌握和有效评估高速公路的交通运行状况,为高速公路的改扩建规划以及优化提升提供有效的决策支撑。

参考文献

[1] 王启明,胥津铭,苏建,等.基于改进 ALEXNet 模型的路面状况识别方法研究[J].公路交通科技,2023,40(3):209-218.

[2] 袁鸽鸽.基于深度卷积神经网络的高速公路拥堵识别[D].西安:长安大学,2020.

[3] 刘宇峰.基于卷积神经网络的高速公路拥堵识别技术研究[J].信息与电脑,2022,14:82-85.

[4] 贾若,戴昇宏,黄霓,等.交通拥堵判别方法研究综述[J].华南理工大学学报(自然科学版),2021,49(4):124-139.

[5] BEKIARIS L N, RONCOLI C, PAPAGEORGIOU M. Highway traffic state estimation per lane in the presence of connected vehicles[J]. Transportation Research Part B: Methodological,2017,106:1-28.

[6] 沈家军,王昊,董长印,等.基于卡尔马滤波的高速公路交通状态估计模型构建及效果评价[C]//第十一届中国智能交通年会大会论文集.北京:电子工业出版社,2016:364-375.

[7] ZHENG F,JABARI S E,LIU H X,et al. Traffic state estimation using stochastic Lagrangian dynamics[J]. Transportation Research Part B: Methodological, 2018, 115:

- 143-165.
- [8] BENKARAOUDA O, THODI B T, YEO H, et al. Traffic data imputation using deep convolutional neural networks [J]. IEEE Access, 2020, 8: 104740-104752.
- [9] 李晓璐, 于昕明, 杜崇, 等. 基于权值优化的 FCM-MSVM 算法及其在高速公路状态判别中的应用[J]. 北京交通大学学报, 2018, 42(4): 72-78.
- [10] 王昊, 杨万波. 速度梯度模型的高速公路交通流状态估计方法[J]. 哈尔滨工业大学学报, 2015, 47(9): 84-89.
- [11] HERNANDEZ P Y, SAIFUZZAMAN M, BERT E, et al. Unsupervised incident detection model in urban and freeway networks[C]//2018 21st International Conference on Intelligent Transportation Systems (ITSC). Maui, Hawaii, USA: IEEE, 2018: 1763-1769.
- [12] 湛锐. 基于 ETC 的高速公路车检器数据修复与交通异常状态融合识别[D]. 重庆: 重庆大学, 2020.
- [13] 纪宇. 基于 AlexNet 模型的高速公路拥堵状态识别[D]. 西安: 长安大学, 2018.
- [14] 胡继启. 基于收费数据的高速公路交通拥堵判别与定位方法研究[D]. 长沙: 长沙理工大学, 2016.
- [15] 罗裕枫. 基于移动信令数据的高速路拥堵识别及预测模型的研究[D]. 重庆: 重庆邮电大学, 2018.
- [16] 刘廷让. 基于多源数据融合的高速公路路段交通拥堵检测方法研究[D]. 重庆: 重庆大学, 2020.
- [17] 张腾月. 基于收费数据的高速公路交通流状态识别研究[D]. 广州: 华南理工大学, 2017.
- [18] 韩坤林. 基于车检器及收费数据融合的高速公路异常状态识别研究[D]. 重庆: 重庆大学, 2014.
- [19] 陈子瑜. 基于大数据的异常检测研究——以高速公路拥堵检测为例[D]. 福州: 福建理工大学, 2022.
- [20] ESTER M, KRIEGEL H P, SANDER J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise [C]// Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. Portland, OR, USA: AAAI Press, 1996: 226-231.

## Research on Expressway Congestion Discrimination Based on Toll Gantry Data

YU Chang<sup>1</sup>, SU Yuejiang<sup>1,2</sup>, WEI Qingbo<sup>3</sup>

(1. Guangzhou Transport Research Institute Co. Ltd., Guangzhou 510635, China;

2. School of Civil Engineering and Transportation, South China University of Technology, Guangzhou 510614, China;

3. Guangzhou Public Transport Research Institute Co. Ltd., Guangzhou 510620, China)

**Abstract:** It can provide effective evidence for expressway traffic control and construction planning by effectively identifying congestion of expressway. A methodology of congestion discrimination was proposed based on travel time threshold. The travel time collections of all vehicles were extracted based on toll gantry data. The travel time collection was clustered and analyzed by clustering algorithm, and the travel time threshold of different congestion levels was calibrated according to the clustering results. The congestion condition will be discriminated based on the congestion travel time thresholds. The results indicates that the overall operation of Guangzhou expressways is good, with 80% of periods and sections being in a smooth state.

**Keywords:** intelligent transportation; toll gantry data; clustering algorithm; travel time threshold; congestion discrimination