

doi: 10.3969/j.issn.1672-6073.2024.04.007

基于奇异值分解的城轨网络 关键站点识别及其演变研究

王一哲^{1,2}, 邱昱植¹, 王子甲¹

(1. 北京交通大学土木建筑工程学院, 北京 100044;

2. 北京交通大学土木工程国家级实验教学示范中心, 北京 100044)

摘要: 为识别城市轨道交通网络关键站点并研究其多年演化, 构建基于截断奇异值分解(truncated singular value decomposition, TSVD)的关键站点识别方法, 选取北京市 2011—2019 年早高峰时段的 OD 数据, 通过关键特征向量分析网络客流演变并对城轨网络中关键站点进行识别, 将其与复杂网络方法的识别结果进行对比。分析表明: TSVD 法能很好地应用于考虑 OD 分布的网络关键站点识别, 识别结果能更好代表网络客流的空分布。从识别结果看, 北京轨道交通关键站点空间布局呈现多中心发展趋势, 如西北西二旗, 西南丰台科技园等站点逐步形成网络客流中心并相互联系; 东南土桥、东北俸伯等站点也初步呈现网络客流中心的特征。

关键词: 城市轨道交通; 截断奇异值分解; OD 矩阵; 关键站点; 关键特征向量

中图分类号: U231

文献标志码: A

文章编号: 1672-6073(2024)04-0046-06

Identification and Evolution of Key Stations in Urban Rail Networks Using Singular Value Decomposition

WANG Yizhe^{1,2}, QIU Yuzhi¹, WANG Zijia¹

(1. School of Civil Engineering, Beijing Jiaotong University, Beijing 100044;

2. Experimental Center of Civil Engineering, Beijing Jiaotong University, Beijing 100044)

Abstract: To identify key stations in urban rail transit networks and study their evolution over multiple years, we developed a key station recognition method based on truncated singular value decomposition (TSVD). We selected origin-destination (OD) data from the morning peak hours between 2011 and 2019 in Beijing as the dataset. The analysis involved evaluating the evolution of network passenger flow and identifying key stations in subway networks using key eigenvectors. These results were then compared with those obtained from complex network methods. The analysis demonstrates that TSVD can effectively identify key network stations by considering the OD distribution. The outcomes from TSVD better represent the spatial distribution of network passenger flow than traditional methods. The results revealed that the spatial layout of key stations in Beijing's urban rail transit has evolved towards multiple centers, such as Northwest Xi'erqi and Southwest Fengtai Science Park. These stations are gradually forming network passenger flow centers and establishing connections. Additionally, stations like Southeast Tu Qiao and Northeast Fengbo are showing preliminary trends toward becoming network passenger flow centers.

Keywords: urban rail transit; truncated singular value decomposition; OD matrix; key stations; key eigenvectors

收稿日期: 2023-09-18 修回日期: 2024-04-03

第一作者: 王一哲, 男, 本科, 主要研究方向为城市轨道交通大数据应用, 20251049@bjtu.edu.cn

通信作者: 王子甲, 男, 教授, 主要从事城市轨道交通线网规划研究, zjwang@bjtu.edu.cn

基金项目: 国家级大学生创新创业训练计划资助项目(202410004187)

引用格式: 王一哲, 邱昱植, 王子甲. 基于奇异值分解的城轨网络关键站点识别及其演变研究[J]. 都市轨道交通, 2024, 37(4): 46-51.

WANG Yizhe, QIU Yuzhi, WANG Zijia. Identification and evolution of key stations in urban rail networks using singular value decomposition[J]. Urban rapid rail transit, 2024, 37(4): 46-51.

1 研究背景

城市轨道交通刷卡数据包含了乘客出行时间、起讫点(OD)等丰富的信息,是重要的交通数据资源。城市轨道交通网络中关键站点对网络效率及平稳运行具有显著影响,在轨道交通线网中具有重要地位^[1]。随着网络快速发展,关键站点也在发生动态变化;长期的刷卡数据能够反映关键站点分布及其演变趋势。

以往识别关键站点的研究主要从网络拓本身出发,然而这些研究大多未能充分考虑客流需求的空间分布。焦柳丹等^[2]将客流中心性作为权重加入拓网络,表征某个站点在整个轨道交通网络中聚集客流的能力,但这种单一权重并不能准确反映客流的大尺度空间分布内在联系。现有研究中尚缺乏基于长期 OD 数据进行网络演化研究的内容。

城市轨道交通网络的关键站点识别方法大致可以归纳为两类:第一类基于复杂网络理论^[3],依据网络拓扑特性识别关键站点,这种方法虽然所需数据量较小,但难以反映需求侧的信息;第二类从交通网络运行性能角度出发,考虑实际客流量、用户出行时间等因素,但需要大量统计数据支撑,且计算复杂度较高。

奇异值分解(singular value decomposition, SVD)作为一种重要的线性代数方法,在图像压缩、推荐系统、信号处理等领域得到广泛应用。而图像中像素和图(Graph)中的节点都具有代表其属性的特征值,且元素间具有某种特定的邻接关系。因此,可以借鉴 SVD 处理图像数据的思路,将复杂的 OD 矩阵分解转化为更易于处理和解释的低维空间,进而捕捉 OD 数据中重要的客流模式和规律,挖掘轨道交通网络发展的主要趋势和关键站点的动态变化。

识别城市轨道交通网络中的关键站点,有助于服务乘客需求和优化网络运营管理,对合理分配交通资源、有效应对紧急情况以及制定长远的网络规划具有一定意义。本文通过引入奇异值分解的基本原理,构造 OD 矩阵的奇异值分解方法并应用到多年的工作日早高峰 OD 矩阵的数据处理中,研究北京轨道交通网络中关键站点空间分布及其演化进程。

2 SVD 奇异值分解法的构造

2.1 SVD 法原理

假设实矩阵 $A \in R^{m \times n}$, 则必存在两正交矩阵

$$U = [u_1, \dots, u_m] \in R^{m \times m} \quad (1)$$

$$V = [v_1, \dots, v_n] \in R^{n \times n} \quad (2)$$

使得

$$A = U \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} V^T \quad (3)$$

$$\Sigma_r = \text{diag}(\sigma_1, \dots, \sigma_r), \quad 0 < \sigma_r \leq \dots \leq \sigma_1 \quad (4)$$

式中,正交矩阵 U 和 V 分别称为矩阵 A 的左、右奇异矩阵; V^T 为右奇异矩阵的转置; Σ_r 为分解产生的 r 个非零奇异值序列组成的对角阵,其中 r 为矩阵 A 的秩,对角阵中最大的奇异值元素代表矩阵 A 最显著的数量特征; σ_j 为矩阵 A 从大到小排列的第 j 个奇异值。式(3)则为实矩阵 A 的奇异值分解。

SVD 法通过对矩阵 U 与 V 第 r 列之后的列向量进行删除操作,实现以低维矩阵的积的形式表示目标矩阵 A , 进而实现数据存储层面的无损失降维。

2.2 基于截断 SVD 的关键站点识别方法

具体地,令包含 n 个站点的原始 OD 矩阵($n \times n$ 维)作为输入矩阵 A , 通过 SVD 法将 OD 矩阵分解为 2.1 节 3 个矩阵的积的形式。在此分解中,矩阵 U 中各行向量为起点站(O)进站客流的低维向量;矩阵 V^T 中各列向量为终点站(D)出站客流的低维向量。两矩阵共同捕捉了客流数据的核心信息和结构特征。

矩阵 Σ_r 对角线上奇异值的大小衡量了 OD 客流水平的高低。根据 OD 矩阵为方阵的构造特征,定义 OD 矩阵表中第 i 个站点($0 < i \leq n$)的关键特征向量 s 为

$$s = \frac{u_{\text{row}} + v_{\text{col}}}{2} \quad (5)$$

式中, $u_{\text{row}} = U[i, :]$, $v_{\text{col}} = V[i, :]$, 分别表征轨道交通站点 i 在网络中作为 O 端与 D 端的客流特征。关键特征向量综合反映了单个站点对全网客流量的总体贡献水平。

同时,由 2.1 节可知, SVD 产生的尾端奇异值较小,因此,通过保留较大的 k 项奇异值和相应的特征向量($0 < k < r = \text{rank}(A)$), 可以进一步提取宏观客流分布中的关键特征,过滤较小、较分散客流量的影响,实现 OD 矩阵的“截断”降维^[4]。在此过程中,定义 OD 矩阵的截断 SVD 法如下:

$$A = U_k \begin{bmatrix} \Sigma_k & 0 \\ 0 & 0 \end{bmatrix} V_k^T \quad (6)$$

式(6)为矩阵 A 的截断奇异值分解(TSVD),其中 k 的大小通过奇异值元素的累积水平确定。

3 线网构建与数据预处理

3.1 城轨线网构建

北京轨道交通建设始于 1965 年,1971—1999 年间,先后建成了 1 号线、2 号线等 4 期工程,总里程突破 50 km。2000—2010 年,城轨建设进入快速发展期,先后建成了 10 号线、5 号线等 5 条线路,总里程达到 200 km。2011 年以来,线网建设进入高峰期,先后建成了 6 号线、9 号线、房山线等线路,总里程突破 700 km,车站总数突破 400 座^[5-6]。以 2018 年为例,北京轨道交通运营线路总长达到 636.8 km;运营车站达 391 座,工作日客运量达到 1 200 万人次/d。

本研究利用地图 API 获取站点空间位置信息以及站间距离数据,使用 Python3 中的 networkx 包构建了北京城市轨道交通网络的有向图(digraph)模型,作为网络参数计算的底图基础。

图 1 展示了北京城市轨道交通 2011—2019 年的线路建设情况,截至 2019 年底,线路开通运营总数达到了 23 条^[7]。

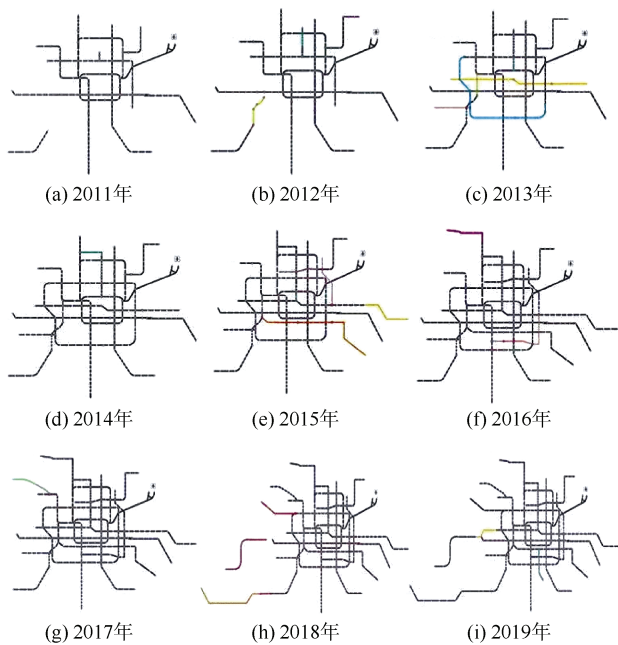


图 1 2011—2019 年北京城市轨道交通网络演化
Figure 1 Evolution of Beijing urban rail transit network from 2011 to 2019

3.2 OD 矩阵数据

城市轨道交通 OD 客流矩阵记录了一定时间段内从特定起点站到特定终点站的乘客数量。

本文使用的 OD 矩阵数据表通过北京轨道交通 2011—2019 年同期某月工作日早、晚高峰 1 h 的进出

站 AFC 系统刷卡数据转换得到。

为研究多年的客流演变,本文选择规律性较强的早高峰 OD 客流为研究对象,同时对比历年同期北京轨道交通工作日早高峰客流的空间分布差异^[8]。

以 2019 年某日为例,如表 1 所示,早高峰 OD 流量中不高于 4 的 OD 流占比约为 65%,该类较小的 OD 流大多系随机产生,规律性较低,通过截断 SVD 法能够对该类 OD 流实现有效过滤。

表 1 早高峰 OD 流统计
Table 1 Morning peak hour OD flow

OD 流	07:00—08:00
0	49.23%
(0, 4]	25.85%
(4, 8]	8.71%
(8, 12]	4.47%
(12, +∞]	11.74%

3.3 多年轨网 OD 的截断 SVD 分解

为识别早高峰背景下的 OD 需求与客流结构特征,首先通过 SVD 法对多年早高峰 OD 客流矩阵进行分解降维。如 2.2 节所述, SVD 法得到的 OD 矩阵的秩 r 代表了非零奇异值数量的同时,也反映了轨道交通网络建设规模特征。

统计历年 OD 矩阵的秩 r 的平均水平,如图 2 所示。可以观察到,2011 至 2019 年间北京轨道交通网络建设进入高峰期,站点总数的不断增加导致 OD 矩阵维度迅速增大;随着轨道交通网络布局日益复杂,考虑实际客流量与用户出行时间等因素的数据规模及计算复杂度表现出超线性增长趋势。

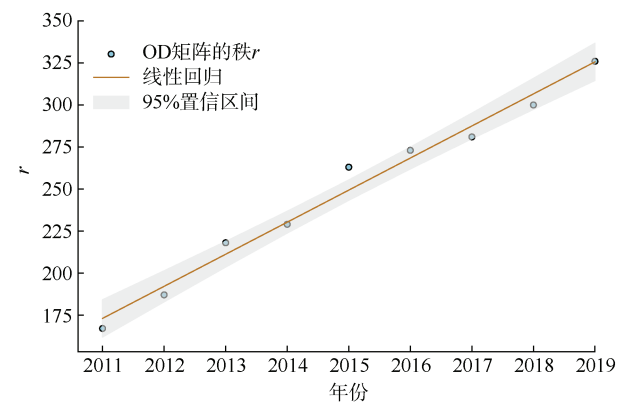


图 2 2011—2019 年秩 r 的平均水平
Figure 2 Average level of r value from 2011 to 2019

在城市轨道交通的时空 OD 矩阵中,奇异值充当了评估不同需求模式重要程度的量化指标,较大的奇

异值表明相应需求模式包含了更丰富的 OD 信息。为选择合适的截断 k 值, 需要进一步确定 SVD 法产生的奇异值的累积水平。

统计标准化处理后的奇异值分布情况, 如图 3 所示, 可以观察到, 最大的奇异值元素明显超过其他值, 后迅速减小至较低水平。以 2011 年为例, 取最大的 20 项奇异值即可表征 70% 的原始 OD 信息。

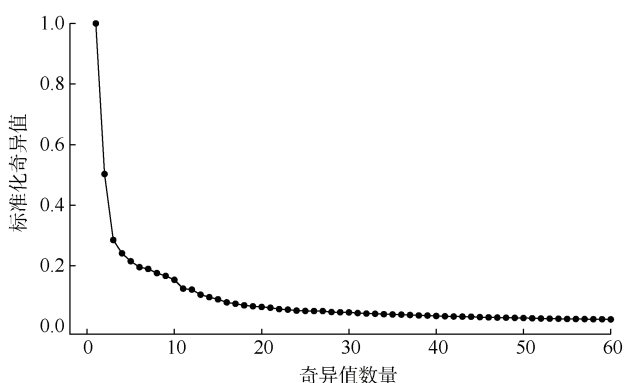


图 3 北京轨道交通 OD 矩阵标准化奇异值分布
Figure 3 Distribution of normalized singular values of OD matrix of Beijing urban rail transit

如图 4 所示, 在 OD 矩阵的秩 r 平均以 19/年的速度增长的情况下(图 2), 保持 70% 以上水平描述 OD 矩阵特征的截断 k 值平均增速约为 1/年。

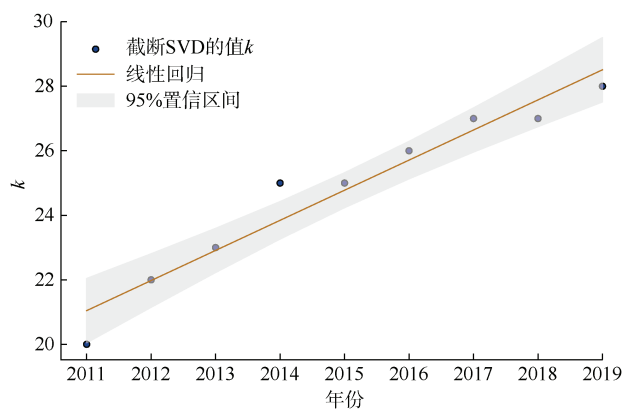


图 4 2011—2019 历年 k 值平均水平
Figure 4 Average level of k value from 2011 to 2019

截断分解结果侧面反映出, 尽管 OD 矩阵的维度增长迅速, 但主要 OD 信息集中在小部分站点且数量稳定。随着轨道交通网络布局逐渐向低需求区域扩张, 网络中较少的关键站点服务了较大比例的客流需求。

4 关键站点识别结果与方法对比

4.1 截断 SVD 法识别结果

在 SVD 法分解产生的数百个非零奇异值中, 通过

截断并保留最大的 k 项奇异值(占 $1/12 \sim 1/8$)即可捕获绝大部分 OD 客流信息, 矩阵的维度从 $O(n^2)$ 压缩至 $O(n)$, 显著降低了直接分析原始 OD 矩阵的复杂度。

针对 SVD 法得到的 OD 矩阵中每个站点对应的关键特征向量, 保留通过截断 SVD 法得到的前 k 个元素, 如 3.3 节所述, 以 2011 与 2019 年为例, 分别保留了关键特征向量中前 20、前 28 个元素。

以截断分解得到的各站点对应的关键特征向量及奇异值累积水平为依据, 识别具有工作日早高峰属性的关键站点。2012、2015 及 2018 年分别通过截断 SVD 法识别得到的关键站点的空间分布如图 5 所示。

对北京轨道交通全网所有站点的关键特征向量分布进行降维可视化^[9], 如图 6 所示。虽然降维处理后的分布无法完整表征特征向量的全体元素, 但能够基本反映出关键特征向量的元素数值大多均匀分布在原点附近的特点。这样的分布特点良好地避免了关键特征向量中单一元素值过大(极端客流)对站点客流水平评价的负面影响, 这是 SVD 法可行的重要支撑。

4.2 识别结果评估与对比

基于复杂网络理论, 引入介数中心性、接近中心性、特征向量中心性三项中心性指标以及 PageRank 法计算的节点重要性(佩奇排名值)共同评价站点关键性水平^[10]。

通过 Critic 客观赋权法对 4 项指标进行权重分配, 以其耦合权值作为复杂网络方法评价站点关键性的理论依据, 识别历年线网中的关键站点。

以 2017 年为例, 北京轨道交通网络中复杂网络指标耦合权值排名最高的 5 个关键站点及其高峰小时进出站的平均客流量水平如表 2 所示, 可以看到, 识别结果均为网络中双线/多线换乘站。

同时, 通过截断 SVD 法识别得到的 2017 年关键站点中, 工作日早高峰小时平均客流量排名在前 5 位的站点及其平均客流量如表 3 所示。

传统的复杂网络方法侧重于网络拓扑结构分析, 强调站点在网络中的结构位置和连接重要性, 在理解线网初期规划方面有一定意义, 然而该方法无法充分捕捉客流时空分布特征。

相比之下, TSVD 法更注重从客流数据中提取信息, 能更准确地反映客流需求, 在应对实际客流模式优化和现实运营挑战方面更具优势。

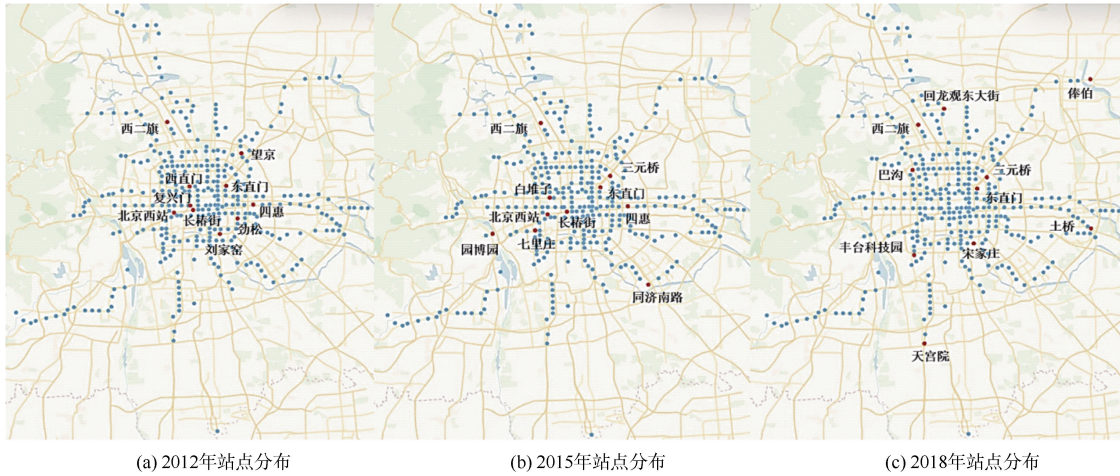


图 5 工作日关键站点分布

Figure 5 Distribution of key stations on weekdays

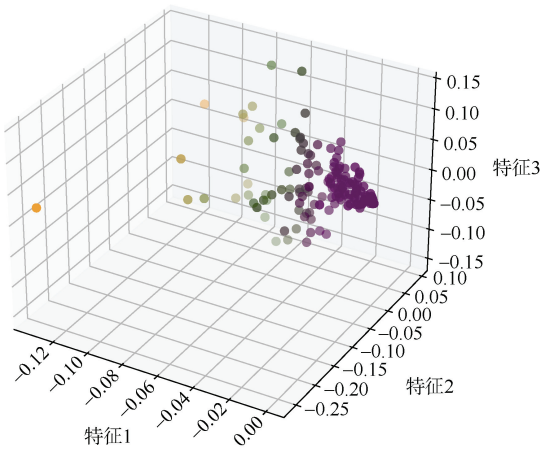


图 6 关键特征向量分布

Figure 6 Distribution of key eigenvectors

表 2 2017 年前 5 名站点指标

Table 2 Metrics of top 5 stations in 2017

车站名称	佩奇排名值	介数中心性	接近中心性	特征向量中心性	Critic 权值	高峰小时客流量/(人次/h)
西直门	0.005 90	0.250 28	0.095 48	0.181 55	0.914 61	15 350
朝阳门	0.004 71	0.170 39	0.096 83	0.263 09	0.831 19	18 329
建国门	0.004 68	0.113 26	0.095 13	0.291 41	0.799 55	7 891
车公庄	0.004 72	0.192 25	0.097 55	0.175 29	0.791 01	7 934
东四	0.004 82	0.135 12	0.097 52	0.222 42	0.790 72	5 408

表 3 2017 年前 5 名站点客流-TSVD

Table 3 Top 5 stations' passenger flow-TSVD

复杂网络法	客流量/(人次/h)	TSVD 法	客流量/(人次/h)
朝阳门	18 329	西二旗	21 878
西直门	15 350	东直门	14 687
车公庄	7 934	北京西站	11 023
建国门	7 891	回龙观东大街	9 864
东四	5 408	四惠	8 466

5 结语

1) 本文基于 SVD 原理, 构造了基于 TSVD 的关键站点识别方法, 通过实例应用到考虑 OD 分布的北京轨道交通关键站点识别研究; 识别结果在一定程度上更好反映了网络客流真实的空间分布。

2) 研究结果显示, 北京轨道交通网络快速扩张过程中, 关键站点数量整体呈现先增加后逐步稳定的规律, 表明线网发展到一定程度, 供需两方面共同反映的网络核心逐步稳定。与此同时, 关键站点空间布局逐渐呈现多中心发展趋势, 如西北西二旗、西南丰台科技园等站逐步形成早高峰网络客流中心并相互联系; 东南土桥、东北俸伯等站点也初步出现网络客流中心特征。

3) 考虑到 OD 矩阵数据的高维特性, 降维处理具有必要性。在本研究中, 少量被保留的奇异值(约 10%)能够有效捕捉到原矩阵约 70% 的 OD 分布信息, 凸显了 TSVD 法在数据压缩和信息提取方面的优势。

参考文献

[1] 樊明浩, 洪于亮, 戴冀峰, 等. 复杂网络下 TOD 耦合机制引导城市发展策略研究: 以北京为例[J]. 都市轨道交通, 2022, 35(4): 74-80.
FAN Minghao, HONG Yuliang, DAI Jifeng, et al. Urban development strategy guided by TOD coupling mechanism under complex network: a case study of Beijing[J]. Urban rapid rail transit, 2022, 35(4): 74-80.

[2] 焦柳丹, 罗敏, 谌微微, 等. 基于复杂网络和 VIKOR 的城市轨道交通网络站点重要性研究[J]. 重庆交通大学学报(自然科学版), 2022, 41(10): 16-25.
JIAO Liudan, LUO Min, CHEN Weiwei, et al. Station importance of urban rail transit network based on complex

- network and VIKOR[J]. Journal of Chongqing Jiaotong University (natural science), 2022, 41(10): 16-25.
- [3] 路庆昌, 崔欣, 谢驰, 等. 城市轨道交通网络关键站点识别方法对比与分析[J]. 北京交通大学学报, 2022, 46(3): 18-25.
LU Qingchang, CUI Xin, XIE Chi, et al. Comparison and analysis of identification methods for critical stations in urban rail transit networks[J]. Journal of Beijing Jiaotong University, 2022, 46(3): 18-25.
- [4] 何凌晖, 余庆, 李玮峰, 等. 新冠肺炎疫情影响下的城际交通运输需求分析[J]. 城市交通, 2020, 18(3): 51-61.
HE Linghui, YU Qing, LI Weifeng, et al. Inter-city transportation demand under the COVID-19 pandemic[J]. Urban transport of China, 2020, 18(3): 51-61.
- [5] 侯秀芳, 冯晨, 左超, 等. 2022年中国内地城市轨道交通线路概况[J]. 都市轨道交通, 2023, 36(1): 9-13.
HOU Xiufang, FENG Chen, ZUO Chao, et al. Statistical analysis of urban rail transit in Chinese mainland in 2022[J]. Urban rapid rail transit, 2023, 36(1): 9-13.
- [6] 中国城市轨道交通协会. 城市轨道交通 2022 年度统计和分析报告[J]. 城市轨道交通, 2023(4): 13-15.
China Association of Metros. Annual statistics and analysis report of urban rail transit in 2022[J]. China metros, 2023(4): 13-15.
- [7] 韩宝明, 习喆, 孙亚洁, 等. 2022 年世界城市轨道交通运营统计与分析综述[J]. 都市轨道交通, 2023, 36(1): 1-8.
HAN Baoming, XI Zhe, SUN Yajie, et al. Statistical analysis of urban rail transit operation in the world in 2022: a review[J]. Urban rapid rail transit, 2023, 36(1): 1-8.
- [8] 杨志刚, 贾慧慧, 陈峰, 等. 基于长期客流数据的北京城市轨道交通客流演变分析[J]. 城市轨道交通研究, 2022, 25(9): 126-131.
YANG Zhigang, JIA Huihui, CHEN Feng, et al. Passenger flow evolution analysis of Beijing urban rail transit network based on long-term passenger flow data[J]. Urban mass transit, 2022, 25(9): 126-131.
- [9] 王晨, 石俊刚, 席苏路, 等. 基于 AFC 数据的南昌轨道交通车站精细化分类[J]. 都市轨道交通, 2023, 36(6): 49-56.
WANG Chen, SHI Jungang, XI Sulu, et al. Accurate classification of Nanchang urban rail transit stations using AFC data[J]. Urban rapid rail transit, 2023, 36(6): 49-56.
- [10] WANG Zijia, LI Jingqi, HUANG Liping, et al. Discovering the evolution of Beijing Rail Network in fifty years[J]. Modern physics letters B, 2020, 34(21): 2050212.

(编辑: 傅依萱)

(上接第 37 页)

参考文献

- [1] 侯秀芳, 冯晨, 燕汉民, 等. 2023 年中国内地城市轨道交通运营线路概况[J]. 都市轨道交通, 2024, 37(1): 10-16.
HOU Xiufang, FENG Chen, YAN Hanmin, et al. Overview of urban rail transit operational lines in Chinese mainland in 2023[J]. Urban rapid rail transit, 2024, 37(1): 10-16.
- [2] 住房和城乡建设部城市交通基础设施监测与治理实验室, 中国城市规划设计研究院. 2022 年度中国主要城市通勤监测报告[R]. 北京, 2022.
- [3] 王新如, 常利, 裴斐, 等. 北方城市某地铁站内颗粒物浓度分布实测和分析[J]. 暖通空调, 2023, 53(7): 35-39.
WANG Xinru, CHANG Li, PEI Fei, et al. Measurement and analysis of particle concentration distribution in a subway station in northern cities[J]. Heating ventilating & air conditioning, 2023, 53(7): 35-39.
- [4] ANALITIS A, BARRATT B, GREEN D, et al. Prediction of PM_{2.5} concentrations at the locations of monitoring sites measuring PM₁₀ and NO_x, using generalized additive models and machine learning methods: a case study in London[J]. Atmospheric environment, 2020, 240: 117757.
- [5] CAO S J, KONG X R, LI L Y, et al. An investigation of the PM_{2.5} and NO₂ concentrations and their human health impacts in the metro subway system of Suzhou, China[J]. Environmental science: processes & impacts, 2017, 19(5): 666-675.
- [6] CHANG L, CHONG W T, WANG X R, et al. Recent progress in research on PM_{2.5} in subways[J]. Environmental science processes & impacts, 2021, 23(5): 642-663.
- [7] 池东. 天津地铁环境颗粒物浓度测试与水平分析[C]//中国环境科学学会 2019 年科学技术年会: 环境工程技术创新与应用分论坛论文集(三). 西安, 2019: 370-373.
- [8] TARGINO A C, KRECL P, BRIMBLECOMBE P, et al. Spatio-temporal variability of airborne particulate matter in the São Paulo subway[J]. Building and environment, 2021, 189: 107526.
- [9] ZHAO Q, XIAO Y M, LIN J Q, et al. Measurement-based evaluation of the effect of an over-track-exhaust ventilation system on the particulate matter concentration and size distribution in a subway[J]. Tunnelling and underground space technology, 2021, 109: 103772.
- [10] MORENO T, PÉREZ N, RECHE C, et al. Subway platform air quality: assessing the influences of tunnel ventilation, train piston effect and station design[J]. Atmospheric environment, 2014, 92: 461-468.

(编辑: 王艳菊)