

DOI:10.20035/j.issn.1671-2668.2023.04.005

# 考虑交通事故时空效应的高速公路交通事故 严重程度预测

吴楚恩, 罗世奎, 巫立明, 许佳乐

(广东联合电子服务股份有限公司, 广东 广州 510075)

**摘要:**为提高高速公路交通事故处置效率,降低交通事故造成的生命财产损失,提出一种基于卷积神经网络(CNN)的高速公路交通事故严重程度预测方法。首先确定交通事故严重程度的影响因素,并根据交通事故损失情况和对事故路段的影响将事故划分为轻微事故、一般事故和严重事故三类;然后采用梯度提升决策树法计算各影响因素对交通事故严重程度影响的权重等级,根据权重等级将影响因素特征矩阵转化为灰度图像,即将交通事故数据的单一特征关系转换为包含并行组合关系的方形特征矩阵;以事故数据的时空特性作为输入,构建基于CNN的高速公路交通事故严重程度预测模型,并对模型进行评估与检验。结果表明,在高速公路交通安全事故严重程度评估中,相较于随机森林机器模型和逻辑回归模型,文中模型的识别率明显提高,对严重交通事故的测试准确率超过80%。

**关键词:**交通安全;交通事故;高速公路;严重程度预测;方形特征矩阵;卷积神经网络(CNN)

中图分类号:U491.3

文献标志码:A

文章编号:1671-2668(2023)04-0022-06

近年来,随着高速公路的不断发展,高速公路面临的交通安全压力不断增大,交通事故数量居高不下,交通事故已成为造成人身伤害和经济损失的重大威胁<sup>[1]</sup>。对高速公路交通事故严重程度进行预测可为合理制定交通事故应急处置策略提供参考。

目前交通事故严重程度预测方法主要有统计模型方法和机器学习算法两类。李庚凭构建了基于有序Logit模型和多项Logit模型的道路交通事故严重程度预测模型,检验结果表明多项Logit模型的性能好于有序Logit模型<sup>[2]</sup>;陈昭明等构建混合Logit模型分析高速公路交通事故严重程度,检验结果表明考虑参数间相关性的混合Logit模型比多项Logit模型具有更好的拟合优度,且能更合理地反映各因素对事故严重程度的作用效果<sup>[3]</sup>。然而,交通事故是人、车、路和环境相互作用的结果,交通事故数据具有高维度、非线性和多重共线性的特点,由若干个特征组成<sup>[4]</sup>。统计模型通常对解释变量和响应变量有严格的假设,而且不能全面考虑数据的特征,其预测性能弱于机器学习算法。孙维富等采用支持向量机理论建立高速公路交通事故严重程度分类识别模型,多角度、多层次地研究了高速公路交通事故与驾驶员、车辆、道路、环境等因素的关系<sup>[5]</sup>;童璐璐应用贝叶斯网络和参数学习算法,利用因树

传播算法构建了基于贝叶斯网络的高速公路交通事故预测模型<sup>[6]</sup>;Iveta M.等对比分析了随机森林模型、前馈神经网络模型和统计模型对交通事故严重程度的预测精度,结果显示机器学习算法具有更高的预测精度<sup>[7]</sup>;Chen X. N.提出一种卷积神经网络(CNN)预测模型,并与决策树分类预测模型、Logit回归预测模型和支持向量机预测模型进行对比分析,结果表明基于CNN的预测模型具有更好的性能<sup>[8]</sup>;吕璞等考虑到深度卷积神经网络可以高效处理图像问题,将事故影响因素图像化,提出一种融合深度反残差与注意力机制的山区高速公路交通事故严重程度预测模型,验证结果表明相比传统的机器学习模型,该模型的识别准确率明显提高<sup>[9]</sup>。但现有机器学习模型只是简单地输入事故数据特征,未考虑数据特征之间的关联特性,同时构建数据集时忽略了交通的时空效应对事故严重程度的影响,不利于模型预测性能的提高。

基于以上分析,本文根据广东省某高速公路近2年的交通事故数据,以交通事故严重程度作为因变量,考虑事故数据的时空特性,以事故发生前30 min的车速、密度、交通量等短时交通数据和道路、环境等参数作为自变量,根据影响因素权重等级将事故影响因素特征转化为包含并行组合关系的方形数据

集,考虑 CNN 在提取高维度特征方面的优势,提出一种基于 CNN 的考虑交通事故时空效应的高速公路交通事故严重程度预测模型,实现高速公路交通事故严重程度预测。

## 1 高速公路交通事故影响因素选择

### 1.1 数据采集

采集广东省某高速公路近 2 年的交通事故统计数据,包括发生时间、发生季节、事故类型、涉事车

型、事故路段车道数、事故路段非直线率、事故路段建筑设施、死亡人数、受伤人数及事故发生前 30 min 的交通流数据(平均车速、交通量、密度、大车比等)。其中事故路段、事故发生时间、事故类型、伤亡情况等数据在事故发生现场采集获取,事故发生前的交通流数据通过高速公路门架信息获取,表 1 为高速公路门架提取的事故发生前交通流状态数据样例。对事故数据进行整理、筛选和预处理,得到本文所需事故数据。

表 1 事故发生前 30 min 的交通流状态数据样例

事故发生前时间区间/min	平均车速/(km·h <sup>-1</sup> )	交通量/(辆·h <sup>-1</sup> )	交通密度/(辆·km <sup>-1</sup> )	大型车比例
0~5	95.62	220.5	5.204 460 967	0.040 000 000
6~10	93.88	196.5	4.237 918 216	0.069 114 471
11~15	93.77	243.0	5.377 942 999	0.062 906 725
16~20	96.39	251.5	5.650 557 621	0.048 016 701
21~25	93.73	237.5	5.303 593 556	0.082 105 263
26~30	96.95	225.0	5.377 942 999	0.057 939 914

### 1.2 交通事故严重程度分类

分析事故占据车道数与事故数据的关系,发现占据不同车道数时事故数据差别明显,故根据事故对高速公路的影响即事故占据车道数、伤亡人数将事故划分为轻微事故、一般事故和严重事故三类(见表 2)。

表 2 交通事故严重程度分类

事故类型	事故占据车道数/条	受伤人数/人	死亡人数/人
轻微事故	1	0	0
一般事故	2	1~2	0
严重事故	>2	>2	>0

共采集该高速公路 441 起事故数据,其中轻微事故 276 起、一般事故 81 起、严重事故 84 起。将事故数据集按照 7:3 的规则划分为训练集和测试集,

训练集中轻微事故、一般事故和严重事故的数据量分别为 193 起、56 起、58 起,测试集中轻微事故、一般事故和严重事故的数据量分别为 83 起、25 起、26 起。

### 1.3 交通事故严重程度影响因素

高速公路交通事故是由多种因素共同作用的结果。考虑高速公路上车辆运行特征,选取可能对事故严重程度造成显著影响的车、路、环境、事故特征方面共计 12 个因素<sup>[10-15]</sup>。其中:车对交通事故严重程度的影响因素主要包括事故发生路段的动态交通流情况和车辆类型<sup>[16]</sup>;道路对交通事故严重程度的影响因素包括事故路段的长度、车道数、线形、类型;环境对交通事故严重程度的影响因素包括事故发生时段、季节;事故特征对交通事故严重程度的影响因素主要为事故类型和伤亡人数(见表 3)。

表 3 高速公路交通事故严重程度影响因素

因素类型	影响因素	说明
车	车速	事故发生前路段的平均车速
	交通量	事故路段到达交通量和离去交通量的均值
	交通密度	事故发生前路段的平均交通密度
	车辆类型	大型车比例
路	道路长度	事故地点上下游高速公路门架的行程距离
	车道数	事故发生路段的车道数
	道路线形	以事故路段非直线率表示,为事故地点上下游高速公路门架的直线距离与行程距离的比值
	道路类型	以事故路段的建筑设施表示,有收费站时取 1,有服务区时取 2,有收费站和服务区时取 3

续表3

因素类型	影响因素	说明
环境	事故发生时段	事故发生所处时段,0:00—12:00取0,12:00—24:00取1
	事故发生季节	事故发生所处季节,春季取0,夏季取1,秋季取2,冬季取3
事故特征	事故类型	碰撞事故取0;追尾事故取1;翻车事故取2
	事故伤亡人数	事故造成的伤亡人数

1.4 影响因素权重计算

以事故严重程度为因变量,以轻微事故、一般事故和严重事故为输出变量,以表3中12个影响因素为自变量即输入变量。对已知事故数据进行统计分析,计算各影响因素所占比例,利用梯度提升决策树法<sup>[17]</sup>对影响因素的权重等级进行计算并排序,父权重等级用大写字母A~D降序表示,子权重等级用小写字母a~l降序表示(见表4)。

表4 高速公路交通事故严重程度影响因素取值及权重等级

因素类型	父权重等级	影响因素	子权重等级
车	A	车速	a
		交通量	b
		交通密度	c
		车辆类型	d
路	C	道路长度	j
		车道数	h
		道路线形	g
		道路类型	i
环境	D	事故发生时段	k
		事故发生季节	l
事故特征	B	事故类型	e
		事故伤亡人数	f

2 交通事故严重程度预测模型

针对传统的机器学习算法未考虑数据之间组合

关系及数据时空特征的缺陷,结合CNN的特点,将数据之间的组合关系联系起来,考虑交通事故时空特性,构建图1所示基于CNN的高速公路交通事故严重程度预测模型。

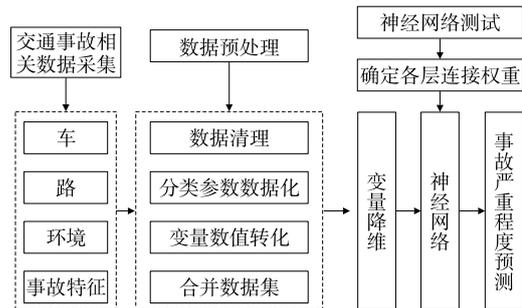


图1 基于交通事故时空效应的高速公路事故严重程度预测模型

2.1 影响因素的图像化

如图2所示,数据输入为5×5的方形矩阵,3×3的矩阵为卷积核,卷积步长为2,数据输出为2×2的矩阵,其中特征13共参与4次卷积运算,特征1、5、21、25只参与一次卷积运算,越靠近矩阵中心的特征卷积运算次数越多,即矩阵的中心位置比边缘位置更有助于从CNN中提取特征信息。考虑到特征的权重等级越高,对预测结果的影响越大,将权重等级较高的特征填充到全零矩阵的中心位置、权重等级较低的特征填充到全零矩阵的边缘位置,以提高预测模型的性能和精度。

依据表4中各数据特征的权重等级,结合特征

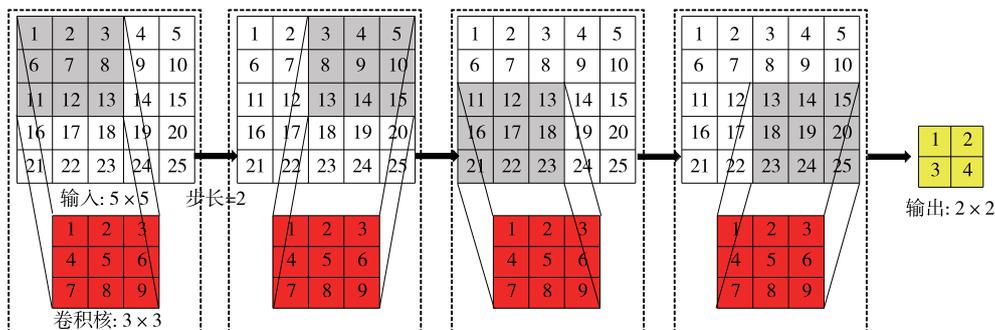


图2 卷积运算过程

权重将特征填充到全零矩阵中,最后转换为方形特征矩阵即灰度图作为 CNN 的输入。将单个特征关系转换为包含数据特征组合关系的方形矩阵的步骤:1) 高速公路交通事故严重程度有 4 个父特征、12 个子特征,根据父特征及对应子特征的数量建立一个 4×4 的全零矩阵;2) 根据最大权重在中心且上权重大于下权重的规则<sup>[18]</sup>将所有父特征按降序排列,所有父特征沿纵轴由全零矩阵的中心向上下

两侧逐渐填充;3) 根据最大权重位于中心且左权重大于右权重的规则<sup>[18]</sup>将所有子特征由全零矩阵的中心沿水平轴向左右两侧逐渐填充;4) 通过调用图形处理函数和扩展矩阵形状的方式将矩阵转换为灰度图像(见图 3),得到一组数据的灰度图像后,使用并行方法将数据集中 441 组数据的特征矩阵并行转换为方形特征矩阵。

另外,考虑到事故数据的时空特性,将车辆的交



图 3 生成方形特征矩阵的流程

通流数据以 5 min 为间隔划分为 6 个时段,即将事故发生前 30 min 至事故发生时的时间划分为 6 个时间切片,并以时间切片作为模型的输入高度即模型的通道数进行模型输入,模型输入的维度为 4×4×6。

## 2.2 卷积神经网络

CNN 包括输入层、卷积层、池化层、全连接层和输出层,其中卷积层、池化层、全连接层属于神经网络的隐藏层(见图 4)。

CNN 模型的输入层为 4×4×6 的图像,隐藏层

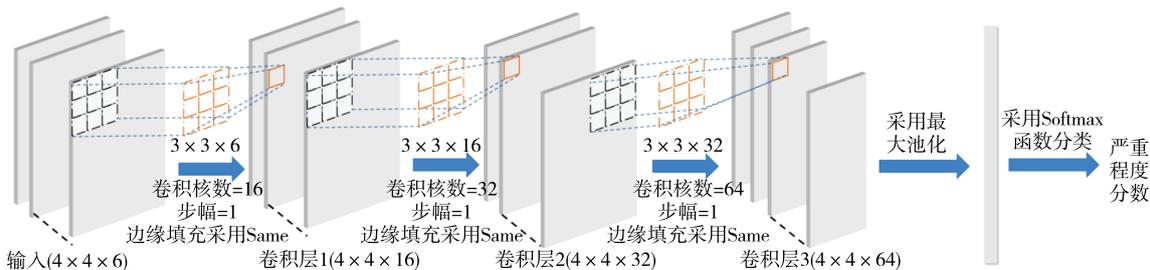


图 4 卷积神经网络模型的结构

包括 3 个卷积层、1 个池化层和 2 个全连接层。卷积层是 CNN 的核心,其目的是提取交通事故数据

集中的抽象特征,CNN 的卷积层为局部连接,不是完全连接,即输出神经元只与局部相邻的输入神经

元相连。为清楚地描述卷积层的计算过程,将每个方形特征矩阵的每个像素编号为 $P_{c,i,j}$ (代表 $c$ 通道图像第 $i$ 行、第 $j$ 列中的像素元素), $c$ 通道滤波器的 $u$ 行和 $v$ 列的权重用 $w_{c,u,v}$ 表示,按式(1)进行卷积运算。池化层一般放在卷积层之后,可以对数据进行压缩,降低维度,减少参数量,防止出现过拟合现象。全连接层将最终特征转换为一维向量,方便输出。模型输出层使用 Softmax 激活函数对交通事

故严重程度进行分类。

$$r_{i,j} = f\left(\sum_{c=1}^C \sum_{u=1}^F \sum_{v=1}^F w_{c,u,v} P_{c,i+u,j+v}\right) \quad (1)$$

式中: $r_{i,j}$ 表示卷积计算中特征矩阵第 $i$ 行、第 $j$ 列元素; $C$ 为滤波器通道数; $F$ 为卷积核维度。

### 3 模型评估结果与检验

#### 3.1 交通事故严重程度预测模型的超参数(见表5)

表5 交通事故严重程度预测模型的超参数

网络层	输入	输出	通道数	卷积核数	卷积核	步长
卷积层1	$4 \times 4 \times 6$	$4 \times 4 \times 16$	6	16	$3 \times 3$	$1 \times 1$
卷积层2	$4 \times 4 \times 16$	$4 \times 4 \times 32$	16	32	$3 \times 3$	$1 \times 1$
卷积层3	$4 \times 4 \times 32$	$4 \times 4 \times 64$	32	64	$3 \times 3$	$1 \times 1$
池化层	$4 \times 4 \times 64$	$2 \times 2 \times 64$	64	64	$2 \times 2$	$2 \times 2$

#### 3.2 模型的评价指标

采用深度学习通常采用的准确率( $R_{\text{accuracy}}$ )、精确率( $R_{\text{precision}}$ )、召回率( $R_{\text{recall}}$ )对高速公路交通事故严重程度预测模型的性能进行评价,计算公式如下:

$$R_{\text{accuracy}} = \frac{Q_{\text{TP}} + Q_{\text{TN}}}{Q_{\text{TP}} + Q_{\text{FP}} + Q_{\text{TN}} + Q_{\text{FN}}} \quad (2)$$

$$R_{\text{precision}} = \frac{Q_{\text{TP}}}{Q_{\text{TP}} + Q_{\text{FP}}} \quad (3)$$

$$R_{\text{recall}} = \frac{Q_{\text{TP}}}{Q_{\text{TP}} + Q_{\text{FN}}} \quad (4)$$

式中: $Q_{\text{TP}}$ 为将正例预测为正例的数量; $Q_{\text{TN}}$ 为将负例预测为负例的数量; $Q_{\text{FP}}$ 为将负例预测为正例的数量; $Q_{\text{FN}}$ 为将正例预测为负例的数量。

#### 3.3 模型的预测结果

在 Tensorflow2.1-keras 环境中运行高速公路交通事故严重程度预测模型,设置模型的迭代次数为100次、一次抓取的样本量为16个、学习率为0.0001。模型训练和测试结果的损失函数、准确率、召回率见图5~7。从图5~7可以看出:经过

100次迭代后,基于CNN的高速公路交通事故严重程度预测模型的测试准确率为80%左右,测试召回率为80%左右,预测性能较好。

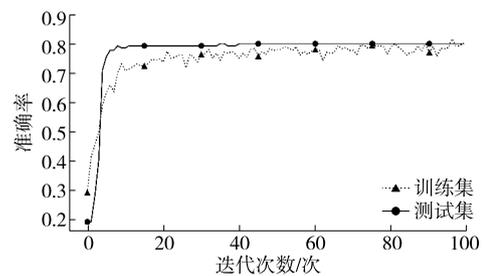


图6 高速公路交通事故严重程度预测模型训练与测试结果的准确率

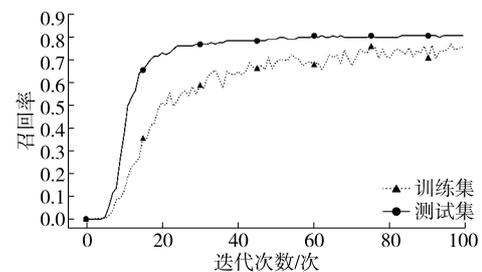


图7 高速公路交通事故严重程度预测模型训练与测试结果的召回率

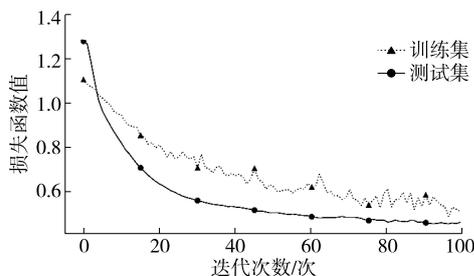


图5 高速公路交通事故严重程度预测模型训练与测试结果的损失函数

基于CNN的考虑交通事故时空效应的高速公路交通事故严重程度预测模型预测结果的混淆矩阵见图8,依据混淆矩阵得出不同严重程度交通事故预测结果的召回率和精确率(见表6)。从图8、表6可以看出:对于严重事故,该模型的预测精度(召回率)为88.46%,超过80%;对于一般事故,其预测精

度为 72%。该模型对一般事故的识别率较低,对严重事故的识别率较高。

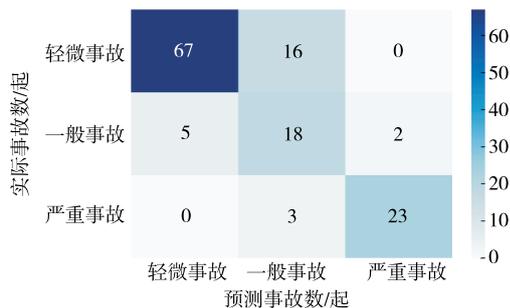


图 8 交通事故严重程度预测模型测试结果的混淆矩阵

表 6 不同严重程度交通事故预测结果的精确率和召回率

事故类型	精确率/%	召回率/%	事故类型	精确率/%	召回率/%
轻微事故	93.06	80.72	严重事故	92.00	88.46
一般事故	48.65	72.00			
平均				77.90	80.39

### 3.4 不同预测模型的比较

为了验证基于 CNN 的高速公路交通事故严重程度预测模型(以下简称 CNN 模型)对交通事故严重程度的预测性能,将该模型与随机森林模型、逻辑回归模型进行对比分析,3 种模型的准确率、召回率、精确率见表 7。

表 7 不同预测模型的性能对比

预测模型	准确率/%	召回率/%	精确率/%
CNN 模型	80.60	80.39	77.90
随机森林模型	78.36	78.12	77.46
逻辑回归模型	76.12	76.25	76.08

从表 7 可以看出:3 种模型中,CNN 模型的准确率、召回率和精确率最高,准确率比随机森林模型、逻辑回归模型分别提高 2.24%、4.48%。这可能是因为统计模型处理交通事故数据时忽略了交通事故数据集特征之间的组合关系,机器学习模型无法从模型结构的角度分析交通事故数据集特征之间的组合关系,而 CNN 模型充分提取了交通事故数据集特征之间的组合关系。

## 4 结语

本文考虑交通事故严重程度各影响因素之间的组合关系,采用方形特征矩阵表征影响因素的特征

组合和特征相关性,结合交通事故数据的时空特性,建立基于 CNN 的高速公路交通事故严重程度预测模型,经过训练后,该模型能有效区分轻微事故、一般事故和严重事故。与随机森林模型、逻辑回归模型相比,该模型的准确率分别提升 2.24%、4.48%,可满足对交通事故严重程度的预测要求。但该模型对轻微事故和一般事故的分类存在一定局限,原因可能是数据样本量较少,轻微事故和一般事故的特征差异性不明显,要提升该模型的预测精度,可能需要更大的数据集参与训练,且该模型未考虑人的因素对交通事故严重程度的影响,有待改进。另外,仅将 CNN 模型与随机森林模型和逻辑回归模型进行比较,未能全面地对比其他机器学习算法和统计模型,模型性能还有待进一步验证。

### 参考文献:

- [1] PEDEN M M, KHAYESI M. Save LIVES technical package: 22 interventions that could make a difference[J]. Injury Prevention, 2018, 24(5): 381-383.
- [2] 李庚凭.基于有序 Logit 和多项 Logit 模型的高速公路交通事故严重程度预测[D].西安:长安大学,2018.
- [3] 陈昭明,徐文远,曲悠扬,等.基于混合 Logit 模型的高速公路交通事故严重程度分析[J].交通信息与安全, 2019, 37(3): 42-50.
- [4] ZONG F, ZHANG H Y, XU H G, et al. Predicting severity and duration of road traffic accident[J]. Mathematical Problems in Engineering, 2013, 2013: 1-9.
- [5] 孙维富,席建锋,郭虹宇.高速公路交通事故严重程度影响因素分析[J].黑龙江交通科技, 2018, 41(5): 200-201.
- [6] 童璐璐.基于贝叶斯网络的高速公路交通事故严重程度预测研究[D].北京:北京交通大学,2018.
- [7] IVETA M, RADOVAN A, MIHALJEVIĆ B. Prediction of traffic accidents severity based on machine learning and multiclass classification model[C]//2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO). IEEE, 2021: 1701-1705.
- [8] CHEN X N. Traffic crash severity prediction with deep learning[J]. Journal of Physics: Conference Series, 2021, 1883(1): 012141.
- [9] 吕璞,柏强,陈琳.融合深度反残差与注意力机制的山区高速公路事故严重程度预测模型[J].中国公路学报, 2021, 34(6): 205-213.

(下转第 34 页)

盔等措施,降低事故的严重程度。 $V_{1,1,1}$ 、 $B_{1,1}$ 、 $B_{1,3}$ 属于中等风险影响、中等风险概率的事故类型,预防这类事故应同时降低其发生概率和严重程度,可采取设置机动车-非机动车分离护栏、交叉口进口道拓宽、设置交通信号等方式减少车辆间的冲突。

(3) 信号控制交叉口 3 种需要关注的事故类型中,除  $B_{7,1}$  事故的发生机理与无信号控制交叉口相似外, $P_1$ 、 $B_{1,3}$  事故大多由事故中的一方未遵守信号规则而引发,防范此类事故,一方面可通过调整信号时间适配过街习惯,另一方面可通过查处、教育、警示等手段提高交通参与者的信号遵守率。

参考文献:

[1] 王雪松,罗棟.信号控制交叉口交通冲突特征与影响因素[J].城市交通,2016,14(6):60-66.  
 [2] 王玉全,邢芳,郭伟伟.信号交叉口混合交通冲突研究[J].中国安全科学学报,2016,26(6):47-51.  
 [3] 巫诚诚,陈大伟.交叉口非机动车冲突易发点空间预测模型[J].中国安全科学学报,2021,31(8):165-171.  
 [4] 刘辉,章国鹏,王羿童,等.无控制交叉口交通安全影响因素分析[J].公路交通科技,2019,36(9):110-116.  
 [5] 孙重静.城市交叉口行人与右转车冲突风险评估[J].交通与运输(学术版),2018(2):64-67.  
 [6] 周智文,马健霄,王冠森.信号交叉口左转非机动车交通冲突分析[J].合肥工业大学学报(自然科学版),2021,44(1):95-100.  
 [7] 于瑞康.信号交叉口右转机动车与非机动车冲突行为分析[D].北京:北京交通大学,2018.

[8] 张开冉,邱谦谦.平面交叉口电动自行车交通冲突模型研究[J].安全与环境学报,2017,17(2):422-425.  
 [9] YAN X D,RADWAN E,ABDEL-ATY M.Characteristics of rear-end accidents at signalized intersections using multiple logistic regression model[J].Accident Analysis & Prevention,2005,37(6):983-995.  
 [10] MITRA S,CHIN H C,QUDDUS M A.Study of intersection accidents by maneuver type[J].Transportation Research Record:Journal of the Transportation Research Board,2002,1784(1):43-50.  
 [11] 谢晓莉,刘洪启,李平生.无信号控制公路交叉口行人死亡事故特征及机理[J].中国安全科学学报,2013,23(8):96-101.  
 [12] 谷旭佳.交叉口人车事故风险及伤害程度统计分析方法[D].北京:北京交通大学,2021.  
 [13] 张一驰.城市道路交叉口货运车辆交通事故成因分析及预防对策[D].长春:吉林大学,2016.  
 [14] YANG N,LI Y,LIU T,et al.Analysis of fatal factors influencing accidents involving two-wheel electric vehicle drivers at intersections [J]. Legal Medicine, 2020,45:101696.  
 [15] 罗聪,徐克,刘潜,等.安全风险分级管控相关概念辨析[J].中国安全科学学报,2019,29(10):43-50.  
 [16] 朱启超,匡兴华,沈永平.风险矩阵方法与应用述评[J].中国工程科学,2003,5(1):89-94.  
 [17] 高惠璇.应用多元统计分析[M].北京:北京大学出版社,2005:252-258.

收稿日期:2022-09-28

\*\*\*\*\*

(上接第 27 页)

[10] 刘鑫,肖乐,金思雨.基于 NAIS 的高速公路严重交通事故特征及致因分析[J].公路与汽运,2022(5):32-37.  
 [11] 王文成.考虑时空效应的道路交通事故建模及其影响因素分析[D].北京:北京交通大学,2020.  
 [12] 由冰玉.高速公路交通事故影响因素研究[D].哈尔滨:哈尔滨工业大学,2020.  
 [13] 魏凌峰,姜文龙.基于支持向量机的交通事故影响因素分析[J].山东交通科技,2022(1):84-87.  
 [14] 孙轶轩,邵春福,岳昊,等.基于 SVM 灵敏度的城市交通事故严重程度影响因素分析[J].吉林大学学报(工学版),2014,44(5):1315-1320.

[15] 杨文臣,周燕宁,田毕江,等.基于聚类分析和 SVM 的二级公路交通事故严重度预测[J].中国安全科学学报,2022,32(5):163-169.  
 [16] 侯树展,孙小端,贺玉龙,等.高速公路交通事故严重程度与交通流特征的关系研究[J].中国安全科学学报,2011,21(9):106-112.  
 [17] 何泊延.基于 GBDT 的数据分类方法的研究及其应用[D].吉林:东北电力大学,2021.  
 [18] ZHENG M,LI T,ZHU R,et al.Traffic accident's severity prediction: A deep-learning approach-based CNN network[J].IEEE Access,2019,7:39897-39910.

收稿日期:2022-09-09