

基于可视化技术的交叉口短时交通流预测研究<sup>\*</sup>

何矿元, 李慧, 徐小金, 胡蓉, 黄丹

(西华大学 汽车与交通学院, 四川 成都 610039)

**摘要:** 为通过视觉图形实现交通流时序特征可视化, 精准掌握交通大数据驱动下交叉口交通主体的移动趋势, 构建交叉口短时交通流可视化预测系统。通过 Python 中的 Matplotlib 实现交叉口交通流时序可视化, 利用 ARIMA 模型进行短时交通流预测, 并以 OpenITS 合肥市示范区黄山路—科学大道交叉口数据进行实例验证。结果表明, 该系统可实时、在线实现不同时段交通流分布规律可视化, 并能有效提取交通流时序特征, ARIMA(1,1,0)模型的 3 个评价指标的预测误差均小于 10%, 具有较高的预测精度。

**关键词:** 城市交通; 交通流预测; 交叉口; 可视化技术; ARIMA 模型

**中图分类号:** U491.14

**文献标志码:** A

**文章编号:** 1671-2668(2023)01-0022-05

通过挖掘交通大数据深层次特征, 探究平面交叉口交通流分布规律, 对平面交叉口短时交通流进行精准预测, 提高交叉口通行能力, 是缓解城市交通拥堵的热门课题。短时交通流预测时间不超过 15 min。目前主要有模型驱动、人工智能及数据驱动等短时交通流预测方法<sup>[1-2]</sup>。张艺铭等建立轨道交通短时客流预测模型, 通过改进灰狼优化算法优化 BP 神经网络的权值和阈值, 提高了模型的预测精度和稳定性<sup>[3]</sup>。田保慧等提出一种基于时空特征分析的短时交通流预测模型, 以长春市局部路网的实测数据为基础验证了模型的有效性和可行性<sup>[4]</sup>。陈喜群等利用图卷积网络模型进行短时交通流预测, 并以城市大规模路网真实浮动车数据对模型进行了测试<sup>[5]</sup>。Hou Q. Z. 等利用模糊逻辑分析将 ARIMA (差分整合移动平均自回归模型) 和 WNN 两个模型预测交通流加权得到最终预测值, 证明了混合模型的预测性能比单一模型更好<sup>[6]</sup>。钟颖等基于极端梯度上升算法分别构建目标路段时间序列预测模型及时空序列预测模型, 实现了对路段未来时段平均旅行时间的预测<sup>[7]</sup>。交叉口短时交通流预测是在大数据分析的基础上建立交通流预测模型, 现有研究侧重于通过改进预测模型来提高预测精度及效率, 而对微观交通流所呈现的时间变化特性关注较少。考虑到 Python 语言对于大数据挖掘可视化的普及性, 本文通过 Python 可视化工具建立交通流可视

化预测系统, 将微观交通流时间变化特性以直观、图像化语言进行表达, 对交叉口短时交通流进行预测, 为不同交通场景预测提供可移植方案。

## 1 可视化预测系统

### 1.1 系统模块

可视化预测系统由数据输入、模型标定、预测输出 3 个模块构成。如图 1 所示, 数据输入模块以交通流时间序列作为输入层, 应用 Python 下 Numpy、Pandas、Matplotlib 等数据分析工具挖掘交通流数据的时空特性, 可视化交通流时间序列, 验证数据的平稳性, 并对非平稳数据进行平稳化处理; 模型标定

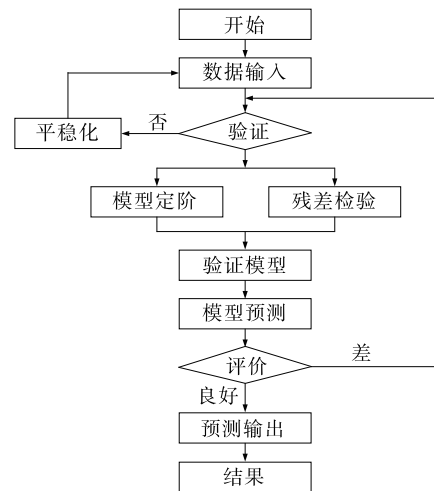


图 1 可视化预测系统流程

<sup>\*</sup> 基金项目: 第 31 届世界大学生夏季运动会交通保障工作培训服务(212293)

模块实现对数据的内部分析,以 AIC(赤池信息准则)、BIC(贝叶斯信息准则)、残差 QQ 图可视化建模过程,拟合 ARIMA 预测模型,确定最优模型参数;预测输出模块将预测交通流以时序图可视化呈现。

一般可视化流程主要为数据收集→数据分析和挖掘→可视化,实现从原始交通流数据到可视化图像。本文数据可视化过程见图 2。其中 ARIMA 模型作为系统的执行层,是系统的核心环节,通过它完成模型标定,实现对交通流的预测。

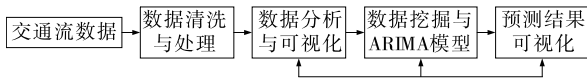


图 2 交通流预测的可视化过程

## 1.2 ARIMA 预测模型

ARIMA 模型将预测对象随时间推移而形成的数据序列视为一个随机变量,随机变量的相关性由历史数据与随机干扰因素结合建立时间序列模型进行识别,并认为这种相关性代表了未来的发展趋势,从而通过已有历史数据及模型来预测未来值。

ARIMA( $p, d, q$ ) 中,AR 为自回归, $p$  为自回归项数;MA 为移动平均, $q$  为移动平均项数。滞后项为  $p$  阶的 AR 模型记为 AR( $p$ ),表达式如下:

$$Y_t = \mu + \beta_1 \times Y_{t-1} + \beta_2 \times Y_{t-2} + \cdots + \beta_p \times Y_{t-p} + \epsilon_t \quad (1)$$

式中: $Y_t$  为时间序列  $t$  时刻的值; $\mu$  为时间序列均值; $\beta$  为权重; $\epsilon_t$  为白噪声序列。

滞后项为  $q$  阶的 MA 模型记为 MA( $q$ ),表达式如下:

$$Y_t = \mu - \theta_1 \times Y_{t-1} - \theta_2 \times Y_{t-2} - \cdots - \theta_q \times Y_{t-q} + \epsilon_t \quad (2)$$

式中: $\theta$  为权重; $\epsilon_t$  为白噪声序列,即随机误差项。

上述两种模型的组合即为 ARMA( $p, q$ ) 模型,表达式如下:

$$Y_t = \mu + \beta_1 \times Y_{t-1} + \beta_2 \times Y_{t-2} + \cdots + \beta_p \times Y_{t-p} - \theta_1 \times Y_{t-1} - \theta_2 \times Y_{t-2} - \cdots - \theta_q \times Y_{t-q} + \epsilon_t \quad (3)$$

该模型中,序列中的每个观测值用变量自身的历史数据及自回归项中的误差项累加组合来表示。

## 1.3 建模步骤

(1) 平稳性检验。平稳性要求经样本时间序列所得到的拟合曲线在未来一段时间内仍能按照现有形态的惯性延续下去。通过观察时间序列图、单位

根检验(ADF)可判断样本时间序列的平稳性。

(2) 平稳性处理。如果样本时间序列不平稳,则采取多阶差分、移动平均等方法进行处理,使序列平稳。

(3) 白噪声检验。白噪声序列在任何两个时点的随机变量都不相关,没有研究意义。通过 Python 实现白噪声检验,对 acorr\_ljungbox 函数下返回值 LB 统计量和 Box-Pierce 进行判断,若统计量  $p \leq 0.05$ ,则有效拒绝原假设,序列为非白噪声序列。

(4) 模型定阶。由差分的次数确定  $d$ ,根据自相关函数(ACF)图和偏自相关函数(PACF)图初步判定  $p, q$  的取值范围,其最优模型参考 AIC、BIC 和 HQIC(汉南一奎因信息准则)准则。在 ARMA 模型中,如果 ACF 和 PACF 都是拖尾,则模型是 ARMA;如果 ACF 是拖尾,PACF 是截尾,则模型简化为 AR;如果 ACF 和 PACF 都是截尾,则模型简化为 AR;如果 ACF 是截尾,PACF 是拖尾,则模型简化为 MA。

(5) 参数估计与残差分析。参数估计采用最大似然估计法、最小二乘法。残差是实际值与估计值的差,如果回归模型正确,可将残差作为误差的预测值,为未来模型假设的合理性及数据可靠性提供依据。通过残差分析验证残差是否为白噪声序列,通过相关图法和残差 QQ 图直观检验残差序列是否为正态分布。若残差为正态分布,说明残差序列是白噪声序列,模型的信息提取充分。

(6) 模型预测及效果评价。利用 ARIMA 模型对短时交通流进行预测。通过平均绝对误差  $\delta_{MAE}$ 、平均绝对百分比误差  $\delta_{MAPE}$  和均方根误差  $\delta_{RMSE}$  评价模型的预测精度。

## 2 实例验证

### 2.1 数据来源及处理

以合肥黄山路—科学大道交叉口(见图 3)交通

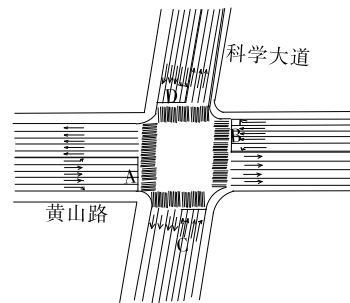


图 3 合肥市黄山路—科学大道交叉口示意图

流数据作为试验数据,包含该交叉口 A 进口道 2016 年 10 月 11—13 日 6:00—22:00 时的交通流数据。统计时间间隔为 3 min 内通过交叉口 A 道路断面的交通流。对于 13 日 11:00—12:00 时缺失的极少部分数据,采用相邻数据平均法补充。

## 2.2 数据可视化分析

在 Python 中导入交通流数据,利用 Matplotlib 包对交通流数据进行描述性分析,以时间序列图的形式实现交通流数据的时变规律可视化。如图 4~6 所示,该交叉口进口道交通流走势呈马鞍形,6:00—7:00、19:00—22:00 时处于时序图波谷,

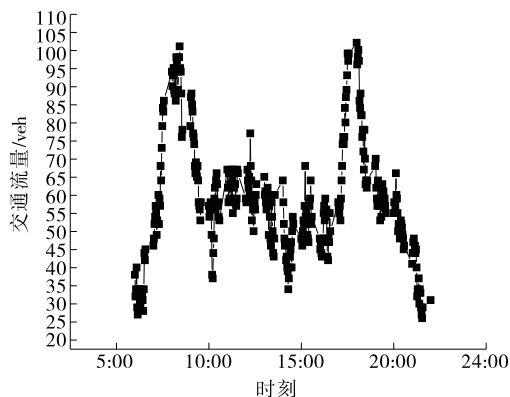


图 4 11 日交通流

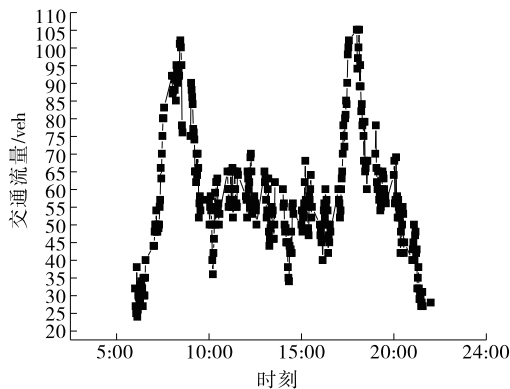


图 5 12 日交通流

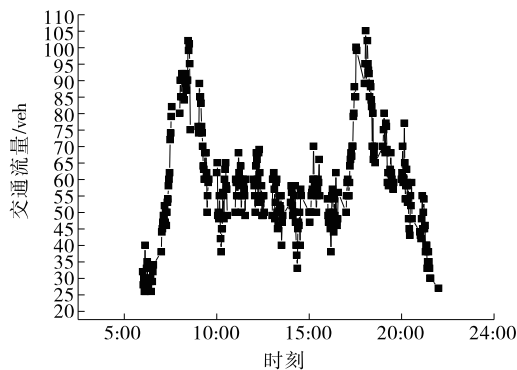


图 6 13 日交通流

8:00—9:00、18:00—19:00 时处于时序图波峰。该进口道 11—13 日的交通流高峰小时系数分为 0.861、0.968、0.957,表明交通流在短小时内迅速增长的现象高于其他时间的某个小时,反映了短时交通流的非线性、不确定性特征。

## 2.3 模型构建

以 11 日交通流  $Q_{11}$  作为训练集,12 日交通流  $Q_{12}$ 、13 日交通流  $Q_{13}$  作为测试集。用  $Q_{11}$  标定 ARI-MA 模型,确立模型参数,预测  $Q_{12}$ 、 $Q_{13}$ ,并将预测值与实际值进行对比分析。

### 2.3.1 序列平稳性检验及平稳化处理

通过 ADF 检验测试数据是否具有时间序列平稳性特征。若检验序列中存在单位根,则回归分析中存在伪回归,序列为非平稳时间序列。

以 Python 中 Statsmodels 包下 Adfuller 进行 ADF 检验,若 T 统计量显著小于 3 个置信度(1%, 5%, 10%)下临界统计值,且  $p$  值非常接近零,则时间序列平稳。检验结果显示 T 统计量明显大于 3 个置信度下临界统计值,且  $p$  值为 0.167,表明原始序列具有非平稳性,需进行差分处理。一阶差分后时间序列见图 7。差分序列  $dQ$  的 ADF 单位根检验结果见表 1。

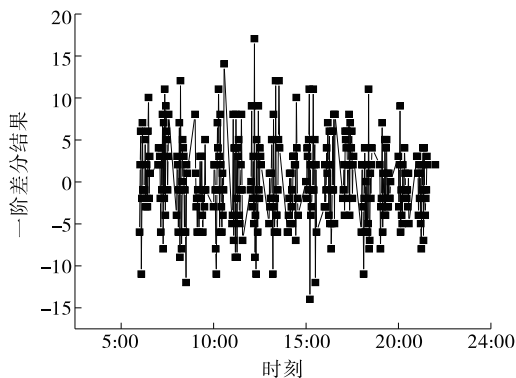


图 7 原始序列一阶差分结果

表 1 ADF 单位根检验结果

交通流	T 统计量	不同置信度(%)下临界统计值			$p$ 值
		1	5	10	
$Q$	-1.997 9	-3.451 0	-2.870 6	-2.571 6	0.287 4
$dQ$	-10.558 0	-3.451 2	-2.870 7	-2.571 7	0.000 0

由表 1 可知:差分后 ADF 检验完全拒绝原假设,时间序列平稳,符合 ARIMA 模型对时间序列数据平稳可逆的要求。确定模型  $ARIMA(p, d, q)$  中  $d=1$ 。

2.3.2 模型定阶

采用相关图法及 AIC、BIC 和 HQIC 准则确定模型  $p$  和  $q$  参数值,准则值越小,预测效果越好。

首先通过差分后序列的 ACF 和 PACF 函数图初步拟定模型阶数(见图 8 和图 9),然后从低阶到高阶逐个进行尝试,挑选最优模型参数。将 ARIMA(0,1,1)、ARIMA(1,1,0)、ARIMA(1,1,1)、ARIMA(2,1,1)、ARIMA(2,1,2)5 个模型作为试验模型进行测试,结果见表 2。由表 2 可知: $p=1$ 、 $q=0$  时,3 个准则的值最小。原始序列拟合为 ARIMA(1,1,0)。

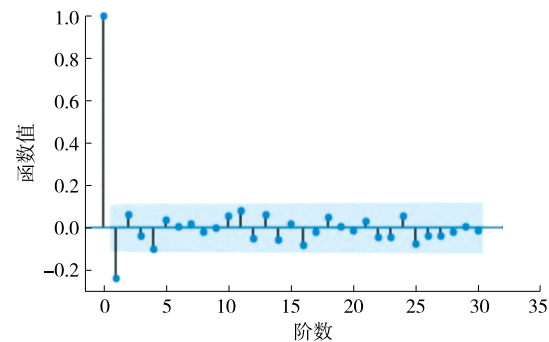


图 8 自相关函数图

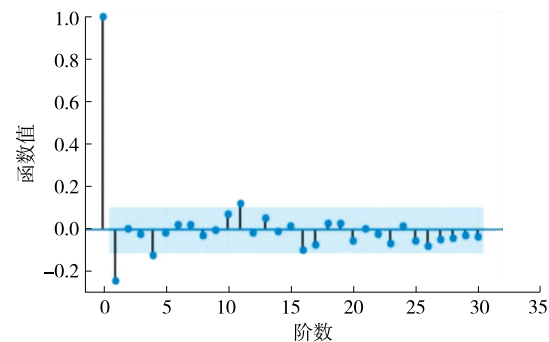


图 9 偏自相关函数图

表 2 模型不同阶数组合测试结果

试验模型	AIC 值	BIC 值	HQIC 值
ARIMA(0,1,1)	1 974.994 5	1 986.299 4	1 979.508 8
ARIMA(1,1,0)	1 974.062 6	1 985.367 5	1 978.576 9
ARIMA(1,1,1)	1 976.032 6	1 991.105 8	1 982.051 6
ARIMA(2,1,1)	1 974.968 6	1 993.810 2	1 982.492 4
ARIMA(2,1,2)	1 979.271 4	2 001.814 0	1 988.300 0

2.3.3 残差检验

模型定阶后,对残差进行白噪声检验。若残差满足白噪声序列,则模型拟合有效。检验结果见图 10 和图 11。从图 10 和图 11 可以看出:拟合残差 ACF、PACF 是截尾,其残差序列为白噪声;KDE 线与  $N(0,1)$  平行,残差呈正态分布(见图 12)。

ARIMA(1,1,0)为有效模型。

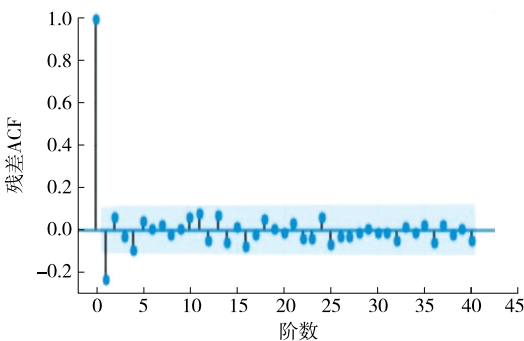


图 10 拟合后残差 ACF

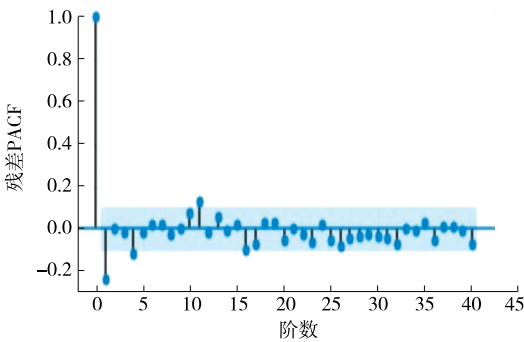


图 11 拟合后残差 PACF

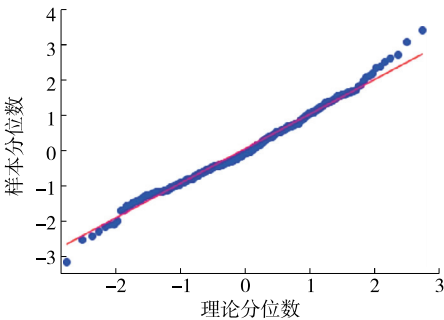


图 12 残差 QQ 图

2.3.4 模型预测及效果评价

用标定并通过残差检验的 ARIMA(1,1,0)模型进行预测,并与 12 日、13 日实际值进行对比,结果见图 13、图 14。

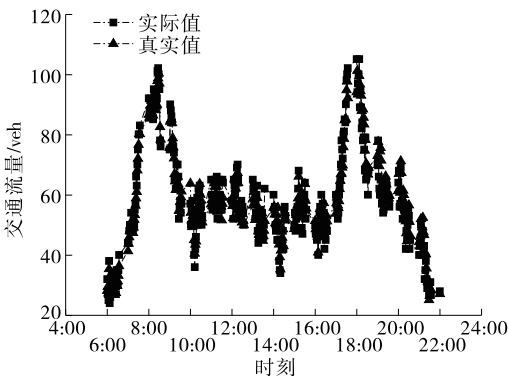


图 13 12 日交通流趋势图

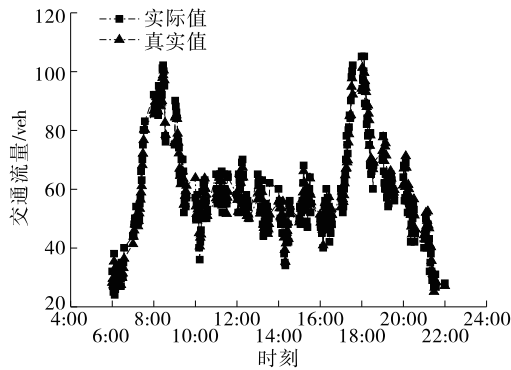


图 14 13 日交通流趋势图

由图 13、图 14 可知:预测值的变化规律与实际值具有较好的一致性。但由于交通流的不确定性、非线性特征,在交通流波动较大的时段拟合效果较差。

为进一步验证模型的预测性能,引入  $\delta_{MAE}$ 、 $\delta_{MAPE}$  和  $\delta_{RMSE}$  进行评价。计算公式见式(4)~(6),计算结果见表 3。

$$\delta_{MAE} = \frac{1}{l} \sum_{i=1}^l |Q_i - u_i| \times 100 \quad (4)$$

$$\delta_{MAPE} = \frac{1}{l} \sum_{i=1}^l \left| \frac{Q_i - u_i}{Q_i} \right| \times 100 \quad (5)$$

$$\delta_{RMSE} = \sqrt{\frac{1}{l} \times \sum_{i=1}^l \left( \frac{Q_i - u_i}{Q_i} \right)^2} \times 100 \quad (6)$$

式中: $l$  为观测次数; $Q_i$  为真实交通流量; $u_i$  为预测交通流量。

表 3 ARIMA 模型预测精度验算结果 单位: %

时间	$\delta_{MAE}$	$\delta_{MAPE}$	$\delta_{RMSE}$
12 日	43.8	7.99	5.62
13 日	48.8	8.80	6.04

由表 3 可知:12 日、13 日交叉口交通流预测值的 3 个评价指标均小于 10%,说明 ARIMA(1,1,0) 模型具有较好的预测效果和较强的适用性; $\delta_{MAE}$  为 43.8%、48.8%,模型预测精度还有待提高。

### 3 结语

本文通过 Python 可视化工具及 ARIMA 模型构建能实时、在线进行交叉口短时交通流可视化预测的系统。该系统将交通流时序以非量化、直观的形式进行处理,实现交通流时间规律的可视化。ARIMA 模型有效提取了交通流时序特征,预测精度较高,对于高速公路、城市快速路等场景下短时交通流预测具有可移植性。未来可通过深度学习进行实验与分析,对不同车道的交通流特性进行深入分析,进一步精细化研究交通流预测。

### 参考文献:

- [1] 赵宏,翟冬梅,石朝辉.短时交通流预测模型综述[J].都市快轨交通,2019,32(4):50-54.
- [2] 张伟斌,张帅,郭海峰,等.基于交通因子状态网络的城市交叉口交通流预测[J].中国公路学报,2021,34(12):217-228.
- [3] 张艺铭,陈明明,石磊,等.基于 IGWO-BP 算法的轨道交通短时客流预测[J].交通信息与安全,2021,39(3):85-92.
- [4] 田保慧,郭彬.基于时空特征分析的短时交通流预测模型[J].重庆交通大学学报(自然科学版),2016,35(3):105-109+182.
- [5] 陈喜群,周凌霄,曹震.基于图卷积网络的路网短时交通流预测研究[J].交通运输系统工程与信息,2020,20(4):49-55.
- [6] HOU Q Z, LENG J Q, MA G S, et al. An adaptive hybrid model for short-term urban traffic flow prediction [J]. Physica A: Statistical Mechanics and Its Applications, 2019, 527: 121065.
- [7] 钟颖,邵毅明,吴文文,等.基于 XGBoost 的短时交通流预测模型[J].科学技术与工程,2019,19(30):337-342.

收稿日期:2022-04-08

\*\*\*\*\*

(上接第 4 页)

- [7] 黄超智,龙军,周金应,等.基于 JT/T 1242-2019 的 AEB 测试研究[J].公路与汽运,2020(3):1-6.
- [8] 辛佳庚,杨复钰,张宝迪,等.基于融合算法的电动汽车 AEB 控制策略[J].北京交通大学学报,2021,45(6):77-86+93.
- [9] 石晏丞,李军.汽车自动驾驶领域的传感器融合技术[J].装备机械,2021(3):1-6+12.

- [10] 牛成勇,苏占领,吴昆仑,等.危险品运输车主动防撞系统性能测试分析[J].华东交通大学学报,2019,36(4):32-38.
- [11] 牛成勇,曾杰,徐建勋,等.汽车装备自动紧急制动系统后的性能测试评价[J].河北工业大学学报,2018,47(5):75-81.

收稿日期:2022-08-16