

吴明珠<sup>1</sup> 冯楷<sup>1</sup> 翁剑成<sup>1</sup> 魏瑞聪<sup>2</sup> 王晶晶<sup>3</sup> 钱慧敏<sup>3</sup>

# 基于可解释机器学习的大型活动场馆 周边路网运行状态影响研究

## 摘要

举办大型活动会导致周边受影响区域在短时间内集中大量人群和车辆,场馆周边路网与常态交通具有差异化特征。为探究大型活动对场馆周边路网运行状态的影响机理,解析活动规模、路段与活动场馆的空间距离等因素的影响特征,构建融合 XGBoost 算法与部分依赖图的可解释机器学习模型,以捕捉不同因素的非线性效应与协同影响。以北京市为例开展了实证研究,单因素的异质性影响表明:路段与活动场馆的空间距离及活动规模对场馆周边路网运行状态的影响较大,其相对重要度分别达到 27.1% 和 25.4%,距离活动开始/结束的时间对场馆周边路网运行状态存在明显非线性特征,在活动开始前 30~60 min,以及活动结束后 30 min 内,场馆周边 3 km 以内的路段将受到显著影响。二维因素的协同影响表明:当活动规模大于 3 万人时,节假日和不利天气对场馆周边路网运行状态呈负面影响,而在降雨和雾霾天气下,场馆周边路网运行状态受时空影响较大,影响范围为活动开始前 60 min 与结束后 40 min 内距离活动场馆 2.5 km 内的路段。相关研究结论可为大型活动期间道路拥堵成因辨别及制定科学有效的路网管控策略提供量化的决策依据。

## 关键词

城市交通;短时大型活动;路网运行状态;影响关系;XGBoost 模型;部分依赖图

中图分类号 U491.1;TP183

文献标志码 A

收稿日期 2023-06-22

资助项目 国家自然科学基金(52072011)

作者简介

吴明珠,女,硕士生,主要从事智能交通与大数据、交通流建模等研究。wumz1210@163.com

翁剑成(通信作者),男,博士,教授,主要从事交通信息与大数据、交通出行行为等研究。youthweng@bjut.edu.cn

## 0 引言

短时大型活动在有限空间范围和较短时间内集中大量人流与车辆,具有交通强聚散性和需求骤增等特征。活动期间,在天气、日期等外部因素以及活动规模、活动性质等属性因素的共同作用下,场馆周边区域路网的运行压力大幅增加,存在短时持续性交通拥堵、风险突发等问题。因此,为保障活动期间路网运行的畅通与稳定,有必要把握关键要素对大型活动场馆周边路网运行状态的影响机理,实现活动期间周边路网交通态势的精细化推演,以期制定科学合理的管控方案。

研究人员从定性和定量的角度讨论了路网交通流特征和影响机理,但是大型活动所引发的诱增交通需求导致场馆周边路网运行状态与常态化条件存在明显差异,具体表现在交通需求量、流量分布规律及服务层次性等方面<sup>[1]</sup>。研究表明大型活动期间交通需求和交通流具有潮汐性、临时性和非平衡性等特征<sup>[2-3]</sup>。

在场馆周边交通状态影响机理研究方面,通常关注活动期间的客流出行影响与路网交通流运行影响。其中,客流出行影响主要结合出行需求来考量,如洪于亮等<sup>[4]</sup>基于用户画像方法,结合实际客流需求调查,讨论了大型活动的客流出行特征和需求,发现不同群体对步行、排队和等待时间以及出行方式选择等方面存在一定差异。在路网交通流运行影响的刻画方面,王振报等<sup>[5]</sup>考虑交通流时空特征,建立目标函数模型刻画大型活动对路段交通的影响;杨子帆等<sup>[6]</sup>将影响评价指标细化,综合考虑道路拥堵和客流强度,构建大型活动交通影响指数,客观反映大型活动对场馆周边区域路网运行状态的影响情况;Pulugurtha 等<sup>[7]</sup>以大型活动举办地点为端点,利用出行时间度量大型活动对出行时间效率指标的时空影响,评估大型活动当天的出行时间效率指标与正常日的差异;Niu 等<sup>[8]</sup>分区域探究了大型活动对基本图的影响,发现在活动入口处路网交通流平均速度和密度会产生显著波动。明确大型活动突发事件对交通运行的时空影响范围与程

1 北京工业大学 交通工程北京市重点实验室,北京,100124

2 福建省高速公路联网运营有限公司,福州,350001

3 北京市交通运行监测调度中心,北京,100161

度,也能为活动期间交通应急预案的生成与优化提供依据<sup>[9]</sup>.

近年来,离散选择模型、统计回归、机器学习等方法广泛应用于影响关系分析,其中,机器学习在拟合因素的非线性特征时具有更优的建模能力,且能够结合可解释方法挖掘变量对模型结果的潜在影响,现有研究主要引入 SHAP、部分依赖图(Partial Dependence Plots)等解释机器学习模型. Qi 等<sup>[10]</sup>结合 XGBoost 和 SHAP 方法构建可解释框架,探讨了拥堵指数、车道数和天气等因素对高速公路出口交通安全风险的影响; Kashifi 等<sup>[11]</sup>利用 LightGBDT 和 SHAP 构成的可解释机器学习框架分析得到行程距离、出行者年龄等显著影响出行者决策的因素;许奇等<sup>[12]</sup>基于 XGBoost 模型刻画城市轨道交通起讫点客流与建成环境的非线性关系,并利用部分依赖图可视化双变量的交互效应;刘柯良等<sup>[13]</sup>利用部分依赖图可视化梯度提升树(GBDT)模型中自变量与因变量之间的非线性关系.

综上所述,目前对于影响机理的分析中,大部分研究仅从交通流时空影响角度出发,确定因素对路网运行状态的正负关系与影响程度,而忽略了因素局部存在的非线性关联与协同效用,不足以支撑相关部门根据影响关系动态准确地优化管控策略以提高路网运行质量.因此需要考虑更为全面的影响因素,并量化分析影响因素对场馆周边路网运行状态的非线性效应和因素间的协同作用,以制定更加精准合理的交通管控策略.

基于此,为深入剖析和度量多维因素对活动场馆周边路网运行状态的潜在影响关系,本文首先运用 XGBoost 算法建立影响关系模型,探究因素的影响程度.为克服机器学习的黑箱问题,本文进一步引入可解释机器学习方法部分依赖图,以解读影响模型,捕捉活动性质、活动场地和时空影响等要素的非线性效应与异质性影响,并量化二维影响因素组合之间的协同作用机理,从而为大型活动期间周边路网交通组织优化提供支持.

## 1 数据基础与影响因素集

为深入探究活动属性因素和交通流时空特征等多维因素对大型活动场馆周边路网运行状态的影响机理,需要依赖路网基础设施数据、浮动车数据和大型活动信息等多源数据.

### 1.1 研究数据基础

基于北京市基础路网结构数据,结合 2019 年 6 月 1 日—11 月 30 日的浮动车数据与大型活动信息数据,借助 ArcGIS 实现活动期间场馆周边路段速度数据的时空关联.其中:路网结构静态数据主要包括路段的地理位置、道路类型、路段长度、车道数量、车流方向等信息;基于浮动车的路段信息数据,其时间统计周期为 5 min,数据列表如表 1 所示;大型活动信息数据共包含北京市工人体育场、凯迪拉克中心、国家体育馆等主要大型活场馆举办的 57 场短时性大型活动,数据列表如表 2 所示,大型活动场馆的分布位置如图 1 所示.

表 1 北京市大型活动周边路段信息字段说明示例

Table 1 Field description of road section around large-scale events in Beijing

路段名称	时间	路段长度/m	路段速度/(km/h)	路段起点	路段终点	路段方向
玉渊潭南路	20191120T1700	581	34	万寿路	万寿庄路	E—W
玉渊潭南路	20191120T1705	396	24	西翠路	万寿庄路	W—E
复兴路	20191120T1745	977	42	西翠路	万寿路	W—E
太平街	20191120T1830	281	34	南纬路	陶然亭路	S—N
西翠路	20191120T2300	283	47	复兴路	万寿路西街	S—N

表 2 北京市大型群众性活动情况样例

Table 2 Sample table of large-scale mass activities in Beijing

序号	活动名称	举办地点	活动类型	规模/人	起止时间	日期
1	2019 罗大佑巡回演唱会	工人体育场	文艺演出	40 000	19:00—22:00	20190608
2	踏冰筑梦	工人体育场	文艺演出	5 300	19:00—21:30	20190705
3	818 从辛出发群星演唱会	奥体中心体育场	文艺演出	9 000	18:00—21:30	20190818
4	2019 中国好声音巅峰之夜	国家体育馆	文艺演出	45 000	19:00—22:00	20191007
5	2019 年中超联赛	工人体育场	体育比赛	35 000	19:35—22:00	20191026

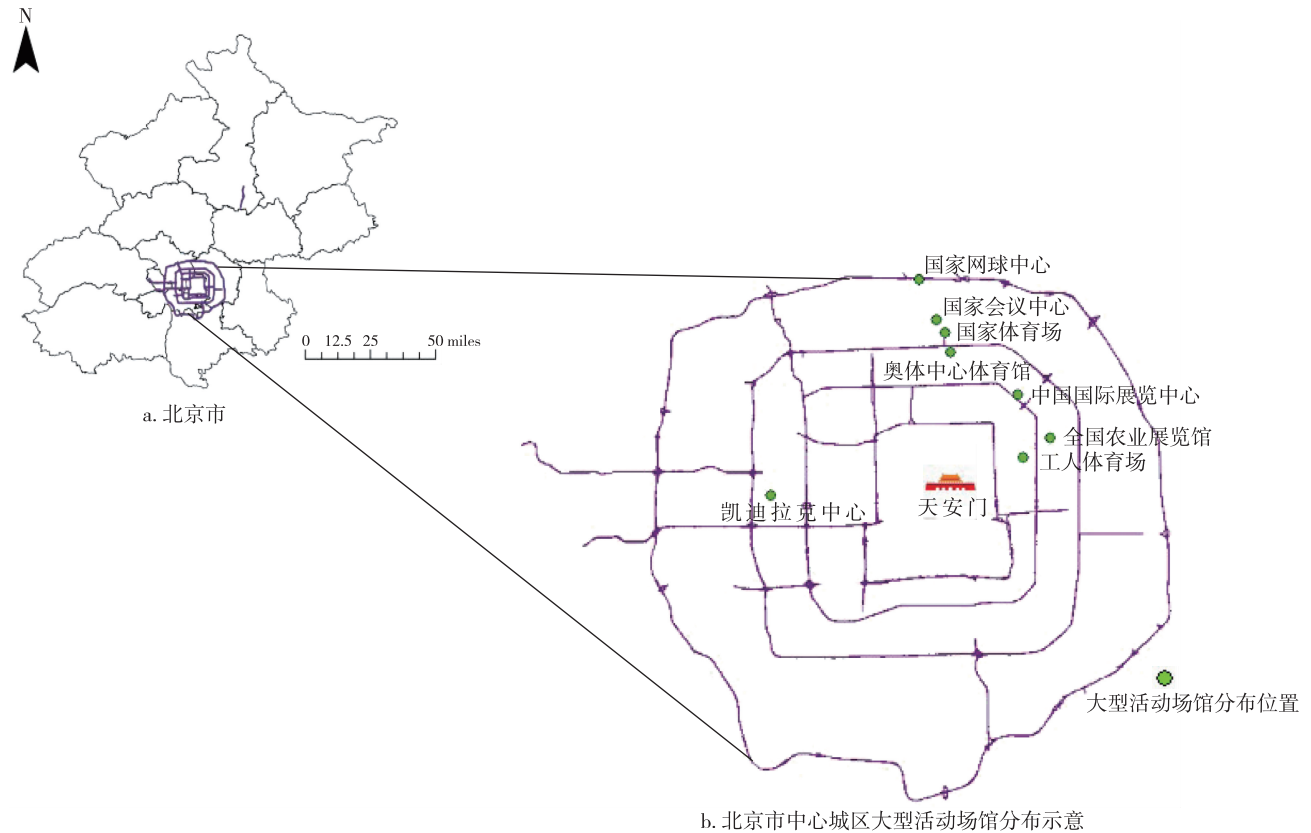


图 1 北京市部分大型活动场馆分布

Fig. 1 Distribution map of some large-scale event venues in Beijing

1.2 影响因素集

考虑到大型活动场馆周边路网的交通需求特征和运行状态与活动固有性质关联性较强<sup>[14]</sup>, 且具有明显时空差异性. 因此, 除考虑天气、日期等影响常态化交通的因素外, 本文围绕大型活动属性与时空维度两个方面构建影响因素集, 其中: 活动属性因素

主要从活动规模、活动性质和场馆区位等内因, 以及日期属性、活动日天气和活动举办时段等外因构建活动因素集; 时空维度因素主要考虑距离活动开始与结束的时间、路段与活动场馆的空间距离. 本文构建的大型活动场馆周边路网运行状态影响因素集如表 3 所示.

表 3 影响因素集

Table 3 Set of influencing factors

维度	影响因素	属性值
大型活动属性	内部因素	
	活动规模	以不同的活动规模进行赋值
	活动性质	体育活动=1; 文艺活动=2; 商业活动=3
	活动场馆区位	根据不同场馆所属区域分别赋值
	活动日期属性	工作日=1; 周末=2; 节假日=3
外部因素	活动日天气状况	晴天=0; 多云=1; 阴天=2; 明显降雨=3; 雾霾=4; 降雪=5
	活动举办时段	白天时段(08:00—18:00)=1; 夜晚时段(18:00—24:00)=2
	时空影响属性	
	距离活动开始/结束的时间	以距离活动开始/结束的时间进行赋值(5 min 粒度)
	路段与活动场馆的空间距离	根据 ArcGIS 生成的路段拓扑点到大型活动场馆的路径距离进行赋值(以米为单位)

1.3 数据处理与关联融合

本文利用线性插值法对速度数据集中丢失或错

误数据进行修复与填充, 之后将北京市静态路网结构、路段速度数据和大型活动场馆信息输入到

ArcGIS 中,并运用其中的标识功能,设路段名称为融合属性,对路段属性、历史速度与场馆周边区域底图属性表关联融合,得到与场馆周边路网匹配后的路段速度.

将提取后的路段速度与表 3 中同时刻的影响因素属性值进行关联匹配;同时,为降低量纲差异的影响,选用 min-max 标准化方法对数据进行归一化处理.当数据集中样本容量为  $n$  时,归一化处理公式如下:

$$y_i = \frac{x_i - \min_{1 \leq j \leq n} \{x_j\}}{\max_{1 \leq j \leq n} \{x_j\} - \min_{1 \leq j \leq n} \{x_j\}} \quad (1)$$

式中,  $y_1, y_2, \dots, y_n \in [0, 1]$ .

## 2 大型活动场馆周边路网运行状态的影响模型

路网运行状态具有复杂性,受到多类因素的综合作用<sup>[15-16]</sup>,机器学习模型不仅能够刻画影响因素的非线性作用机制,还能提供变量的相对重要度<sup>[17]</sup>,有利于对结果的解释,一定程度上克服了线性模型的局限性.同时,为解决机器学习的黑箱问题,需引入可解释机器学习框架以分析响应变量与解释变量之间的耦合作用机理<sup>[18]</sup>.因此,本文选取 XGBoost 建立影响关系模型,量化影响因素对场馆周边路网运行状态的影响程度,并结合部分依赖图

解析单变量的非线性关系与多变量间的协同影响,本文提出的影响模型框架如图 2 所示.

### 2.1 基于 XGBoost 的影响关系模型

XGBoost 是基于梯度提升树 (GDBT) 的改进模型<sup>[19]</sup>,与传统回归模型相比,XGBoost 在刻画非线性关系时性能更优,且不需要预设模型结构,能够有效量化并输出影响因素的相对重要度以体现解释变量对因变量的影响程度.因此,采用 XGBoost 模型探究因素与场馆周边路网运行状态的影响依赖关系.

以表 3 中大型活动属性和时空影响属性作为 XGBoost 的解释变量  $x = [x_1, x_2, \dots, x_8]$ ,被解释变量为路段运行速度.XGBoost 通过前向迭代集成每棵决策树的预测结果作为最终输出预测值,其函数定义如式(2)所示:

$$\hat{y}_i^{(t)} = \sum_{k=1}^t f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i). \quad (2)$$

式中:  $\hat{y}_i^{(t)}$  为  $t$  轮迭代后模型预测第  $i$  个样本的路段运行速度;  $\hat{y}_i^{(t-1)}$  为已知  $t-1$  轮迭代后第  $i$  个样本的预测值;  $f_k(x_i)$  为第  $k$  棵决策树的预测结果;  $f_t(x_i)$  为第  $t$  轮迭代对残差的预测值;  $x_i$  为第  $i$  个样本的特征向量.

XGBoost 的目标函数见式(3).为防止模型过拟合,在损失函数中加入正则项  $\Omega(f_k)$ .

$$O(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k). \quad (3)$$

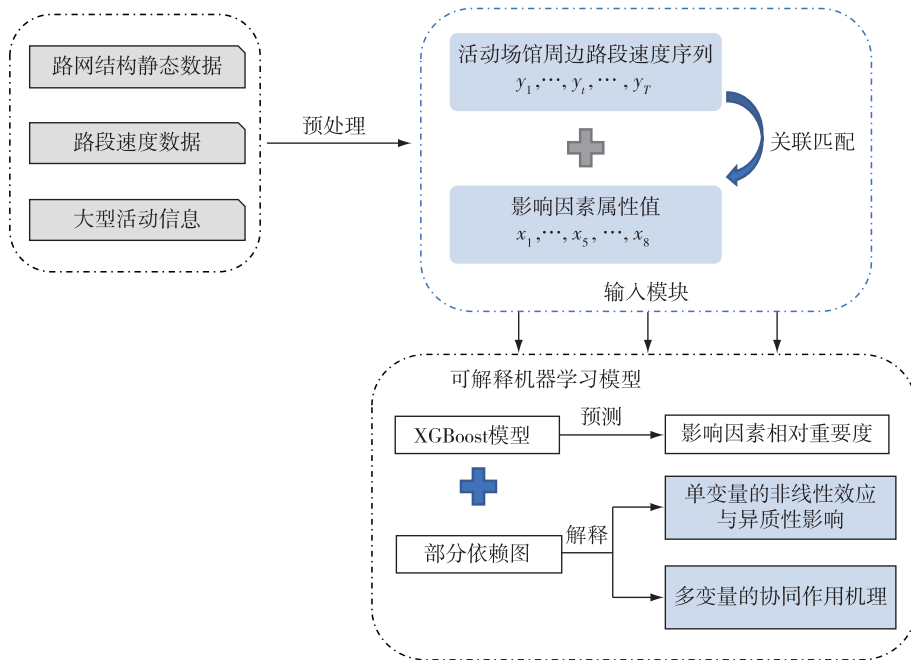


图 2 可解释机器学习影响模型框架

Fig. 2 Framework of impact model via explainable machine learning



式中:  $O(\theta)$  为 XGBoost 的目标函数;  $l(y_i, \hat{y}_i)$  为样本  $x_i$  的损失函数;  $\Omega(f_k)$  表示第  $k$  棵树的正则项, 用于控制模型复杂度, 其计算公式见式(4);  $K$  为学习器个数.

$$\Omega(f_k) = \zeta T_k + \frac{1}{2} \lambda \|\omega_k\|^2. \quad (4)$$

式中:  $T_k$  为第  $k$  棵决策树;  $\omega_k$  为第  $k$  棵树叶子节点的预测值;  $\zeta$  和  $\lambda$  为权重系数.

根据模型最终训练结果计算各个解释变量的相对重要性, 具体计算见式(5).

$$R_l^2 = \frac{1}{K} \sum_{k=1}^K R_l^2(T_k). \quad (5)$$

式中:  $R_l^2$  为变量  $l$  对路段速度的相对重要度;  $J$  为该树的叶节点数.

## 2.2 基于部分依赖图的影响机理解析

虽然 XGBoost 在准确性和泛化能力等方面优于回归模型, 但其可解释性低. 因此, 本文引入部分依赖图解释黑箱机器学习模型中因素间的影响机理, 同时可视化单变量对场馆周边路网运行状态的非线性效应和异质性影响, 以及二维因素组合的协同作用机理. 部分依赖图的核心思想是通过控制其他变量不变, 计算单个或两个解释变量对模型中目标函数的边际贡献<sup>[20]</sup>, 利用部分依赖相对值表示一组变量影响目标函数时的交互作用程度, 具体的计算方法如式(6)所示:

$$\bar{f}_S(x_S) = \frac{1}{N} \sum_{j=1}^N f(x_S, x_C^{(j)}). \quad (6)$$

式中:  $f_S(x_S)$  是部分依赖函数,  $\bar{f}_S(x_S)$  绝对值越大, 说明该特征组合对目标函数的影响程度越强;  $x_S$  为  $S$  中二维特征取值;  $S$  为影响因素属性变量  $x = [x_1, x_2, \dots, x_8]$  的子集, 包含两个选定的特征  $(x_a, x_b)$ ;  $x_C^{(j)}$  为

其他特征在数据集中第  $j$  个实例上对应的取值;  $N$  是数据集中实例数; 特征向量  $x_S$  和  $x_C$  组成总特征空间  $x$ ;  $f$  为 XGBoost 模型.

## 3 实证分析

本文以北京市 57 场大型活动与活动日当天场馆周边路段速度数据为基础, 利用 XGBoost 和部分依赖图分析各类特征与场馆周边路网运行状态的影响关系. 采用网格搜索法和交叉验证法搜索 XGBoost 模型超参数的最优组合, 发现当模型包含 50 棵最大深度为 5 的决策树、学习参数为 0.05、最小样本权重为 3 时, 模型具有最佳拟合效果.

### 3.1 影响因素的相对重要度

将表 3 中的 8 个影响因素作为影响关系模型的输入特征, 学习模型输出各类特征的相对重要度, 其值越大, 特征贡献度越大, 表明该指标对路网运行状态的影响越大.

图 3 是 8 个影响因素的相对重要度. 易知, 大型活动属性和时空影响属性对场馆周边路网运行状态影响具有明显差异. 具体而言, 路段与活动场馆的空间距离的重要度排序第一, 达到 27.1%, 其次是活动规模和距离活动开始/结束的时间, 其重要度分别为 25.4% 和 21.3%, 说明场馆周边路网资源的时空聚集程度和参加活动人群集聚规模对路网运行状态的影响显著, 而活动性质和活动举办时段的影响不显著.

在明确各类因素对场馆周边路网运行状态的定量影响基础上, 本文选取相对重要度较高的因素应用于部分依赖图的影响机理解析, 分别讨论单一变量对路网运行状态的非线性效应与多变量的协同作用.

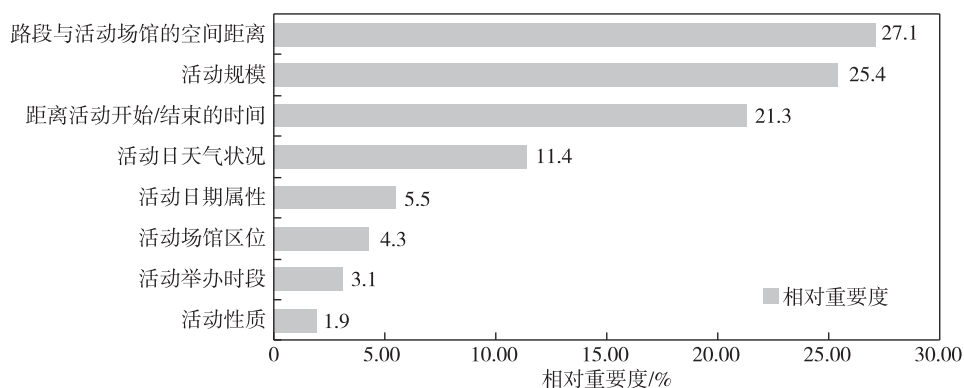


图 3 影响因素的相对重要度

Fig. 3 Relative importance of influencing factors

### 3.2 单变量对场馆周边路网运行状态的非线性效应与异质性影响

根据重要度排序选取路段与活动场馆的空间距离、活动规模、距离活动开始/结束的时间、活动日天气状况、活动日期属性共5个变量,利用部分依赖图方法,分别计算每个变量对路段速度的边际贡献,以探究其对场馆周边路网运行状态的非线性关系或异质影响。

#### 1) 路段与活动场馆的空间距离

路段与活动场馆的空间距离对场馆路网运行状态的影响关系如图4所示.当空间距离小于2 km时,处于较高影响状态,当空间距离大于2 km时,整体呈正相关趋势,而当空间距离大于3 km后,其影响程度显著降低.这表明越接近于活动场馆,路段中交通流的聚集程度越高,路网运行质量越低,该结论同样反映了大型活动场馆对路网运行状态的影响半径约为3 km,这为大型活动交通运输保障的空间布局提供参考。

#### 2) 活动规模

活动规模变量对场馆路网运行状态的影响关系如图5所示.可以看出:当活动规模低于2万人次时,路网运行状态与活动规模呈显著负相关;当活动规模为2万~4万人次时,路网运行质量呈现波动式降低,场馆周边路段运行速度降低约6 km/h;之后其影响不再随着规模的增加而明显变化.这可能是由于当活动规模超过4万人时场馆周边路网资源已趋于饱和,道路拥堵严重,路段运行速度变化较小。

#### 3) 距离活动开始/结束的时间

距离活动开始/结束的时间对场馆路网运行状态的影响关系如图6所示。

从活动开始前2.5 h,路段运行速度开始下降,

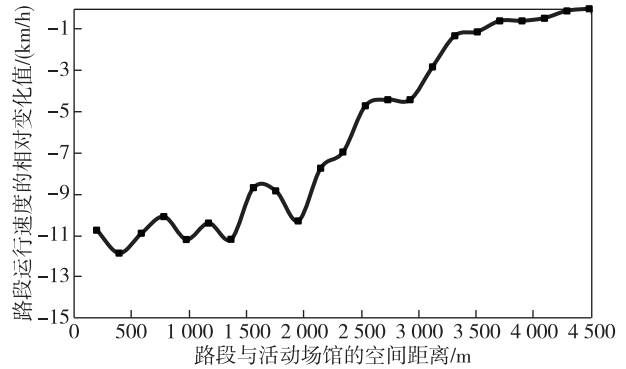


图4 路段与活动场馆的空间距离的非线性影响  
Fig. 4 Nonlinear influence of spatial distance from road to the event venue

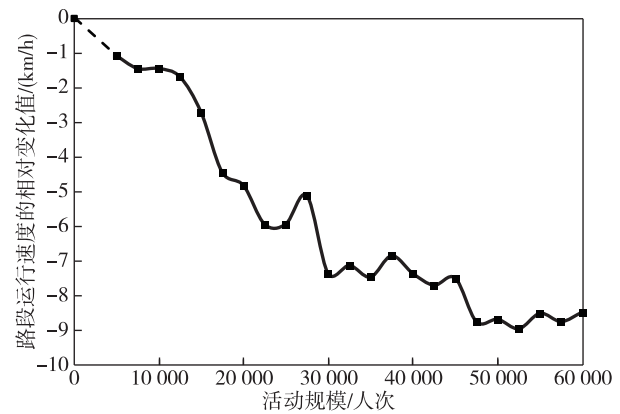


图5 活动规模的非线性影响  
Fig. 5 Nonlinear influence of event scale

在活动开始前1~2 h内,路段运行速度下降速率变缓.但在开始前30 min至1 h内,路段运行速度下降幅度增大,最高降低了6.8 km/h,而在活动开始前30 min内,路网运行状态逐渐恢复至正常水平。

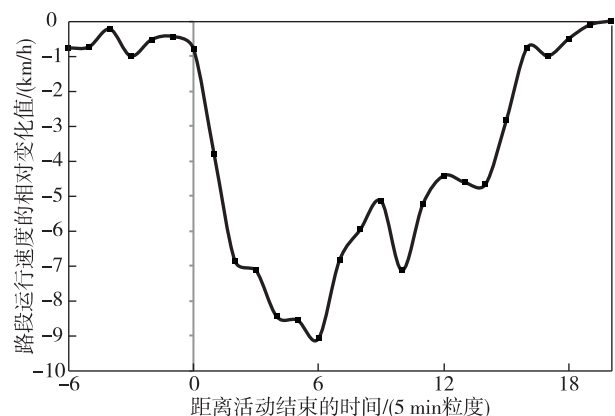
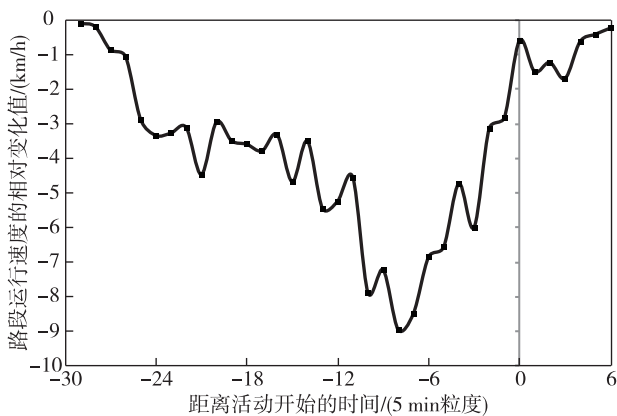


图6 距离活动开始/结束的时间对场馆周边路网运行状态的非线性影响

Fig. 6 Nonlinear influence of time before start and after end of the event on traffic flow around the venue

在活动结束后,时间变量与路网运行状态呈负相关,路段运行速度开始迅速下降,平均降低了 5.6 km/h,持续约 30 min.在活动结束后的 30~90 min 内,路网运行状态逐渐恢复到正常水平.这种特征说明,在活动开始前 30 min 至 1 h 内和活动结束后的 30 min 内客流的集散程度增强,导致路网运行质量降低.

#### 4) 活动日天气状况

活动日天气状况对场馆路网运行状态的影响关系如图 7 所示.晴天和多云对路网运行状态较为一致且影响不显著;阴天对路网运行状态具有抑制作用,但影响程度较弱.相较于晴天,雾霾天气下场馆周边路段运行速度平均降低了 6.6 km/h,原因可能是中度及以上雾霾天气的能见度会缩小驾驶员视野范围,对其驾驶行为产生影响,从而导致路段运行速度明显降低.降雨天气下活动场馆周边路网运行状态降幅明显.这表明雾霾和降雨天气举办大型活动会增加场馆周边路网交通拥堵和事故风险,在进行交通运输保障时可制定针对性路网管控策略.

#### 5) 活动日期属性

活动日期属性对场馆路网运行状态的影响关系如图 8 所示.相比工作日,在周末和节假日举办大型活动时场馆周边路段运行速度分别下降 12.6% 和 14.7%,这可能是大型活动客流与周末和法定节假日期间居民弹性出行的交通流叠加,出现较为明显的车流与人流增长.

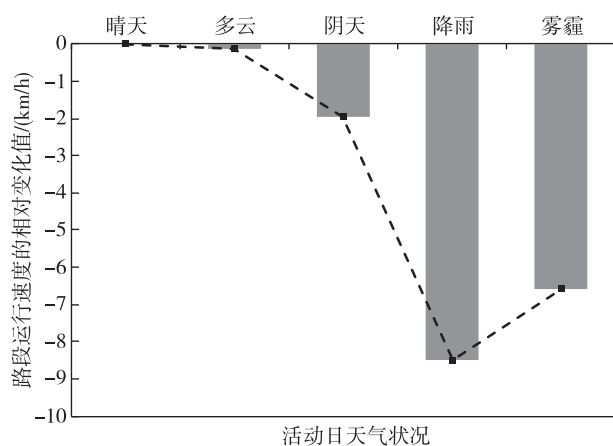


图 7 活动日天气状况的异质性影响

Fig. 7 Heterogeneous influence of weather conditions

### 3.3 场馆周边路网运行状态的二维协同作用机理分析

场馆周边路网运行状态具有复杂矢量性,与大

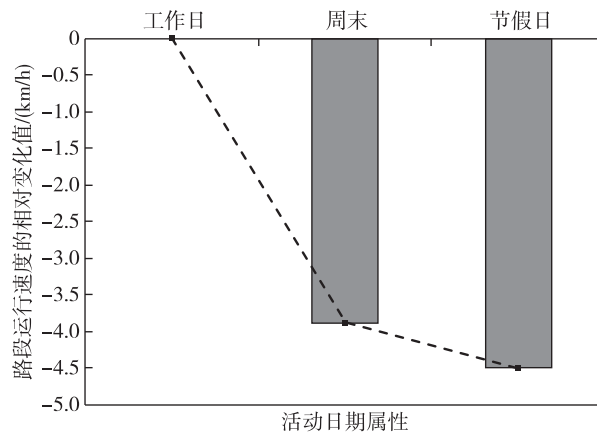


图 8 活动日期属性的异质性影响

Fig. 8 Heterogeneous influence of event date attributes

型活动属性密切相关,同时又受时空影响.为进一步刻画两个单变量组合成双变量后对场馆周边路网运行状态的交互影响,本文结合单变量的相对重要度与其非线性特征,选取活动规模和活动日天气状况、活动规模和活动日期属性、活动日天气状况和路段与活动场馆的空间距离、活动日天气状况和距离活动开始/结束时间共 4 类因素组合,采用二维部分依赖图可视化其协同作用机理.

#### 1) 活动规模和活动日天气状况

活动规模与活动日天气状况对场馆周边路网运行状态的协同作用关系如图 9 所示.当活动规模低于 2 万人次时,活动日天气对路网运行状态的整体影响不明显;当活动规模高于 3 万人次时,雾霾和降雨天气对路网运行质量的负面影响逐渐显著,而相比雾霾天气,降雨天气与活动规模对路网运行状态的协同影响更明显.

#### 2) 活动规模和活动日期属性

活动规模与活动日期属性对场馆周边路网运行状态的协同作用关系如图 10 所示.在节假日与活动规模高于 3 万人次的共同影响下,路网运行质量降低幅度更大.

#### 3) 活动日天气状况和路段与活动场馆的空间距离

活动日天气状况和路段与活动场馆的空间距离对场馆周边路网运行状态的协同作用关系如图 11 所示.在降雨和雾霾天气条件下,路段与活动场馆的空间距离对场馆周边路网运行状态具有显著影响,尤其是距离场馆 2 500 m 内的路段.

#### 4) 活动日天气状况和距离活动开始/结束时间

活动日天气状况和距离活动开始/结束时间对

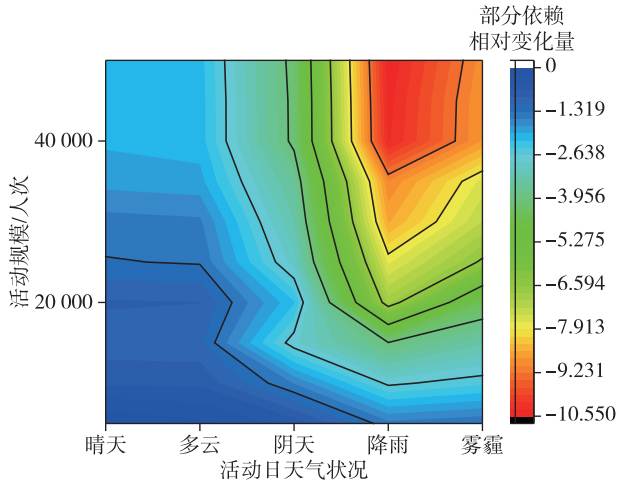


图 9 活动规模与活动日天气状况的协同作用  
Fig. 9 Synergy of event scale and weather conditions

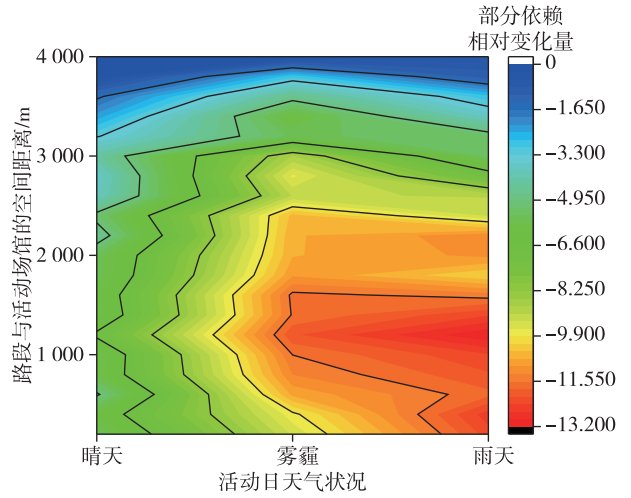


图 11 活动日天气状况和路段与活动场馆的空间距离的协同作用

Fig. 11 Synergy of weather conditions and spatial distance from road to event venue

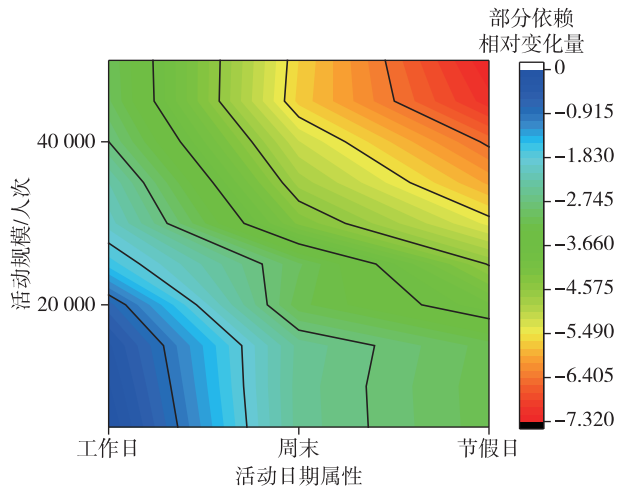


图 10 活动规模与活动日期属性的协同作用  
Fig. 10 Synergy of event scale and date attributes

场馆周边路网运行状态的协同作用关系如图 12 所示.与路段和活动场馆的空间距离类似,在降雨和雾霾天气下,距离活动开始/结束的时间距离对场馆周边路网运行状态的影响较为显著.其中:降雨天气对路网运行状态的影响作用时间约为活动开始前 2.5 h 至活动开始后的 15 min 内以及活动结束后 10 min 至结束后 90 min 内;雾霾天气对路网运行状态的影响时间短于降雨天气,在临近活动开始时路网运行状态基本恢复正常水平.

#### 4 结论与展望

本文基于路网数据、浮动车数据及大型活动信

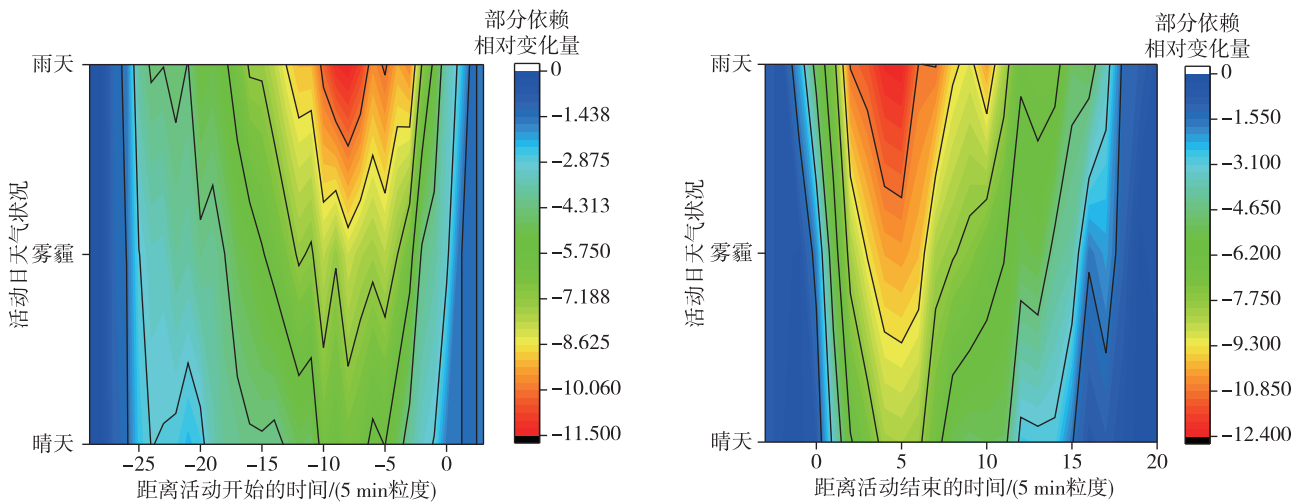


图 12 活动日天气状况和距离活动开始/结束时间的协同作用  
Fig. 12 Synergy of weather conditions and time before start and after end of the event



息等多源异构数据, 构建基于 XGBoost 的影响关系模型, 结合部分依赖图解释多维因素对大型活动场馆周边路网运行状态的非线性关系、异质性影响和协同作用机理, 并以北京市的实际大型活动多源数据为基础开展模型实证分析, 得到主要结论如下:

1) 路段与活动场馆的空间距离和活动规模是影响场馆周边路网运行状态的主要因素, 其相对重要度分别为 27.1% 和 25.4%.

2) 单变量对场馆周边路网运行状态的非线性效应与异质性影响表现为: 距离活动开始/结束的时间对场馆周边路网运行状态存在明显非线性特征; 大型活动对场馆周边路网运行状态的影响半径约为 3 km, 主要影响时段为活动开始前 30 min 至 1 h 内与活动结束后 30 min.

3) 多变量对场馆周边路网运行状态的协同影响表现为: 节假日和活动日不利天气在活动规模大于 3 万人时与场馆周边路网运行状态具有负相关性, 并且其负相关性随活动规模的增大而愈发显著; 在雾霾和降雨天气条件下, 活动开始前 1 h 与活动结束后 40 min 内对距离活动场馆 2 500 m 内路段运行状态呈显著负面影响.

解析多维因素对大型活动场馆周边路网运行状态的非线性影响关系与协同作用机理, 有助于精准制定路网管控策略. 研究结果表明, 大型活动交通保障不仅需从交通生成角度分析活动本身对场馆周边路网运行状态的影响, 还应考虑大型活动场馆的潜在影响时空范围, 从交通分布角度研究大型活动期间场馆周边路网交通资源协同配置, 进而改善路网运行质量. 本文的研究重点关注路段运行速度对路网状态的影响, 未来可结合交通指数、道路负荷度等指标探究因素对场馆周边路网运行状态的影响, 更加全面地反映场馆周边交通流运行特征和影响机制, 进而设置科学合理的交通专项保障实施方案.

## 参考文献

### References

- [ 1 ] 戴铃. 大型活动交通运行多源风险协同评估方法 [ D ]. 南京: 东南大学, 2021  
DAI Ling. Study on the method of collaborative assessment of traffic operation multisource risk for big event [ D ]. Nanjing: Southeast University, 2021
- [ 2 ] 梁兴柳. 大型活动对区域道路路网稳定性的影响分析方法研究 [ D ]. 石家庄: 石家庄铁道大学, 2021  
LIANG Xingliu. Research on the analysis method of the influence of large-scale activities on the stability of regional road traffic network [ D ]. Shijiazhuang: Shijiazhuang Tiedao University, 2021
- [ 3 ] 崔洪军, 陆建, 魏连雨, 等. 大型活动散场交通流的时空消散规律 [ J ]. 中国公路学报, 2007, 20(2): 102-107  
CUI Hongjun, LU Jian, WEI Lianyu, et al. Traffic flow time space dissipation rule under large-scale activity [ J ]. China Journal of Highway and Transport, 2007, 20(2): 102-107
- [ 4 ] 洪于亮, 许宁, 秦焕美, 等. 基于用户画像的大型冬季户外活动客流需求特征挖掘: 以哈尔滨和长春国际滑冰雪马拉松为例 [ J ]. 科学技术与工程, 2023, 23(10): 4400-4407  
HONG Yuliang, XU Ning, QIN Huanmei, et al. Characteristics mining of passenger flow based on user portrait in large winter outdoor events: take Harbin and Changchun international skating marathon as an example [ J ]. Science Technology and Engineering, 2023, 23(10): 4400-4407
- [ 5 ] 王振报, 李金山, 陈艳艳. 大型活动期间交通影响分析方法研究 [ J ]. 武汉理工大学学报 (交通科学与工程版), 2010, 34(4): 758-761  
WANG Zhenbao, LI Jinshan, CHEN Yanyan. Research on traffic impact analysis method under special events [ J ]. Journal of Wuhan University of Technology (Transportation Science & Engineering), 2010, 34(4): 758-761
- [ 6 ] 杨子帆, 徐海辉, 钱慧敏, 等. 基于多源数据的城市大型活动交通影响评价方法 [ J ]. 交通工程, 2022, 22(3): 7-12  
YANG Zifan, XU Haihui, QIAN Huimin, et al. Traffic impact evaluation method for urban large-scale activities based on multi-source data [ J ]. Journal of Transportation Engineering, 2022, 22(3): 7-12
- [ 7 ] Pulugurtha S S, Duddu V R, Venigalla M. Evaluating spatial and temporal effects of planned special events on travel time performance measures [ J ]. Transportation Research Interdisciplinary Perspectives, 2020, 6: 100168
- [ 8 ] Niu X J, Zhao X M, Xie D F, et al. Impact of large-scale activities on macroscopic fundamental diagram: field data analysis and modeling [ J ]. Transportation Research Part A: Policy and Practice, 2022, 161: 241-268
- [ 9 ] 沈凌, 陆建, 王成晨. 面向大型活动的交通应急预案快速生成与动态优化方法 [ J ]. 交通信息与安全, 2021, 39(3): 33-40  
SHEN Ling, LU Jian, WANG Chengchen. Rapid generation and dynamic optimization of traffic emergency plans for large-scale events [ J ]. Journal of Transport Information and Safety, 2021, 39(3): 33-40
- [ 10 ] Qi H, Yao Y, Zhao X H, et al. Applying an interpretable machine learning framework to the traffic safety order analysis of expressway exits based on aggregate driving behavior data [ J ]. Physica A: Statistical Mechanics and its Applications, 2022, 597: 127277
- [ 11 ] Kashifi M T, Jamal A, Kashifi M S, et al. Predicting the travel mode choice with interpretable machine learning techniques: a comparative study [ J ]. Travel Behaviour and Society, 2022, 29: 279-296
- [ 12 ] 许奇, 李雯茜, 陈越, 等. 建成环境对城市轨道交通起讫点客流的非线性影响及阈值效应 [ J ]. 交通运输系统工程与信息, 2023, 23(4): 290-297

- XU Qi, LI Wenqian, CHEN Yue, et al. Nonlinear and threshold effects of built environment on origin-destination flows of urban rail transit [J]. *Journal of Transportation Systems Engineering and Information Technology*, 2023, 23(4): 290-297
- [13] 刘柯良, 陈坚, 祝焯, 等. 社区建成环境对小汽车使用行为的非线性影响模型 [J]. *北京交通大学学报*, 2022, 46(3): 49-56
- LIU Keliang, CHEN Jian, ZHU Ye, et al. Non-linear impact model of community built environment on car usage behavior [J]. *Journal of Beijing Jiaotong University*, 2022, 46(3): 49-56
- [14] 付宇, 翁剑成, 钱慧敏, 等. 基于 XGBoost 算法的大型活动期间轨道进出站量预测 [J]. *武汉理工大学学报 (交通科学与工程版)*, 2020, 44(5): 832-836
- FU Yu, WENG Jiancheng, QIAN Huimin, et al. Prediction of metro passenger flow during large-scale activities based on XGBoost algorithm [J]. *Journal of Wuhan University of Technology (Transportation Science & Engineering)*, 2020, 44(5): 832-836
- [15] Zheng Y, Wang S Y, Dong C J, et al. Urban road traffic flow prediction: a graph convolutional network embedded with wavelet decomposition and attention mechanism [J]. *Physica A: Statistical Mechanics and its Applications*, 2022, 608: 128274
- [16] Nigam A, Srivastava S. Hybrid deep learning models for traffic stream variables prediction during rainfall [J]. *Multimodal Transportation*, 2023, 2(1): 100052
- [17] Ullah I, Liu K, Yamamoto T, et al. Modeling of machine learning with SHAP approach for electric vehicle charging station choice behavior prediction [J]. *Travel Behaviour and Society*, 2023, 31: 78-92
- [18] Angelini M, Blasilli G, Lenti S, et al. A visual analytics conceptual framework for explorable and steerable partial dependence analysis [J]. *IEEE Transactions on Visualization and Computer Graphics*, 2023: 1-16
- [19] Chen T Q, Guestrin C. XGBoost: a scalable tree boosting system [C] // *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, CA, USA. New York, NY, USA: ACM, 2016: 785-794
- [20] Hastie T, Tibshirani R, Friedman J H. *Data mining, inference, and prediction* [M]. 2nd ed. New York: Springer, 2009

## Research on impact of short-term large-scale events on nearby traffic flow via interpretable machine learning

WU Mingzhu<sup>1</sup> FENG Kai<sup>1</sup> WENG Jiancheng<sup>1</sup> WEI Ruicong<sup>2</sup> WANG Jingjing<sup>3</sup> QIAN Huimin<sup>3</sup>

<sup>1</sup> Beijing Key Laboratory of Traffic Engineering, Beijing University of Technology, Beijing 100124, China

<sup>2</sup> Fujian Expressway Group Co., Ltd., Fuzhou 350001, China

<sup>3</sup> Beijing Municipal Transportation Operations Coordination Center, Beijing 100161, China

**Abstract** The large number of people and vehicles gathered in a short period of time around large-scale events will lead to a differentiated traffic flow. Here, an interpretable machine learning model integrating XGBoost algorithm and partial dependence plots is proposed to capture the nonlinear effects and synergistic influences of large-scale events and their characteristics on the operation of nearby road network, and an empirical study has been conducted in Beijing. The heterogeneity of single factors shows that the distance of road section away from event venue and the event scale have great impact on nearby traffic flow, with relative importance of 27.1% and 25.4%, respectively; time before start and after end of the event has obvious nonlinear characteristics, and the road sections within 3 km from the venue will be significantly affected within 30–60 minutes before the event and 30 minutes after the event. The synergistic effect of two-dimensional factors shows that, if an event attracted more than 30,000 people, holidays and adverse weather have a negative impact on the nearby traffic flow; in rain or haze weather, the road section within 2.5 km from the venue will be affected within 60 minutes before the event and 40 minutes after the event. The findings can provide quantitative data support for identifying the causes of road congestion and formulating reasonable and effective road network control strategies during large events.

**Key words** urban transportation; short-term large-scale events; road network operation state; influencing relationship; eXtreme Gradient Boosting (XGBoost); partial dependence plots