

图9 C60混凝土坍落度、扩展度
Fig. 9 Slump and spread of C60 concrete

混凝土更优。这表明采用降黏措施优化后 C60 预应力混凝土力学性能均得到一定提升,均满足 7 d 达设计强度 60 MPa 的要求。

2.5.3 干缩性能

3 组混凝土干燥收缩率随时间变化结果如图 10 所示,由图 10 可知:C60 基准混凝土干缩稍大,120 d 混凝土干缩率达到了-0.040 1%,而 T1、T2 组混凝土各龄期干缩率与基准混凝土略有减小,但总体差别不大;T1、T2 组混凝土 120 d 干缩率分别为-0.039 4%和-0.038 4%,分别减小了 1.7%和 4.2%。结果表明,降黏措施对 C60 混凝土干缩性能影响不大。

3 结 语

(1)掺入 5~10 mm 粒径小石、I 级粉煤灰、引气剂、触变剂、降黏剂等 5 种低黏化措施均可使拌和物坍落度、扩展度增大,黏度、屈服剪切应力降低。其中掺入 2%触变剂对坍落度、扩展度改善效果最明显;掺入 15%的 I 级粉煤灰、0.001%液体降黏剂对降低拌和物黏度效果较为明显,黏度在120~127

表 11 C60 混凝土力学性能

Tab. 11 Mechanical properties of C60 concrete

编号	3 d			7 d			28 d		
	立方抗压强度/MPa	轴心抗压强度/MPa	轴压弹模/GPa	立方抗压强度/MPa	轴心抗压强度/MPa	轴压弹模/GPa	立方抗压强度/MPa	轴心抗压强度/MPa	轴压弹模/GPa
基准组	66.2			71.9	54.9	41.0	77.3	62.0	44.1
T1 组	67.1			72.1	56.3	42.9	81.8	63.7	45.7
T2 组	64.9			75.3	58.4	44.5	85.0	64.4	47.3

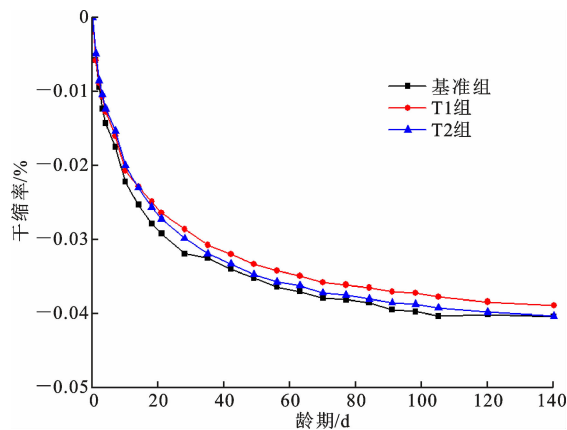


图 10 C60 混凝土干缩性能
Fig. 10 Dry shrinkage performances of C60 concrete

Pa·s,较基准组降低 21%~25%;而掺入 0.005%引气剂(调节含气量至 4.8%)对降低屈服剪切应力效果最佳,屈服剪切应力可低至 194 Pa,较基准组

降低 70%。

(2)低流态混凝土在泵管中以“柱塞流”的形式流动,其拌和物流变参数——黏度、屈服剪切应力的增大均会引起混凝土在管道中流动时泵送压力损失的增大;且对单位长度压力损失起重要影响的流变参数是黏度,屈服剪切应力更多影响流动启动时的起始压力。

(3)五种低黏化措施均可降低混凝土在管道中流动时的压力损失,其中掺入 0.001%液体降黏剂时作用效果最佳,30 和 50 m³/h 泵送速率下单位长度压力损失较基准组可分别降低 53%和 50%。

(4)掺入 5~10 mm 粒径小石、I 级粉煤灰、触变剂、降黏剂对混凝土抗压强度均无不利影响,混凝土后期强度反而会有所提高,满足 7 d 达设计强度 60 MPa 的要求;在保证混凝土强度的前提下,也可通过适当提高含气量的方式来改善混凝土拌和物的

流动性和流变性。

(5)优化推荐的2组低流态高性能C60预应力混凝土拌和物黏度在120~135 Pa·s之间、屈服剪切应力在235~252 Pa之间,7 d抗压强度在72.1~75.3 MPa之间,120 d干缩率约为-0.04%。

(6)基于室内试验和CFD仿真分析,本文研究了不同低黏化措施对高强混凝土流变性能、高扬程泵送压力损失影响规律,后期将结合实际工况优化泵管几何模型与边界条件、实测泵管各部位压力参数,进一步研究低黏化措施对混凝土在泵管中流速、压强变化特征及泵后性能的影响。

参考文献:

References:

- [1] 缪昌文,刘建忠.应用高强混凝土应注意的几个问题[J].施工技术,2013,42(10):1-5.
MIAO Chang-wen, LIU Jian-zhong. Issues for attention during the application of high strength concrete[J]. Construction Technology, 2013, 42(10): 1-5.
- [2] 宋沛霖.高强混凝土用高保坍降粘型聚羧酸减水剂的试验研究[D].济南:山东建筑大学,2019.
SONG Pei-lin. Experimental study on the use of high anti-collapse and viscosity-reducing polycarboxylate superplasticizer for high strength concrete[D]. Jinan: Shandong Jianzhu University, 2019.
- [3] KWAN A K H, WONG H H C. Effects of packing density, excess water and solid surface area on flowability of cement paste[J]. Advances in Cement Research, 2008, 20(1): 1-11.
- [4] ZHAN J B, AN X H, DING N. Effect of fine aggregate characteristics on the thresholds of self-compacting paste rheological properties[J]. Construction and Building Materials, 2016, 116: 355-365.
- [5] 张磊,王超,李京川,等.泵送混凝土压力损失试验研究[J].施工技术,2014,43(6):50-53.
ZHANG Lei, WANG Chao, LI Jing-chuan, et al. Experimental study on pressure loss of pumping concrete[J]. Construction Technology, 2014, 43(6): 50-53.
- [6] FERRARIS C F, OBLAK H, HILL R. The influence of mineral admixtures on the rheology of cement paste and concrete[J]. Cement & Concrete Research, 2001, 31(2): 245-255.
- [7] 沈锐.降低UHPC拌合物粘度的技术措施[D].重庆:重庆大学,2018.
- SHEN Rui. Technical measures to reduce the viscosity of UHPC mixtures[D]. Chongqing: Chongqing University, 2018.
- [8] STRUBLE L, SUN G K. Viscosity of Portland cement paste as a function of concentration[J]. Advanced Cement Based Materials, 1995, 2(2): 62-69.
- [9] 徐永波.高强混凝土的性能特点分析及其改进措施[J].安徽建筑,2015,22(3):174-176.
XU Yong-bo. Performance features analysis of high strength concrete and its improvement measures[J]. Anhui Architecture, 2015, 22(3): 174-176.
- [10] 魏素梅,关战军.配制高强混凝土的关键技术[J].四川建材,2013,39(4):24-25.
WEI Su-mei, GUAN Zhan-jun. Key technology of preparing high strength concrete[J]. Sichuan Building Materials, 2013, 39(4): 24-25.
- [11] ARORA A, AGUAYO M, HANSEN H, et al. Microstructural packing- and rheology-based binder selection and characterization for ultra-high performance concrete (UHPC)[J]. Cement and Concrete Research, 2018, 103: 179-190.
- [12] 张倩倩,刘建忠,周华新,等.超高性能混凝土流变特性及其对纤维分散性的影响[J].材料导报,2017,31(23):73-77.
ZHANG Qian-qian, LIU Jian-zhong, ZHOU Hua-xin, et al. Rheological properties of ultra-high performance concrete and its effect on the fiber dispersion within the material[J]. Materials Review, 2017, 31(23): 73-77.
- [13] SHI Y X, MATSUI I, FENG N Q. Effect of compound mineral powders on workability and rheological property of HPC[J]. Cement and Concrete Research, 2002, 32(1): 71-78.
- [14] ZHU W Z, GIBBS J C. Use of different limestone and chalk powders in self-compacting concrete[J]. Cement and Concrete Research, 2005, 35(8): 1457-1462.
- [15] WANG C, YANG C H, LIU F, et al. Preparation of ultra-high performance concrete with common technology and materials[J]. Cement and Concrete Composites, 2012, 34(4): 538-544.
- [16] SONG Q L, YU R, WANG X P, et al. A novel self-compacting ultra-high performance fibre reinforced concrete (SCUHPFRC) derived from compounded high-active powders[J]. Construction and Building Materials, 2018, 158: 883-893.
- [17] 刘建忠,张倩倩,刘加平,等.一种高与超高强混凝土

- 降粘剂、其制备方法及其应用:中国,CN104529225A[P]. 2015-04-22.
- LIU Jian-zhong, ZHANG Qian-qian, LIU Jia-ping, et al. Method for preparing viscosity-reduction-type polycarboxylic acid superplasticizer and application of viscosity-reduction-type polycarboxylic acid superplasticizer; China, CN104529225A[P]. 2015-04-22.
- [18] 西川朋孝,宇野亨,山本弘一,等. 多羧酸混凝土外加剂:中国,CN1784369[P]. 2006-06-07.
- TOMOTAKA N, HARU U, KOICHI Y, et al. Polycarboxylic acid concrete admixture; China, CN1784369A[P]. 2006-06-07.
- [19] EDE A N. The resistance of concrete pumped through pipelines[J]. Magazine of Concrete Research, 1957, 9(27):129-140.
- [20] 刘 豫. 新拌混凝土流变的测量、模型及其应用[D]. 北京:中国建筑材料科学研究总院, 2020.
- LIU Yu. Rheological measurement, models and their applications for fresh concrete [D]. Beijing: China Building Materials Academy, 2020.
- [21] 黎梦圆. 新拌水泥基材料非线性流变模型的研究与应用[D]. 北京:清华大学, 2018.
- LI Meng-yuan. Nonlinear rheological modeling of fresh cementitious materials [D]. Beijing: Tsinghua University, 2018.
- [22] 阎培渝,黎梦圆,韩建国,等. 新拌混凝土可泵性的研究进展[J]. 硅酸盐学报, 2018, 46(2):239-246.
- YAN Pei-yu, LI Meng-yuan, HAN Jian-guo, et al. Recent development on pumpability of fresh concrete [J]. Journal of the Chinese Ceramic Society, 2018, 46(2):239-246.
- [23] KAPLAN D, LARRARD F D, SEDRAN T. Design of concrete pumping circuit[J]. ACI Materials Journal, 2005, 102(2):110-117.
- [24] FEYS D, KHAYAT K H, KHATIB R. How do concrete rheology, tribology, flow rate and pipe radius influence pumping pressure? [J]. Cement and Concrete Composites, 2016, 66:38-46.
- [25] 周玉鹏. 新拌混凝土流动行为及泵送吸料性能提升研究[D]. 湘潭:湘潭大学, 2021.
- ZHOU Yu-peng. Research on fresh concrete flow behavior and the improvement of pumping suction performance[D]. Xiangtan: Xiangtan University, 2021.
- [26] 魏子易. 基于 CFD 模拟的新拌混凝土泵送过程中流动行为的研究[D]. 北京:中国建筑材料科学研究总院, 2019.
- WEI Zi-yi. Study on flow behavior of fresh concrete during pumping using CFD simulation[D]. Beijing: China Building Materials Academy, 2019.
- [27] 李 悦,田小璇,金彩云,等. 基于 CFD 的泵送混凝土流变性能数值模拟研究[J]. 硅酸盐通报, 2020, 39(4):1139-1144, 1152.
- LI Yue, TIAN Xiao-xuan, JIN Cai-yun, et al. Numerical simulation of rheological properties of pumped concrete based on CFD[J]. Bulletin of the Chinese Ceramic Society, 2020, 39(4):1139-1144, 1152.
- [28] MORSI S A, ALEXANDER A J. An investigation of particle trajectories in two-phase flow systems [J]. Journal of Fluid Mechanics, 1972, 55(2):193.
- [29] CHOI M, ROUSSEL N, KIM Y, et al. Lubrication layer properties during concrete pumping[J]. Cement and Concrete Research, 2013, 45:69-78.
- [30] JO S D, PARK C K, JEONG J H, et al. A computational approach to estimating a lubricating layer in concrete pumping[J]. Computers, Materials & Continua, 2012, 27(3):189-210.

王建军,李冬怡,王 赛,等. 基于 EWM-TOPSIS 的城市卡口地点车速数据诊断[J]. 长安大学学报(自然科学版),2023,43(3):67-75.
WANG Jian-jun,LI Dong-yi,WANG Sai,et al. Diagnosis for urban bayonet spot speed data based on EWM-TOPSIS method[J]. Journal of Chang'an University(Natural Science Edition),2023,43(3):67-75.

DOI:10.19721/j.cnki.1671-8879.2023.03.007

基于 EWM-TOPSIS 的城市卡口地点 车速数据诊断

王建军,李冬怡,王 赛,李 鹏,刘明雨

(长安大学 运输工程学院,陕西 西安 710064)

摘 要:为解决因交通信息监控设备年久失修导致时空数据不完备,对交通时空数据质量状况进行综合诊断,为后续数据质量恢复与应用研究奠定数据基底。首先利用地点车速数据的时序特性和箱线图判别准则(Boxplot)客观判别了问题数据类型,其次针对不同问题数据类型特点构建了多属性综合评价指标图谱,采用熵权法(EWM)为各评价指标赋予信息熵权值,利用逼近理想解(TOPSIS)评价法得到了各条路段地点车速数据的贴近度并给予排序,最后综合道路属性及贴近度排序结果评价了区域路网地点车速数据质量状况。以 2021 年乌鲁木齐市 1 周的原始卡口交通数据为例,对应不同问题数据类型提出了多个数据质量评价指标,利用多属性评价指标图谱及道路属性对 EWM-TOPSIS 综合评价方法进行验证。研究结果表明:问题数据判别算法能够精确判别出冗余数据、缺失数据及异常数据 3 类问题数据,对应的评价指标均值分别为 5.66%、10.04%、5.04%,综合可用性指标均值为 84.66%,各路段数据质量整体良好;利用 EWM-TOPSIS 综合评价方法得到的路段地点车速数据贴近度均值为 0.25,区域路网地点车速数据质量指数均值为 0.19,道路属性是影响区域路网数据质量的关键因素之一。所提数据质量诊断方法能够辨识原始卡口交通问题数据类型,综合道路属性客观评价区域路网的整体数据质量,为交通数据质量诊断及未来应用服务提供了新的解决思路。

关键词:交通工程;城市交通信息监控设备;EWM-TOPSIS;卡口地点车速;多属性综合评价指标

中图分类号:U491.1

文献标志码:A

文章编号:1671-8879(2023)03-0067-09

Diagnosis for urban bayonet spot speed data based on EWM-TOPSIS method

WANG Jian-jun, LI Dong-yi, WANG Sai, LI Peng, LIU Ming-yu

(School of Transportation Engineering, Chang'an University, Xi'an 710064, Shaanxi, China)

Abstract: In order to solve the incompleteness of spatiotemporal data caused by the disrepair of traffic information monitoring equipment, a comprehensive diagnosis for traffic spatiotemporal data quality was carried out, which lays a solid data base on subsequent data quality recovery and application research. First, the time series characteristics of spot speed data and Boxplot criterion

收稿日期:2022-12-09

基金项目:国家自然科学基金项目(52172338);长安大学研究生科研创新实践项目(300103722058)

作者简介:王建军(1970-),男,陕西乾县人,教授,博士研究生导师,E-mail:wjjun16@chd.edu.cn.

were used to objectively identify the problem data types. Secondly, a multi-attribute comprehensive evaluation indexes map was constructed according to the characteristics of different problem data types, then the entropy weight method (EWM) was utilized to assign information entropy weights for each evaluation indexes. The technique for order preference by similarity to ideal solution method (TOPSIS) was applied to obtain the closeness degree of spot speed data on road segments giving the closeness degree ranking as well. Finally, the spot speed data quality on regional road network was comprehensively evaluated by combining the closeness degree ranking results with the road attributes. The original bayonet traffic data of Urumqi road network for one week in 2021 was taken as an example, multiple data quality evaluation indicators were proposed by different problem data types. The EWM-TOPSIS method was verified through multi-attribute evaluation indexes map and road attribute. The results show that the problem data discrimination algorithm can accurately distinguish redundant data, missing data and abnormal data, meanwhile, the corresponding evaluation indexes are 5.66%, 10.04%, 5.04% respectively, and the average comprehensive availability value is 84.66%. The data quality for each road segment is better as a whole. The mean closeness degree of spot speed data on road segments is 0.25, and the mean quality index of spot speed data on regional road network is 0.19 evaluated by EWM-TOPSIS comprehensive evaluation method. The road attribute is also the one of key factors affecting the spot speed data quality on regional road network. The proposed method can identify the original traffic problem data types, objectively evaluate the spot speed data quality for regional road network by considering the road attributes, and provide a novel solution for traffic data quality diagnosis and future application services. 1 tab, 6 figs, 27 refs.

Key words: traffic engineering; urban traffic information monitoring equipment; EWM-TOPSIS; bayonet spot speed; multi-attribute comprehensive evaluation index

0 引言

以“交通工程+IT”为新技术理念,以海量交通时空数据为引擎驱动实现城市交通智能化综合治堵。高质量交通时空数据能够反映交通运行态势的动态变化规律,为未来交通运行状态研判、评估及分析提供丰富的数据源,为城市智能交通系统的“感知现在、预测未来、面向服务”理念提供最优质的数据支撑。因此,数据质量状况优劣程度对实时交通时空数据的诊断、恢复、挖掘及应用研究有重要影响。然而,交通信息监控设备由于自身性能问题、外界环境影响及数据体量庞大等因素产生损坏、缺失、异常、冗余等若干类问题数据,部分问题数据潜在的时空信息损失会影响道路交通运行态势分析结果的准确性,如何对不完整的原始交通时空数据质量进行全面性、系统性诊断及评估,以监测原始交通时空数据质量已成为一项亟需解决的问题。

数据质量评价作为数据管理过程中的首要环节之一,在政府开放数据、统计学、医疗领域中应用广泛。张晓娟等针对面向政府开放平台元数据建立了数据质量评估指标体系,并基于主成分分析评价方

法对数据质量水平进行评测,最后根据评测结果分析得到了导致政府开放平台元数据总体质量状况不佳的原因^[1]。白永昕等从数据系统偏差、安全性及非随机性方面出发,利用 Bootstrap 方法建立了多维数据质量评估体系,并以统计数据为研究对象进行数据质量评估验证^[2]。D'Ambrosio 等从完整性及可靠性角度出发,以质量得分为评价指标对医疗数据质量进行评价,根据评价结果总结了导致数据质量差异化的原因^[3]。Tute 等应用开源软件 Open CQA,提出了一种基于标准化数据定义的医疗数据质量评价方法,研究结果可为医疗保健中的数据质量评估提供科学的理论基础^[4]。

随着城市道路网规模的不断扩张,交通时空数据当量达到海量级,数据质量评价研究逐渐延伸至交通领域,近年来有关交通信息监控设备及浮动车移动数据质量判别与修复研究较多。Li 等借助 GPS 数据及相邻车道交通流量比率识别了城市交叉口环形检测器的异常数据^[5]。Wang 等基于 2 个指标评价了浮动车数据(FCD)的覆盖能力,并采用改进的离群点算法判别了 FCD 异常数据^[6]。Maghrour 等认为从多个交通流参数阈值角度判别

环形检测器多维交通数据质量是十分有必要的^[7]。谢聪聪等从问题数据类型出发,提出了基于时空流量守恒定律的非直观错误数据判定方法,继而采用了基于认知评价理论的综合评价体系,分析城市快速路检测器时空数据质量^[8]。陆百川等根据传统阈值理论和马氏距离诊断出微波检测器中的异常数据,经剔除后采用灰色-遗传小波神经网络修复了故障数据^[9]。

目前交通数据质量评价研究还处于重视评价指标选取、轻视评价方法构建的阶段。美国交通部于2008年针对交通数据质量评价研究出台了《数据质量白皮书》,从6个维度评价指标体系出发,对交通流数据进行综合评价并应用于智能交通系统中^[10]。Park等在现有评价指标的基础上,提出了时空一致性和数据缺失严重性指标,采用熵权-层次分析法评价高速公路检测器数据质量^[11]。Li等分别基于交通流参数模型及专家建议系统选取了6个评价指标,通过自适应模糊推断理论综合评价了线圈检测器数据质量^[12]。庄广新等提出了准确度、完整度和时效性3个交通数据质量评价指标,基于灰色聚类的层次分析法评价了城市快速路多源交通速度数据质量^[13]。Bae等根据时空特性提出了4种时间尺度数据类型下双R指标(repeatability & reproducibility),分析了2种交通场景下的环形检测器数据精度^[14]。

无线射频识别设备(radio frequency identification, RFID)及城市卡口监控设备作为新兴的交通数据采集手段,已成为学者们重点关注的对象。杨越思等将相同车辆的相邻过车时间差作为判断阈值,解决了不同RFID冗余数据类型的识别问题,并对比并分析了在7种不同时间汇集度下城市路网的RFID时间覆盖率^[15-16]。龙思卿等针对RFID数据特点提出了一套评价指标体系,并将各指标赋权后采用诱导有序加权(IOWA)算子综合评价了RFID数据的有效性^[17]。针对交通数据有效性评价涵盖不全面问题,赵盼明从微观与宏观视角出发,设计了全面的有效性评价指标体系,最后基于TOPSIS的灰色关联分析结合客观自主式评价方法综合评估了RFID数据质量的有效性^[18]。付凤杰分析高清卡口设备数据质量,剖析数据内在特征,进而实现了路段行程时间估算任务^[19]。孙猛等将卡口数据时间关联信息与路网空间逻辑关系结合,构建了交通流多参数实时提取方法,达到了较优的参数提取准确率^[20]。

已有研究对现有交通问题数据判别和数据质量评价研究过程相对独立,尚未形成系统性的数据质量诊断体系;其次,诊断对象主要以传统的交通监控

设备为主,面向城市卡口监控设备的问题数据诊断研究甚少;同时,复杂的城市道路网络层次结构造成现有评价客体的层次较为模糊和评价指标过于片面,尤其对于冗余率指标鲜有考虑,评价方法不具备综合性及客观性。为全面诊断地点车速数据质量状况,本文以乌鲁木齐市某一区域路网卡口交通数据为研究对象,在问题数据类型判别的基础上,通过分析各类问题数据特点构建适用于城市路段和路网级别的数据质量多属性评价指标图谱,其次综合考虑道路属性因素,并基于EWM-TOPSIS综合评价方法开展城市区域路网交通数据整体质量诊断的研究,以期在后续的时空地点车速数据质量恢复等应用方向发挥数据的最大潜在价值^[20-23]。

1 地点车速数据质量诊断

交通数据质量反映了被检测信息的真实性与准确性,为了确保交通数据质量恢复技术的有效性与可靠性,对交通数据质量进行智能诊断是不可缺少的环节。数据质量诊断作为全过程管控的重要组成部分,是解决数据质量问题的首要步骤,交通数据质量诊断旨在检查原始数据集中是否存在“脏数据”,通过评估数据整体的质量进一步反映城市道路数据采集设备的运行状况,进而提出更为合理的数据质量恢复措施。

1.1 问题数据判别算法

“脏数据”伴随数据采集设备故障、车辆违法行驶及恶劣气候等因素而产生,一般包含异常数据、冗余数据及缺失数据。异常数据指的是变化趋势长期偏离正常状态的离群速度数据^[24]。冗余数据是对于某一个观测对象经过某一检测设备时采集到的2条乃至多条重复或相似速度数据。缺失数据指由于检测设备自身故障或其他外界因素影响未收集到完整速度数据。“脏数据”会导致总样本量减少以及整体价值降低,剩余可用样本量会对研究结果产生误导性结论,因此,针对每一种问题数据设计不同的判别算法可精确定位问题数据类型。

卡口地点车速(traffic bayonet spot speed, TBSS)指在某一时间段内通过某一卡口监控设备时所有车辆瞬时速度的调和平均值,能够直观反映道路断面的交通状态,即

$$v_n^t = \left(\frac{1}{K} \sum_{n=1}^N \sum_{i=1}^K v_m \right)^{-1} \quad (1)$$

式中: v_n^t 为采集时间间隔 t 内,断面 n 的交通时空地点车速, $n=1,2,\dots,N$, N 为断面总数; K 为经过某

一卡口断面的车辆总数; v_{ni} 为经过断面 n 处第 i 辆车的瞬时速度, $i=1,2,\dots,K$ 。

图1为TBSS冗余及缺失数据判别流程。首先将原始TBSS数据按照相同车牌分组,继而对相同车牌的采样时间排序,如果出现车牌及采样时间相同的数据则判别为冗余重复数据;其次计算组内相邻采样时间差,如果采样时间差小于设定的阈值,则标记为冗余相似数据,由于车辆以最低速5 km/h通过距离交叉口进口道20 m的监控设备时所用最短采样时间为14 s,为保证数据判别的可靠性,本文将采样时间差阈值设定为60 s^[15,19]。交通信息监控设备一般以固定时间粒度的方式采集数据,将冗余数据约简后,通过扫描固定时间粒度内的数据能够判别缺失数据。将固定时间粒度设置为2 min,对剩余TBSS非冗余数据进行聚合操作,基于此,1 d时间区间可定义为 $T_M=[2(M-1),2M]$, $M=1,2,\dots,720$,实际采样时间粒度为 t_m ,如果 t_m 不在 T_M 内,则将 t_m 时的数据判为缺失数据, m 、 M 分别为实际和理论数据点个数。

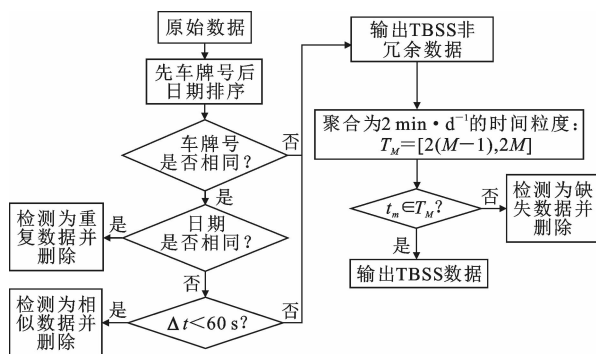


图1 TBSS冗余及缺失数据判别流程

Fig. 1 Discrimination process of TBSS redundancy and missing data

图2为基于Boxplot判别准则的TBSS异常数据检测。3sigma判别准则一般以数据的均值及标准差为判别标准,检测数据以服从正态分布为前提,在置信概率为99.73%下,测量值误差超过3倍标准差时判定为异常数据,但是该判别准则的均值及标准差具有较差抗耐性,使得判别精度粗糙^[25]。Boxplot判别准则无需服从特定分布,且75%位数 Q_3 及25%位数 Q_1 具有一定的耐抗性,四分位距 I_{QR} 具有一定鲁棒性,受异常值影响较小,因此采用Boxplot判别准则更能提高判别异常数据的精度及判别阈值的抗耐性。首先计算图1输出TBSS数据的75%位数 Q_3 、25%位数 Q_1 及四分位距 $I_{QR} = Q_3 - Q_1$,若输出TBSS数据超出阈值范围 $Q_1 - 1.5I_{QR}$

或 $Q_3 + 1.5I_{QR}$,则认为数据异常。

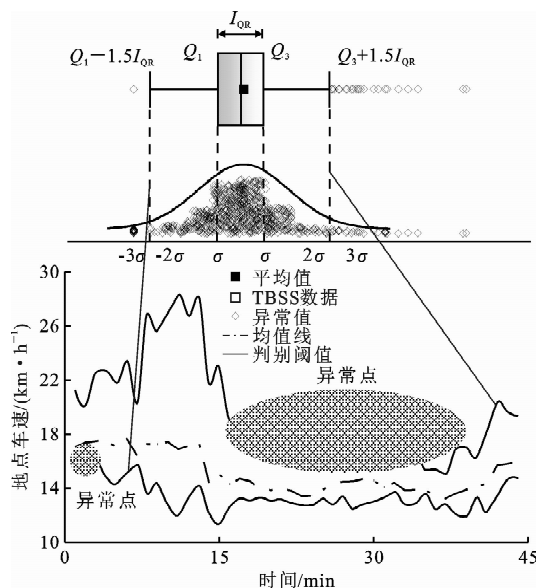


图2 基于Boxplot判别准则的TBSS异常数据检测

Fig. 2 TBSS abnormal data detection based on

Boxplot criterion principle

1.2 TBSS数据质量多属性评价指标图谱构建

城市区域路网由若干等级路段构成,具有鲜明的层次结构,因此本文按照由城市路段至区域路网的空间层次开展TBSS数据质量综合评价研究。针对问题数据类型,从冗余性评价、完整性评价、有效性评价及可用性评价方面构建的TBSS数据质量多属性评价指标图谱如图3所示。

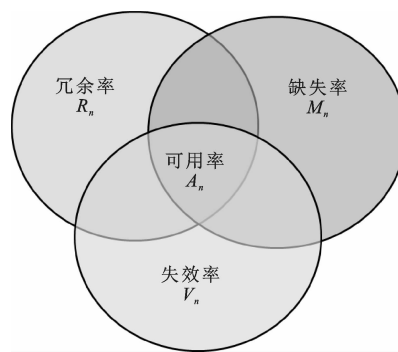


图3 TBSS数据质量多属性评价指标图谱

Fig. 3 Multi-attribute evaluation indexes map of TBSS data quality

多属性评价指标体系可表示为

$$\begin{cases} R_n = \frac{\sum_{n=1}^N \sum_{l=1}^L f_n(l)}{N_n^{Rt}} \times 100\% \\ M_n = \frac{N_n^m}{N_n^{Mt}} \times 100\% \\ V_n = \frac{N_n^i}{N_n^{Vt}} \times 100\% \\ A_n = (1 - R_n)(1 - M_n)(1 - V_n) \end{cases} \quad (2)$$

$$f(l) = \begin{cases} 1 & \Delta t_l = 0 \\ -1 & 0 < \Delta t_l < T \\ \text{其他} & \Delta t_l \geq T \end{cases}$$

式中: R_n 用于冗余性评价,表示第 n 个断面的冗余率; l 为不同的时空过车数据序列, T 为时间阈值上限,一般为 60 s, Δt_l 表示前 1 条采样时间序列 t_l 与后 1 条采样时间序列 t_{l-1} 的差分, $\Delta t_l = 0$ 表示前后 2 条数据序列信息完全相同,属于重复数据, $0 < \Delta t_l < T$ 表示相邻采样时间序列的差分处于合适的时间差阈值内,属于相似数据, $\Delta t_l \geq T$ 表示非冗余数据; $f_n(l)$ 表示第 n 个断面采集到的第 l 条重复或相似数据; N 为断面总数; L 为数据总数; $N_n^{R_1}$ 表示第 n 个断面采集到的原始数据总量; M_n 用于完整性评价,表示第 n 个断面的缺失率,缺失率越大表明数据缺失的越多,数据集完整性越差; N_n^m 为第 n 个断面实际未采集到的数据; $N_n^{M_1}$ 为按照一定时间间隔应采集到第 n 个断面的非冗余数据总量; V_n 用于有效性评价,表示第 n 个断面的失效率,失效率越大表明异常数据越多,数据有效性越差; N_n^i 为第 n 个断面通过 Boxplot 判别准则诊断出的异常数据; $N_n^{V_1}$ 为实际采集到的非冗余且非缺失数据总量; A_n 用于可用性评价,表示第 n 个断面的可用率,可用率越高表明路段的整体数据质量越好。

1.3 基于 EWM-TOPSIS 的 TBSS 数据质量综合评价模型构建

本文所提出的 EWM-TOPSIS 综合评价算法,首先以卡口监控设备断面位置构成的路段为中观层面,利用 EWM 法计算各评价指标信息熵权重,进一步采用 TOPSIS 评价算法评估路段 TBSS 数据质量,最后基于中观层面的评价结果及道路属性综合评估区域路网 TBSS 数据质量^[26-27]。

城市路段 TBSS 数据质量评价的关键步骤在于确定各评价指标权重,EWM 是一种强客观性且可再现的赋权方法,通过计算信息论中信息熵衡量某一指标的离散程度,指标的离散程度越大,其对综合评价结果的影响越大,对应的信息熵也越大。

步骤 1, EWM 定权过程,量纲一化原始数据矩阵 Z 得到 $Z' = (a'_{ij})_{p \times q}$

$$Z = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1q} \\ a_{21} & a_{22} & \cdots & a_{2q} \\ \cdots & \cdots & \cdots & \cdots \\ a_{p1} & a_{p2} & \cdots & a_{pq} \end{bmatrix} \quad (3)$$

$$a'_{ij} = \frac{a_{ij} - \min a_j}{\max a_j - \min a_j} \quad (4)$$

式中: p 为待评价路段数量; q 为评价指标数量; a_{pq} 表示路段 p 的第 q 个评价指标值; a'_{ij} 为数据量纲一化后路段 i 的第 j 个指标值, $i = 1, 2, \dots, p, j = 1, 2, \dots, q$; $\max a_j$ 及 $\min a_j$ 分别为指标 j 所在列的最大值及最小值。

每个指标的信息熵 H_j 表示为

$$H_j = -C \sum_{i=1}^p P_{ij} \ln P_{ij} \quad (5)$$

式中: C 为常数且 > 0 , 取 $C = \frac{1}{\ln p}$; $P_{ij} = \frac{a'_{ij}}{\sum_{i=1}^p a'_{ij}}$ 为

第 p 条待评价路段中第 j 个指标值 a'_{ij} 占总指标值 $\sum_{i=1}^p a'_{ij}$ 的特征比重。

每个指标的权重 w_j 表示为

$$w_j = \frac{1 - H_j}{\sum_{j=1}^q (1 - H_j)} \quad (6)$$

式中: $1 - H_j$ 为第 j 个指标的差异系数,差异系数反映了综合评价指标的作用大小,其值越大,指标的作用就越大。

步骤 2, EWM-TOPSIS 评价路段 TBSS 数据质量过程,首先基于 Z' 及权重 W 矩阵构造加权矩阵 F ,进而计算待评价路段与最优解 f^+ 、最劣解 f^- 之间的欧式距离 D_i^+ 及 D_i^- ,最后得出各路段的相对贴近度 C_i , C_i 越大则待评价路段的 TBSS 数据质量越好,即

$$F = Z' \cdot W = \begin{bmatrix} a'_{11} & a'_{12} & \cdots & a'_{1q} \\ a'_{21} & a'_{22} & \cdots & a'_{2q} \\ \vdots & \vdots & & \vdots \\ a'_{p1} & a'_{p2} & \cdots & a'_{pq} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \cdots \\ w_q \end{bmatrix}^T \quad (7)$$

$$f^+ = [\min_i (w_j a'_{ij}) | j \in n] = [f_1^+, f_2^+, \dots, f_q^+] \quad (8)$$

$$f^- = [\max_i (w_j a'_{ij}) | j \in n] = [f_1^-, f_2^-, \dots, f_q^-]$$

$$D_i^+ = \sqrt{\sum_{j=1}^q (w_j a'_{ij} - f_j^+)^2} \quad (9)$$

$$D_i^- = \sqrt{\sum_{j=1}^q (w_j a'_{ij} - f_j^-)^2} \quad (10)$$

$$C_i = \frac{D_i^-}{D_i^- + D_i^+} \quad (11)$$

步骤 3, 综合评价区域路网 TBSS 数据质量过程,基于路段层面的评价结果 C_i 及道路属性 A_i ,即第 i 条卡口检测路段长度 L_i 占完整道路长度 L_t 的比重和第 i 条路段的等级 B_i ,得到 TBSS 数据质量指数 T_D ,即

$$\begin{cases} A_i = \frac{L_i}{L_t} \times 100\% \\ B_i = 1, 2, 3, 4 \\ T_D = \sum_{i=1}^p (A_i B_i C_i) \end{cases} \quad (12)$$

一般城市路段等级按照快速路为 1、主干路为 2、次干路为 3 及支路为 4 划分, T_D 越大代表区域路网的质量越好。

2 实例验证分析

2.1 TBSS 数据判别分析

表 1 为乌鲁木齐市内由 30 条有向路段构成的某一区域路网空间范围,通过卡口监控设备采集的 2021 年 8 月 30 日~2021 年 9 月 5 日共 475 200 条原始过车数据记录。

表 1 部分卡口监控设备原始过车信息

Tab. 1 Original vehicle informations for part of bayonet monitoring equipment

路段编号	车牌号码	采样时间	车速/(km · h ⁻¹)
1	新 A**** Z	2021-08-30 00:01:32	10
1	新 A**** Z	2021-08-30 00:01:32	55
1	新 A**** 6	2021-08-30 00:01:40	10
1	新 A**** 6	2021-08-30 00:01:40	25
⋮	⋮	⋮	⋮
20	新 A**** 5	2021-09-01 00:00:00	12
20	新 A**** 3	2021-09-01 00:00:04	25
⋮	⋮	⋮	⋮
30	新 A**** 0	2021-09-05 00:00:21	41
30	新 A**** 3	2021-09-05 00:01:18	42

图 4 为路段 20 在工作日(8 月 30 日~9 月 3 日)及双休日(9 月 4 日~9 月 5 日)的地点车速经过问题数据判别算法筛选的结果。图 4(a)的方框为 9 月 1 日内 TBSS 缺失数据,5:00~8:00 期间由于车辆流量较小导致 TBSS 数据缺失程度较为严重,9:00~12:00 期间 TBSS 数据缺失程度适中,而图 4(b)中全天内没有 TBSS 缺失数据。图 4 中的突出点标记分别为连续 5 d 工作日及双休日的 TBSS 异常数据,可以看出 TBSS 异常数据现象普遍存在,且工作日中的 TBSS 异常数据量多于双休日中的 TBSS 异常数据量。

2.2 TBSS 数据质量诊断分析

图 5 为 30 条路段在 8 月 30 日~9 月 5 日的 TBSS 数据质量多属性诊断结果,可见,各路段 TBSS 数据的冗余率、缺失率、失效率及可用率波动走势大致相似,整体冗余性较低,完整性、有效性及

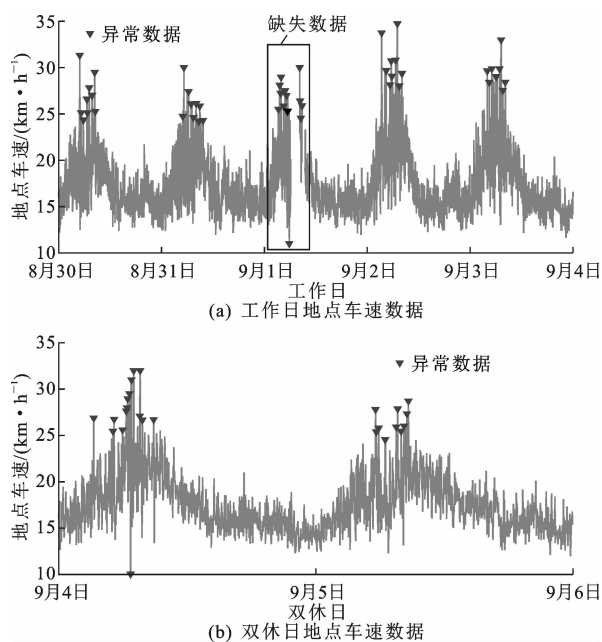


图 4 TBSS 问题数据判别结果

Fig. 4 Discrimination results of TBSS problem data

可用性较优。

可见,图 5(a)中路段的 TBSS 数据平均冗余率为 5.66%,一半以上路段的 TBSS 数据冗余率为 0%,路段 4、路段 12、路段 22、路段 14、路段 3 的 TBSS 数据冗余率皆高于其余路段,分别为 31.48%、29.02%、27.76%、20.47%、15.6%,这是由于受到周边学校和商圈影响,导致路段交通流过于密集,车辆在经过卡口断面时速度过慢或静止,检测到大量 TBSS 冗余数据。图 5(b)中路段的 TBSS 数据平均缺失率为 10.04%,路段 26、路段 27、路段 28 的 TBSS 数据缺失率皆高于其余路段,分别为 39.7%、32.22%、24.56%,这是由于非高峰时期的车辆较少,路段交通流过于疏散,卡口监控设备在 2 min 采样时间粒度下检测到稀疏的 TBSS 数据。图 5(c)中路段的 TBSS 数据平均失效率为 5.04%,路段 4、路段 5、路段 11、路段 17、路段 25、路段 29 的 TBSS 数据平均失效率皆高于其余路段,分别为 8.84%、9.65%、9.32%、9.08%、8.94%、8.9%,这是由于路段周边存在学校和商圈导致交通流过于密集,卡口监控设备检测到大量车辆数据,设备采集并存储数据时出现异常现象。图 5(d)中路段的 TBSS 数据平均可用率为 84.66%,路段 4、路段 12、路段 14、路段 26 的 TBSS 数据可用率皆低于其余路段,分别为 61.4%、64.76%、69.24%、61.33%,这是由于各路段布设的卡口设备冗余性、完整性及有效性在不同程度上存在差异,导致 TBSS 数据质量的综合

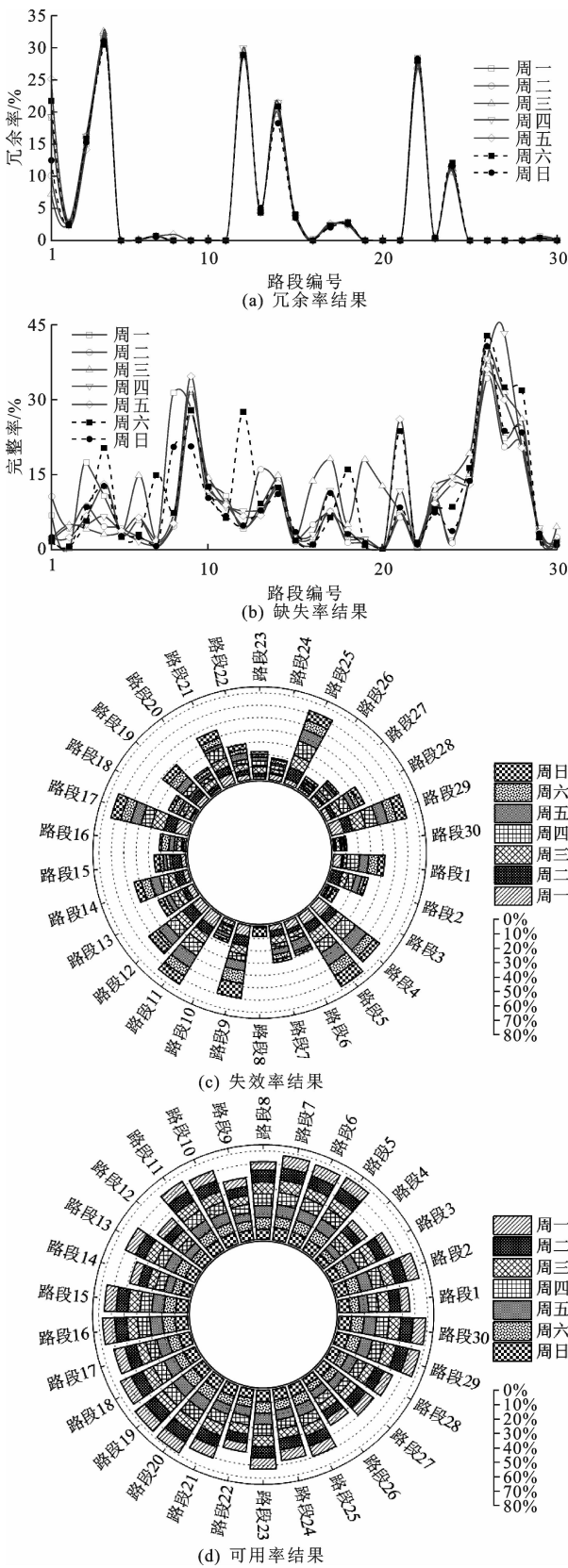


图 5 路段 TBSS 数据质量多属性评价指标诊断结果

Fig. 5 Diagnosis results of multi-attribute evaluation indexes for TBSS data quality on road segments

可用率较差。基于路段空间层次的 TBSS 数据质量

评价结果和道路属性信息综合评估,8 月 30 日~9 月 5 日内区域路网 TBSS 数据质量的结果如图 6 所示。可见:

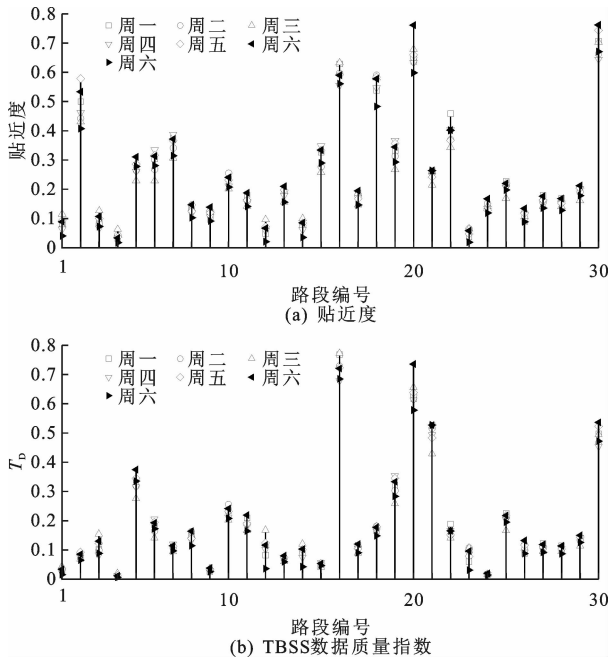


图 6 区域路网 TBSS 数据质量综合诊断结果

Fig. 6 Comprehensive diagnosis results of TBSS data quality on regional road network

图 6(a)中 30 条路段的平均贴近度为 0.25,各路段 TBSS 数据质量整体良好。其中路段 30 和路段 20 在 8 月 30 日~9 月 5 日的贴近度整体保持最优,分别为 0.7 和 0.66,尤其在 9 月 4 日,2 条路段的贴近度最高,皆为 0.76。路段 4 和路段 23 在 8 月 30 日~9 月 5 日的贴近度整体最差,分别为 0.037 和 0.048。

图 6(b)中区域路网在 8 月 30 日~9 月 5 日的平均 T_D 为 0.19,区域路网 TBSS 数据质量整体较优。其中路段 4 和路段 24 的平均 T_D 最低,皆为 0.01,路段 1 和路段 9 的平均 T_D 次之。图 6(a)中路段 24 和路段 9 的贴近度高于路段 4 和路段 1,皆为 0.03,路段 4、路段 24、路段 1 和路段 9 的道路等级均为 2,但由于路段 24 和路段 9 的卡口检测路段长度比重分别为 0.067 以及 0.136,贴近度较优的 2 条同等级路段受卡口检测路段长度比重的影响使得其平均 T_D 降低。

进一步观察图 6(b)可以看出,路段 16、路段 20、路段 21 及路段 30 的平均 T_D 优于其他路段,分别为 0.72、0.63、0.5、0.49,道路等级分别为 2、4、2 和 3,相反,对应的卡口检测路段长度比重分别为 1、0.61、0.24 和 0.23,路段同时受卡口检测路段长度

比重和道路等级影响,使得贴中度较优路段的平均 T_D 低于贴中度相对较低路段。综合图 6 可见,道路属性在一定程度上会联动影响区域路网的 TBSS 数据质量。

3 结 语

(1)本文构建了冗余及缺失数据判别算法和 Boxplot 异常数据判别准则,对原始卡口交通数据进行精确判别,最终得到了 3 种问题数据类型,即冗余数据、缺失数据、异常数据。

(2)针对 3 种问题数据类型特点,本文从冗余性、完整性、有效性及可用性角度设计了一种多属性综合评价指标图谱。该图谱明确了指标之间的相互联系,表明了所选取评价指标的合理性与科学性。

(3)为了综合评价区域路网 TBSS 数据质量状况,本文利用 EWM 赋权法为各评价指标定权后,基于 TOPSIS 评价法评估了路段 TBSS 数据质量,并结合道路属性综合评价了区域路网 TBSS 数据质量,为城市卡口交通数据质量恢复及应用发展提供一定的数据支持。

(4)以乌鲁木齐市某一区域路网 1 周的卡口交通数据为例,应用本文提出的方法对原始卡口交通数据进行明确判别,对路段及区域路网 TBSS 数据质量综合评价进行验证。获得了各路段 TBSS 数据质量状况评价结果,同时发现道路属性在一定程度上影响了区域路网 TBSS 数据质量状况,这为后续的城市道路改扩建提供一定的思路。

(5)本文仅针对卡口地点车速数据质量进行诊断,对于多个交通参数的数据质量融合性诊断需要进一步拓展研究,未来会考虑多类型交通监控设备数据质量的综合诊断研究。

参考文献:

References:

- [1] 张晓娟,谭 婧.我国省级政府数据开放平台元数据质量评估研究[J].电子政务,2019(3):58-71.
ZHANG Xiao-juan,TAN Jing. Research on metadata quality evaluation of provincial government data open platform in China[J]. E-Government,2019(3):58-71.
- [2] 白永昕,闫懋博,田茂再,等.基于 Bootstrap 方法的多维统计数据质量评估[J].统计与决策,2020,36(11):5-9.
BAI Yong-xin,YAN Mao-bo,TIAN Mao-zai, et al. Evaluation on the quality of multidimensional statistics based on Bootstrap method[J]. Statistics & Decision,2020,36(11):5-9.
- [3] D'AMBROSIO A,GARLASCO J,QUATTROCOLO F, et al. Data quality assessment and subsampling strategies to correct distributional bias in prevalence studies[J]. BMC Medical Research Methodology, 2021, 21 (1): 90-104.
- [4] TUTE E,SCHEFFNER I,MARSHOLLEK M. A method for interoperable knowledge-based data quality assessment[J]. BMC Medical Informatics and Decision Making,2021,21(1):1-14.
- [5] LI J,ZUYLEN H J,WEI G. Loop detector data error diagnosing and interpolating with probe vehicle data [J]. Transportation Research Record, 2014 (2423): 61-67.
- [6] WANG X X,XU L H,CHEN K X. Data-driven short-term forecasting for urban road network traffic based on data processing and LSTM-RNN [J]. Arabian Journal for Science and Engineering, 2019, 44 (4): 3043-3060.
- [7] MAGHROUR Z M,TOROK A. Single loop detector data validation and imputation of missing data[J]. Measurement,2018,116:193-198.
- [8] 谢聪聪,邵敏华.城市快速路交通采集数据完整性与有效性评价[J].交通信息与安全,2018,36(4):60-65,89.
XIE Cong-cong,SHAO Min-hua. An evaluation on integrity and validity of traffic data collected from urban expressways[J]. Journal of Transport Information and Safety,2018,36(4):60-65,89.
- [9] 陆百川,张冬梅,舒 芹,等.基于时空特性和灰色残差的交通故障数据诊断与修复[J].重庆交通大学学报(自然科学版),2020,39(9):8-16.
LU Bai-chuan,ZHANG Dong-mei,SHU Qin, et al. Diagnosis and repair of traffic fault data based on spatio-temporal characteristics and grey residual [J]. Journal of Chongqing Jiaotong University (Natural Science),2020,39(9):8-16.
- [10] AHN K,RAKHA H,HILL D. Data quality white paper[R]. Washington DC:Department of Transportation,2008.
- [11] PARK H,YOON M,KIM H,et al. Development of a novel integrated evaluation index for freeway traffic data[J]. Journal of Korean Society of Transportation, 2015,33(4):417-429.
- [12] LI Y D,CHEN D W. A learning-based comprehensive evaluation model for traffic data quality in intelligent transportation systems[J]. Multimedia Tools and Applications,2016,75(19):11683-11698.

- [13] 庄广新,谷远利,马韵楠,等.基于AHP灰色理论的道路交通流数据质量评价[J].山东科学,2017,30(1):69-75.
ZHUANG Guang-xin, GU Yuan-li, MA Yun-nan, et al. Road traffic flow data quality assessment based on grey theory and analysis hierarchical process[J]. Shandong Science, 2017, 30(1): 69-75.
- [14] BAE J, CHOI K. A new perspective for precision evaluation of large-scale traffic sensor data measurement[J]. Transportation Planning and Technology, 2020, 43(6): 571-585.
- [15] 杨越思,杜威,宁丹,等.RFID交通冗余数据检测及分析[J].交通信息与安全,2016,34(3):72-80.
YANG Yue-si, DU Wei, NING Dan, et al. Detection and an analysis of redundant RFID traffic data[J]. Journal of Transport Information and Safety, 2016, 34(3): 72-80.
- [16] 杨越思,杜威,宁丹,等.RFID交通数据时间覆盖率分析[J].交通信息与安全,2017,35(3):43-50,58.
YANG Yue-si, DU Wei, NING Dan, et al. An analysis on time coverage rate of RFID traffic data[J]. Journal of Transport Information and Safety, 2017, 35(3): 43-50, 58.
- [17] 龙思卿,赵盼明,郭建华.基于IOWA算子的RFID数据有效性评价方法[J].南京信息工程大学学报(自然科学版),2019,11(4):474-482.
LONG Si-qing, ZHAO Pan-ming, GUO Jian-hua. An IOWA operator based RFID data validity evaluation method[J]. Journal of Nanjing University of Information Science & Technology (Natural Science Edition), 2019, 11(4): 474-482.
- [18] 赵盼明.典型道路交通数据有效性评价方法[D].南京:东南大学,2019.
ZHAO Pan-ming. The effectiveness evaluation method of typical road traffic data[D]. Nanjing: Southeast University, 2019.
- [19] 付凤杰.基于自动车牌识别数据的城市道路行程时间估计[D].杭州:浙江大学,2017.
FU Feng-jie. Travel time estimation for urban road networks using ANPR data[D]. Hangzhou: Zhejiang University, 2017.
- [20] 孙猛,李建梅,孙锋,等.基于卡口数据的交通参数提取方法研究[J].交通信息与安全,2020,38(6):137-144.
SUN Meng, LI Jian-mei, SUN Feng, et al. A traffic parameter extraction method based on checkpoint data[J]. Journal of Transport Information and Safety, 2020, 38(6): 137-144.
- [21] LIU C J, WANG Q P, CAO Z. An intelligent optimization method for highway route selection based on comprehensive weight and TOPSIS[J]. PLoS One, 2022, 17(2): e0262588.
- [22] 杜威.智能交通数据质量控制关键技术研究[D].南京:东南大学,2017.
DU Wei. Research on the key quality control technology of intelligent transportation system data[D]. Nanjing: Southeast University, 2017.
- [23] 邵毅明,钟颖,吴文文,等.基于熵权TOPSIS法的短时交通流预测模型性能综合评价[J].重庆理工大学学报(自然科学),2020,34(7):205-211,254.
SHAO Yi-ming, ZHONG Ying, WU Wen-wen, et al. Comprehensive evaluation of short-term traffic flow prediction model based on entropy TOPSIS model[J]. Journal of Chongqing University of Technology (Natural Science), 2020, 34(7): 205-211, 254.
- [24] 徐程,曲昭伟,陶鹏飞,等.动态交通数据异常值的实时筛选与恢复方法[J].哈尔滨工程大学学报,2016,37(2):211-217.
XU Cheng, QU Zhao-wei, TAO Peng-fei, et al. Methods of real-time screening and reconstruction for dynamic traffic abnormal data[J]. Journal of Harbin Engineering University, 2016, 37(2): 211-217.
- [25] 张振慧,张正江,胡桂廷,等.基于拉依达准则与线性拟合的改进型无迹卡尔曼滤波粗大误差补偿算法[J].计算机测量与控制,2019,27(11):153-156,162.
ZHANG Zhen-hui, ZHANG Zheng-jiang, HU Gui-ting, et al. Improved unscented Kalman filter based on gross error compensation algorithm with Pauta criterion and linear fitting[J]. Computer Measurement & Control, 2019, 27(11): 153-156, 162.
- [26] 马静,戴维佳.基于AHP-信息熵决策的城市轨道交通建设期交通组织方案评价[J].长安大学学报(自然科学版),2020,40(4):101-108.
MA Jing, DAI Wei-jia. Evaluating on traffic management during traffic construction of urban rail based on AHP-information entropy[J]. Journal of Chang'an University (Natural Science Edition), 2020, 40(4): 101-108.
- [27] 魏勇,李建文,郭亮亮,等.基于TOPSIS的GNSS数据质量评估方法研究[J].大地测量与地球动力学,2016,36(10):892-896.
WEI Yong, LI Jian-wen, GUO Liang-liang, et al. Research on GNSS data quality evaluation based on TOPSIS[J]. Journal of Geodesy and Geodynamics, 2016, 36(10): 892-896.