

文章编号:1005-0523(2024)05-0065-09



道路交通事故严重程度预测及致因分析

严利鑫¹, 胡鑫辉¹, 刘清梅², 金武杰¹

(1. 华东交通大学交通运输工程学院, 江西 南昌 330013; 2. 南昌交通学院交通运输学院, 江西 南昌 330044)

摘要:【目的】为提高交通事故严重程度预测的准确性,明晰事故严重程度的关键影响因素。【方法】从885起道路交通事故案例数据中选取影响交通事故严重程度的人、车、路、环境四方面共14个因素,采用融合通道注意力的卷积神经网络(CA-CNN)构建事故分类预测模型。在此基础上,采用边际效用方法分析得出交通事故严重程度的显著影响因素。【结果】结果表明,相较于卷积神经网络(CNN)、随机森林(RF)、朴素贝叶斯(NaiveBayes)、回归分析(Logistics)、决策表(Decision_table)、引导聚集算法(Bagging)6种预测模型,CA-CNN模型在准确率、查准率、召回率等指标评价下,整体预测性能更优;在交通事故严重程度的影响因素中,季节、是否工作日、道路类型、事故形态、是否违法变更车道、未按规定让行及制动不当7个因素具有显著性($p \leq 0.05$)。【结论】CA-CNN是一种有效的交通事故严重程度预测模型,其分析结果有助于降低交通事故发生率和严重程度。

关键词:交通安全;事故严重程度预测;卷积神经网络;通道注意力;致因分析

中图分类号:U491.31

文献标志码:A

本文引用格式:严利鑫,胡鑫辉,刘清梅,等. 道路交通事故严重程度预测及致因分析[J]. 华东交通大学学报, 2024, 41(5): 65-73.

Road Traffic Accident Severity Prediction and Causation Analysis

Yan Lixin¹, Hu Xinhui¹, Liu Qingmei², Jin Wujie¹

(1. School of Transportation Engineering, East China Jiaotong University, Nanchang 330013, China;

2. College of Transportation, Nanchang Jiaotong Institute, Nanchang 330044, China)

Abstract: 【Objective】To improve the accuracy of predicting the severity of traffic accidents and clarify the key influencing factors of traffic accident severity. 【Method】Fourteen factors related to people, vehicles, roads, and environment that affect traffic accident severity were selected from 885 road traffic accident case data, and a traffic accident classification prediction model was constructed using a convolutional neural network with channel attention (CA-CNN). On this basis, the significant influencing factors of traffic accident severity were analyzed using the method of marginal utility. 【Result】The results show that compared with the 6 prediction models such as convolutional neural network (CNN), random forest (RF), NaiveBayes, logistics regression (Logistics), decision table (Decision_table), and bagging algorithm (Bagging), the convolutional neural network model with channel attention fusion has better overall prediction performance in terms of accuracy, precision, and recall, etc. Among the influencing factors of traffic accident severity, the season, whether it is a workday, road type, accident type, illegal lane change, failure to yield, and improper braking have significant effects ($p \leq 0.05$). 【Conclusion】CA-CNN is an effective

收稿日期:2023-02-28

基金项目:国家自然科学基金项目(52162049, 51805169);赣鄱俊才支持计划-主要学科学术和技术带头人培养项目(20232BCJ23012);华东交通大学创新创业教育研究课题(23hict05)

tive traffic accident severity prediction model, and the analysis results are helpful in reducing the incidence and severity of traffic accidents.

Key words: traffic safety; accident severity prediction; convolutional neural network; channel attention; causation analysis

Citation format: YAN L X, HU X H, LIU Q M, et al. Road traffic accident severity prediction and causation analysis[J]. Journal of East China Jiaotong University, 2024, 41(5): 65–73.

【研究意义】根据世界卫生组织发布的《2023年全球道路安全状况报告》，2021年全球约有119万人因道路交通事故死亡，相当于每10万人中有15人死于道路交通事故，道路交通事故是造成5至29岁年龄段人员死亡的主要原因之一^[1]。近年来，随着我国经济社会的快速发展和城镇化水平的不断提高，道路交通需求急剧扩张，这无疑对日益严峻的交通安全形势构成了更大挑战。因此，提高交通事故严重程度预测的准确性，明晰事故严重程度的关键影响因素，对提高道路安全水平具有重要意义。

【研究进展】在早期的研究中，受计算机运算能力的限制，研究人员主要采用线性回归、负二项式、零膨胀回归等简单耗时的方法来构建道路交通事故预测模型^[2-3]；然而，这类方法未能充分反映出道路交通事故的分布特点，模型预测准确率不高。Sun等^[4]用动态贝叶斯网络模型探索事故发生前车辆的速度数据与事故发生的关系，实现了76.4%的预测精度。孙轶轩等^[5]采用C5.0决策树方法构建交通事故严重程度预测模型，以二分类和三分类作为因变量进行预测，预测准确率分别为70%和61%。Parsa等^[6]采用概率神经网络(PNN)构建模型用以监测交通事故的发生，该模型在正确预测事故方面的表现优于支持向量机(SVM)。Vaiyapuri等^[7]比较了多层感知机、逻辑回归和K最近邻算法在事故严重程度预测上的性能，发现多层感知机的预测性能优于其他算法。综合来看，在评估某种事故严重程度预测模型的预测精度时，通常会将其与其他多种模型进行对比分析，以凸显出该模型的优势。

在交通事故严重程度致因分析领域的研究中，李英帅等^[8]基于随机森林算法构建了一个预测模型，该模型将22个特征因素纳入考虑以预测事故严重程度；研究发现，车辆间事故类型、受伤部位、道路物理隔离类型等因素是影响受伤严重程度的主

要因素。石小林等^[9]基于美国密歇根州Wayne县的直角事故记录数据，通过二元Logit模型分析了影响直角事故伤害严重性的因素；Jalayer等^[10]采用相关随机参数有序Probit模型，探讨了驾驶员错误驾驶行为对交通事故严重性的影响；韩天园等^[11]运用文本挖掘技术提取了重特大道路交通事故调查报告中的事故成因特征；以上研究结果显示驾驶员违法、安全隐患和操作不当是导致重大交通事故的主要原因。目前，对于涉及多种因素和分类的交通事故严重性的研究尚需进一步深化，此前大多研究主要集中在二分类事故严重程度，这也会限制事故数据中有效信息的提取。

【创新特色】在目前关于交通事故严重程度的研究中，研究数据主要来自高速公路，对日常市区通勤发生的交通事故研究较少，为了填补这项空白，本文选取江苏某园区城市道路交通事故数据。机器学习方法已经在道路交通事故的预测中显示良好的应用前景，本文采用改进卷积神经网络，得到模型精度更高、泛化能力更强的交通事故严重程度预测模型。

【关键问题】深入研究交通事故与涉及的人、车辆、道路条件、环境等各要素之间的复杂关系，对于精确预测事故的严重程度至关重要，找出对事故严重程度有显著影响的因素，对提升道路安全水平，改善交通管理策略，促进交通安全政策的制定与执行，推动交通安全技术的进步，增强公众对交通安全重要性的认识具有重要的应用价值。

1 道路交通事故数据

依据公安部2017年颁布的《道路交通事故处理程序规定》，道路交通事故按事故后果严重程度可以分为：死亡事故、伤人事故和财产损失事故。本研究选用苏州某园区一年的交通事故数据进行挖掘分析，经过预处理得到有效数据885条。结合数

据样本特征,将交通事故严重程度等级划分为:轻微事故、一般事故、重大事故3个不同等级。其中轻微事故指的是造成经济财产损失但事故人员不需要医疗干预的事故;一般事故是指在事故中相关人员涉及骨折等较为严重的伤情,需要医疗干预的事故;重大事故是指事故发生导致事故相关人员死亡或1个月内重伤不治的事故。

每条交通事故记录包含3种类型的信息:①事故原因;②事故特征;③环境特征;事故原因又包括驾驶员、车辆、环境3个维度的信息;事故特征包括所涉车辆种类,事故形态等;环境特征包括事故发生的时间、地点、道路环境等。从以上多方面信息中筛选出性别(gender)、年龄段(age group)、季节(season)、是否工作日(weekday)、时间段(time period)、路面情况(pavement condition)、道路类型(road type)、事故形态(accident pattern)、车辆类型(vehicle type)、违反交通信号(traffic signal viola-

tion)、违法变更车道(illegal lane change)、未按规定让行(failure to yield)、制动不当(improper braking)、非机动车逆行(non-motor vehicle retrograde)14个因素作为影响交通事故严重程度的变量,见表1。从数据统计结果可知,一般事故数量最多,占比为37.5%;其次为轻微事故,占比为33.8%;数量最少的为重大事故,占比为28.7%。

2 研究方法

2.1 建模方案设计

将交通事故的严重程度等级作为因变量,同时选取影响交通事故发生的人、车、路、环境四个方面的特征因素作为自变量。考虑到数据中影响因素较多,传统的数据预处理方法,如主成分分析,并不能有效提高模型的预测性能。因此,本研究采用了一种融合通道注意力的卷积神经网络(convolutional neural network with channel attention, CA-CNN)模型来预测交通事故的严重等级。该模型通过调

表1 变量分析与频率统计

Tab.1 Variable analysis and frequency statistics

Variable	Definition	Symbol	Ratio/%	Variable	Definition	Symbol	Ratio/%
Gender	1=male	Gender_1	71.1	Accident pattern	1=collision with a moving vehicle	Shape_1	69.4
	2=female	Gender_2	28.9		2=vehicle and non-motor vehicle accidents	Shape_2	5.8
Age group	1=29 years old and younger	Age_1	28.4		3=other accidents	Shape_3	24.8
	2=30 to 39 years old	Age_2	39.0	Vehicle type	1=non-motor vehicle	Vehicle_1	32.0
	3=40 to 49 years old	Age_3	17.4		2=car	Vehicle_2	53.1
	4=50 years old and elder	Age_4	15.2		3=passenger car	Vehicle_3	10.2
Season	1=spring	Season_1	54.0		4=truck	Vehicle_4	2.3
	2=summer	Season_2	1.7		5=other	Vehicle_5	2.4
	3=autumn	Season_3	0.9	Traffic signal violation	1=yes	R_1	9.3
	4=winter	Season_4	43.4		0=no	-	90.7
Weekday	1=workday	Weekday	78.3	Illegal lane change	1=yes	R_2	9.1
	0=day off	-	21.7		0=no	-	90.9
Time period	1=[0:00, 6:00)	Time_1	1.1	Failure to yield	1=yes	R_3	21.5
	2=[6:00, 12:00)	Time_2	34.5		0=no	-	78.5
	3=[12:00, 18:00)	Time_3	36.1	Improper braking	1=yes	R_4	12.6
	4=[18:00, 24:00)	Time_4	28.3		0=no	-	87.4
Pavement condition	1= dry	Condition_1	84.5	Non-motor vehicle retrograde	1=yes	R_5	6.4
	2= slippery	Condition_2	15.3		0=no	-	93.6
	3= others	Condition_3	0.2	Accident severity	1=minor	-	33.8
Road type	1= ordinary road	Type_1	66.8		2=general	-	37.5
	2= three forks	Type_2	7.6		3=major	-	28.7
	3= four forks	Type_3	25.6				

整不同特征在网络预测中的权重,实现了对不同事故严重等级(如轻微事故、一般事故、重大事故)的有效识别,从而增强了模型的预测性能,并提高了道路交通安全预警的水平。

卷积计算被当作一种特殊的线性运算,常用于提取图像特征。CNN具有训练速度快、准确率高等优点,该模型主要包括卷积层、池化层和分类器层^[12-13]。与传统的特征提取方法不同,CNN可以自动提取特征信息(其结构如图1所示),在特征识别方面同样具有平移不变性的优势,因此广泛应用于时间序列的特征识别和提取。

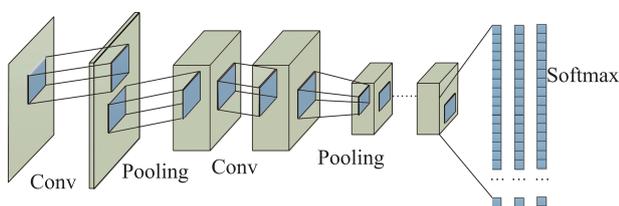


图1 卷积神经网络结构图
Fig. 1 CNN structure diagram

为了应对CNN中普遍存在的通道间信息价值不均衡问题,本文引入文献[14]中提出的通道注意力与卷积神经网络进行组合,构建CA-CNN交通事故严重等级预测模型。

通道注意力通过均值池化全局信息压缩到信息通道维度,以此来建立不同通道间的相互依存关系,对卷积计算得到的特征通道维度加权,提高神经网络对数据特征的表达能,从而增强网络模型对目标的判别能力,提高模型识别准确率。为了能从复杂背景下区分不同特征的重要程度,给予重要特征更高的权重有利于提升神经网络对数据的特征提取能力,进一步获取不同特征的关键程度,将池化后的向量分别送入瓶颈结构的挤压-激励网络,获取不同重要程度特征的权重,通道注意力模块流程如图2所示。

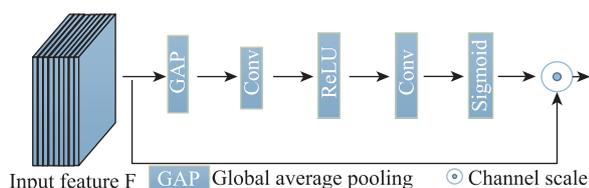


图2 通道注意力机制模块
Fig. 2 Channel attention mechanism module

设送入通道注意力机制的矩阵为 $X \in R^{H \times 1 \times a}$; H 为通道数; a 为数据长度。通过挤压-激励网络输出的两个结果相加,以此得到新的特征矩阵向量 $c' \in R^{H \times 1 \times 1}$ 。最后,通道注意力机制的输出矩阵为

$$C = X \sigma(c') \quad (1)$$

通过全局平均池化层保留通道特征,使用两层 1×1 卷积层学习通道特征权重,并利用sigmoid激活函数进行归一化处理,可以得到不同通道的权值分配信息,有利于特征提取。

$$M_n(F_1) = \sigma(MLP(\text{AvgPool}(F_1))) = \sigma(W_1(W_0(F_{1-\text{avg}}^n))) \quad (2)$$

$$F_0 = M_n(F_1) * F_1 + F_1 \quad (3)$$

式中: σ 为激活函数; MLP 为感知机; $F_{1-\text{avg}}^n$ 为平均池化特征; W_0, W_1 为减少率; n 是特征的通道维度的大小。 $M_n(F_1)$ 是一个通道注意力参数矩阵, F_1 和 F_0 分别表示通道注意力模块特征图的输出和输入。

在通道注意力的基础上,模型将具有不同道路事故严重程度的数据作为输入数据,构建基于融合通道注意力的卷积神经网络识别模型,通道注意力模块用于提取不同特征变量之间的差异,增强或减弱特征在事故严重程度预测中的输入权重。CA-CNN交通事故严重程度预测模型网络结构如图3所示。

在网络结构的设置上,经过多次调试,模型设置多个卷积层、2个通道注意力单元、1个输入层及1个输出层。每个隐含层结构采用卷积层、激活层、卷积层结构;通道注意力单元采用自适应平均池化层、卷积层、激活层、卷积层、激活层的结构;输出层为1个卷积层。模型网络隐含层中采用了ReLU作为激活函数,通道注意力单元中第1个激活层使用ReLU函数,第2个激活层使用sigmoid函数。各卷积层中卷积核大小为1,卷积核移动步长均设置为1。训练模型时,网络采用SGD函数优化学习率,初始学习率设置为0.02。

2.2 模型评价指标

在对模型分类预测模型进行评估时,评价指标种类多样,但不同的指标可能得到不同的结果。本研究采用分类准确率(accuracy)、查准率(precision)、召回率(recall)等5种预测评价指标对事故严重程度预测精度进行评估^[15],其详细表达式见式

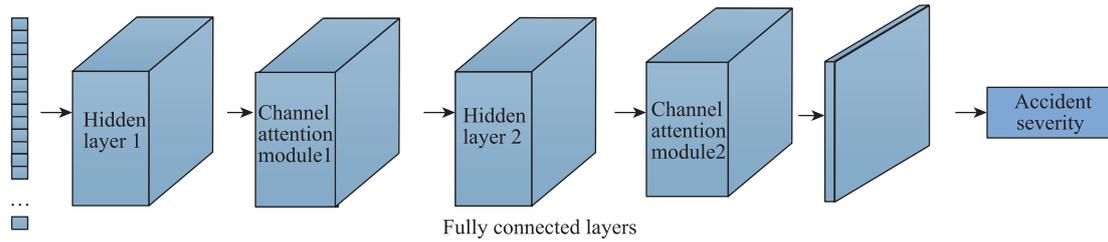


图3 CA-CNN网络结构图
Fig.3 CA-CNN network structure

(4)~式(8)。

1) 准确率,分类准确度衡量了分类模型总体预测准确度,计算公式如下

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \quad (4)$$

式中:TP样本为正类,预测为正类;FP样本为负类,预测为正类;TN样本为负类,预测为负类;FN样本为正类,预测为负类。

2) 查准率计算公式如下

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

3) 召回率计算公式如下

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

4) $F1_score$ 是统计学中用来衡量模型精确度的一种指标,它同时兼顾了分类模型的准确率和召回率

$$F1_score = \frac{2Precision \times Recall}{Precision + Recall} \quad (7)$$

5) MCC 用以衡量分类预测模型性能的指标,该指标考虑了真正类、真负类、假正类和假负类,通常认为该指标是一个比较均衡的指标,计算公式如下

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (8)$$

3 模型结果评估分析

3.1 模型预测结果分析

通过数据的预处理,本研究形成一份包含885个样本,14个特征变量的交通事故数据集。分类变量标签以事故严重程度为标准,划分为3个等级。算法使用python语言在pytorch深度学习框架上进行编程。计算机使用i5-1135G7处理器以及16GB内存硬件配置。考虑到本研究数据量较少,因此选取了90%的数据用于训练模型,10%的数据作为测试集。

3.2 不同模型预测结果对比

根据本文建模方案,将经预处理的数据集分别输入CNN,CA-CNN,RF,NaiveBayes,Logistics,Decision_table,Bagging 7个模型中。为充分验证此模型在事故严重程度预测上具有优势,将CA-CNN模型与以上模型进行比较。

依据各模型预测结果计算出对应混淆矩阵,利用混淆矩阵显示模型对各事故严重程度的预测结果与实际指标之间的差异。在混淆矩阵中,每一行代表特定事故严重程度的实际观察次数,而各列代表预测次数。对角线上的单元格值表示准确预测特定事故严重程度的次数,非对角线单元格则表示错误分类,这可能导致模型估计过高或过低。各模

表2 各模型混淆矩阵对比
Tab.2 Comparison of the confusion matrix of each model

Model	Accident severity	Minor	General	Major	%
CNN	Minor	65.5	10.3	17.2	
	General	8.3	86.1	5.6	
	Major	17.4	13.0	69.6	
CA-CNN	Minor	72.4	6.9	20.7	
	General	8.3	88.9	2.8	
	Major	13.0	8.7	78.3	
RF	Minor	67.7	9.7	22.6	
	General	3.7	96.3	0	
	Major	16.7	10.0	73.3	
NaiveBayes	Minor	58.1	6.5	35.5	
	General	3.7	92.6	3.7	
	Major	23.3	13.3	63.3	
Logistics	Minor	61.3	9.7	29.0	
	General	7.4	92.6	0	
	Major	16.7	13.3	70.0	
Decision_table	Minor	54.8	12.9	32.3	
	General	11.1	88.9	0	
	Major	16.7	13.3	70.0	
Bagging	Minor	51.6	9.7	38.7	
	General	11.1	88.9	0	
	Major	23.3	16.7	60.0	

型预测结果对应的混淆矩阵详见表2。在轻微事故的预测中,CA-CNN模型的预测准确率最高,达到72.4%;在一般事故的预测中,RF模型的预测准确率最高,达到96.3%;在重大事故的预测中,CA-CNN模型的预测准确率最高,达到78.3%。在所有分类预测模型中,CA-CNN模型正确预测事故类别的数量最多,正确预测总数达到71例。其他几种模型正确预测的数量依次为:RF模型69例、CNN模型66例、Logistics模型65例、NaiveBayes模型62例、Bagging模型58例、Decision_table52例。

为从整体上了解模型整体预测效果,依据模型预测结果对应混淆矩阵,计算模型预测效果总体评

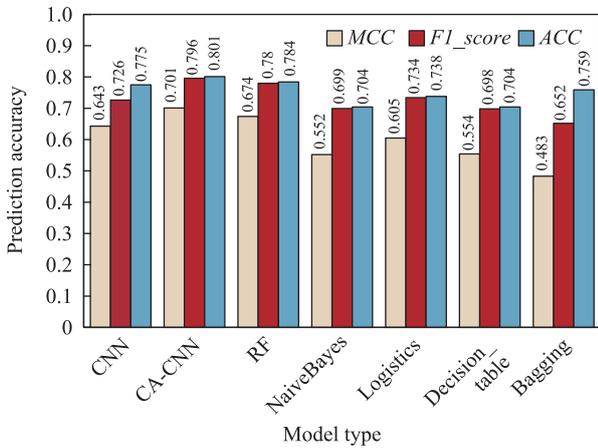


图4 模型预测效果总体评价指标

Fig. 4 Overall evaluation index of model prediction effect

价指标MCC, F1_score, ACC,具体计算结果见图4。

由图4可知,CA-CNN模型的整体预测性能优于其他模型。与其他模型相比,CA-CNN模型的F1_score值最大(0.796),相对于CNN,RF,Naive-Bayes,Logistics,Decision_table, Bagging模型分别提高了0.070,0.016,0.097,0.062,0.098和0.144,对应增长率分别为8.8%,2.0%,12.2%,7.8%,12.3%和18.1%;CA-CNN模型的MCC值最大,为0.701,相对于CNN,RF,NaiveBayes,Logistics,Decision_table, Bagging模型分别高出0.058,0.027,0.149,0.096,0.147和0.218,对应增长率为8.3%,3.9%,21.3%,13.7%,21.0%和31.1%;模型CA-CNN模型准确率为0.801,相对于其他模型分别提高了0.026,0.017,0.093,0.063,0.097和0.042,增长率为3.2%,2.1%,12.1%,7.9%,12.1%和5.2%。由以上分析可知,采用模型CA-CNN的预测效果最佳,其性能较Bagging等模型而言有明显提升。

为更具体地了解模型实际总体分类性能,引入了TP Rate(真正类率),FP Rate(假正类率),Preci-

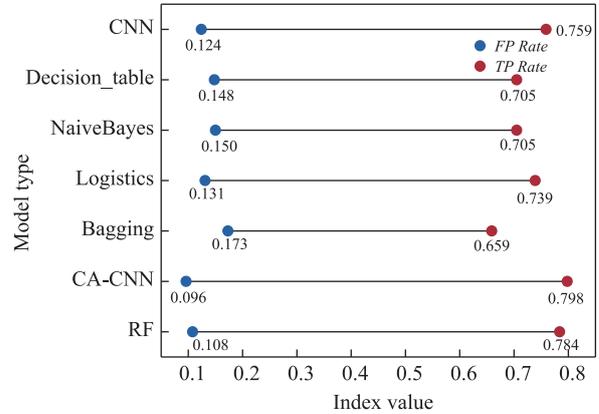


图5 模型预测效果TPR和FPR
Fig. 5 Model prediction effect TPR and FPR

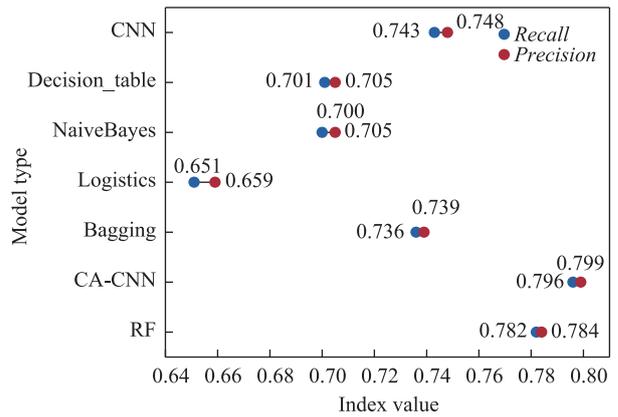


图6 模型预测效果评价指标Precision和Recall值

Fig. 6 Evaluation indicators of model prediction effect of Precision and Recall

sion和Recall评价指标,具体计算结果见图5和图6。

TPR数值越大,FP Rate数值越小,表明模型预测效果越好。从图5的数据可知,CA-CNN模型的TPR数值最大,其对应数值为0.798,其次为RF模型,TPR值为0.784, Bagging模型的TPR值最小,TPR值为0.659。CA-CNN模型的TPR值较Bagging模型和RF模型分别提高0.139,0.014,增长率为21.1%,1.8%。CA-CNN模型的FP Rate数值最小,其对应数值为0.096,RF模型,FP Rate值为0.108, Bagging模型的TPR值最大,FP Rate值为0.173。CA-CNN模型的FP Rate值较Bagging模型和RF模型分别减小0.077,0.012,减小比率为80.2%,12.5%。

由文献[13]可知,Precision和Recall的值越高,模型预测效果越好。从图6可知,CA-CNN模型对

应的 *Precision* 值和 *Recall* 值分别为 0.796 和 0.799, 远高于 Bagging 模型的 0.651 和 0.659, 分别提高了 0.145 和 0.140, 增长率为 22.3% 和 21.2%。

由以上分析可知, 从整体预测效果上看 CA-CNN 模型是一种较好的交通事故预测模型, 其整体预测效果优于 CNN, RF, Logistics 等模型。

3.3 多重共线性检验

为了排除不同影响因素之间存在较强相关性从而影响模型的准确性, 本文基于 Stata15.0 软件对回归模型 14 个自变量做了多重共线性诊断, 诊断结

表3 事故严重程度自变量的多重共线性检验

Tab.3 Multicollinearity test of independent variables of accident severity

Variable	Allowance	VIF
Gender	0.945	1.06
Age group	0.965	1.04
Season	0.874	1.14
Weekday	0.986	1.01
Time period	0.963	1.04
Pavement condition	0.973	1.03
Road type	0.841	1.19
Accident pattern	0.909	1.10
Vehicle type	0.849	1.18
Traffic signal violation	0.851	1.18
Illegal lane change	0.837	1.20
Failure to yield	0.789	1.27
Improper braking	0.846	1.18
Non-motor vehicle retrograde	0.838	1.19

果见表3。

衡量多重共线性程度一个常用的指标是方差膨胀因子 *VIF*。*VIF* 值越大, 表示自变量的容忍度越小, 越有共线性问题。通常, 以 10 作为临界值, 当 *VIF* 值大于 10 时, 则认为变量间存在明显的多重共线性。从上表结果可以看出, 所有变量的 *VIF* 值均远小于临界值 10, 表明它们之间不存在多重共线性, 可以作为模型的自变量。

3.4 模型标定结果分析

基于 Stata15.0 软件, 对交通事故严重程度的事故数据建立多元 Logit 模型, 设置置信度区间为 95%。当显著性水平 *P* 值小于或等于 0.05 时, 表示该变量对事故严重程度具有显著影响, 当 $P > 0.05$ 时, 说明该变量对事故严重程度影响显著性较低。对应的参数估计结果见表4。

表4 多元Logit模型参数估计

Tab.4 Parameter estimation of multivariate Logit model

Variable	Std. Err.	P
Gender	0.054	0.679
Age group	0.024	0.287
Season	0.017	0.000
Weekday	0.058	0.006
Time period	0.030	0.417
Pavement condition	0.065	0.407
Road type	0.030	0.050
Accident pattern	0.020	0.000
Vehicle type	0.030	0.204
Traffic signal violation	0.088	0.323
Illegal lane change	0.090	0.000
Failure to yield	0.065	0.000
Improper braking	0.078	0.011
Non-motor vehicle retrograde	0.106	0.410

从以上结果可以看出, 影响交通事故发生的因素中存在 7 个显著因子, 为季节、是否工作日、道路类型、事故形态、是否违法变更车道、未按规定让行及制动不当。

4 事故严重程度致因分析

基于此前分析得出的显著影响因素结果, 为量化不同影响因素对事故严重程度等级的影响, 本研究在多元 Logit 模型的基础上对存在显著影响的自变量进行了深入的边际效应分析, 具体边际效应结果见表5。

4.1 轻微事故致因分析

对于轻微事故而言, 普通道路、车辆与非机动车事故、夏季这 3 个因素对于该类事故的发生影响更为显著, 其对应的边际效用绝对值高达 43.3%, 43.1% 和 31.2%。在对一般事故以及重大事故机理分析中, 道路类型为普通道路的影响也很显著, 因此合并在一起进行机理分析。

1) 普通道路。此类道路类型会将 3 种事故发生的概率分别增加 43.3%, 46.7% 和 39.8%。道路原因主要表现在混合交通、平面交叉、标志不全、路面障碍和道路不符合标准, 此外, 一般车辆在普通道路上行驶的时间相对于三支以及四支分岔口时间更长, 从这个角度来说, 也会相应增大事故的发生概率。

2) 车辆与非机动车事故。此类事故形态会将

表5 边际效应结果
Tab.5 Marginal effect results

Variable	Accident category	dy/dx	Std. Err.	z	P> z
Season_1	Minor	0.235	0.022	10.57	0.000
	General	0.007	0.024	0.29	0.770
	Major	-0.246	0.022	-11.40	0.000
Season_2	Minor	0.312	0.102	3.07	0.002
	General	-0.100	0.091	-1.10	0.270
	Major	-0.166	0.108	-1.53	0.126
Weekday	Minor	-0.055	0.030	-1.85	0.064
	General	-0.033	0.026	-1.26	0.207
	Major	0.090	0.031	2.91	0.004
Type_1	Minor	0.431	0.042	10.22	0.000
	General	-0.467	0.021	-22.73	0.000
	Major	0.406	0.041	9.84	0.000
Type_2	Minor	0.222	0.069	3.21	0.001
	General	-0.160	0.039	-4.05	0.000
	Major	0.160	0.061	2.64	0.008
Shape_1	Minor	0.101	0.034	3.00	0.003
	General	-0.132	0.028	-4.76	0.000
	Major	0.073	0.038	1.91	0.057
Shape_2	Minor	-0.433	0.104	-4.14	0.000
	General	-0.086	0.049	-1.74	0.083
	Major	0.398	0.055	7.23	0.000
Shape_3	Minor	-0.138	0.079	-1.75	0.080
	General	-0.086	0.064	-1.34	0.179
	Major	0.261	0.065	4.02	0.000
R_2	Minor	-0.127	0.042	-3.02	0.003
	General	-0.111	0.045	-2.48	0.013
	Major	0.196	0.036	5.37	0.000
R_3	Minor	-0.171	0.036	-4.78	0.000
	General	0.011	0.029	0.38	0.706
	Major	0.138	0.030	4.55	0.000
R_4	Minor	-0.063	0.038	-1.67	0.095
	General	-0.030	0.035	-0.86	0.393
	Major	0.107	0.037	2.90	0.004

轻微事故发生的概率增加43.1%。考虑原因为非机动车驾驶员缺乏安全意识,没有注意交通信号和道路标志,违规情况较多,一些道路也存在安全隐患,容易造成事故发生。

3) 夏季。此季节因素会将轻微事故发生的概率增加31.2%。夏季白天时间增长,夜晚时间缩短,人们的睡眠时间也随之缩短。夜间睡眠不足,导致白天犯困、精神不能集中,甚至开车时打盹。此外,天气炎热,温度高,驾驶员心情容易急躁,尤其是对于暴露在外的非机动车驾驶员。

4.2 一般事故致因分析

对于一般事故而言,普通道路对于该类事故的发生影响最为显著,对应的边际效用绝对值为46.7%,其次为三支分岔口、车辆间运动碰撞边际效用绝对值分别为16.0%和13.2%。

1) 三支分岔口。此类道路类型会将一般事故发生的概率减小16.0%。这可能是由于三支分岔口相对于四支分岔口的路况较简单,因而对事故发生的影响较小。

2) 车辆间运动碰撞。此类事故形态会将一般事故发生的概率减小13.2%。由于此类型事故是车辆之间的事故,由于车身外壳以及车辆防护设备,发生事故时往往能对乘车人起到保护作用。

4.3 重大事故致因分析

对于重大事故而言,车辆与非机动车事故和普通道路跟其他因素相比影响更为显著,边际效用绝对值为39.8%和40.6%,其他事故形态的边际效用绝对值为26.1%,违法变更车道的边际效用绝对值为19.6%。

1) 车辆与非机动车事故。此类事故形态会将重大事故发生的概率增加39.8%,由于机动车驾驶速度过快,而非机动车部件老化,甚至存在机械故障,导致安全性较差,事故发生时,非机动车往往因缺乏防护设施导致驾驶员重伤甚至当场死亡。

2) 其他事故形态。其他事故形态会将重大事故发生的概率增加26.1%,其他事故形态包括翻车、坠落、爆炸等,这些类型的事故危害极大,往往造成巨大的生命财产损失。

3) 违法变更车道。违法变更车道会将重大事故发生的概率增加19.6%,在车辆进行车道变换时,驾驶员需要执行一系列复杂的操作,这不仅会耗费他们的精力,还可能诱发超速行为。

5 结论

1) 提出CA-CNN的交通事故分类预测模型。在7种分类预测模型中,CA-CNN模型在准确预测事故类型方面表现最佳,准确预测了71起事故的类别,能够有效预测出轻微事故、一般事故和重大事故。

2) CA-CNN模型在7种模型中的TPR数值最大,FPR数值最小,Precision值和Recall值也最大,从整体预测效果上看CA-CNN模型是一种效果较好的交通事故预测模型。

3) 置信度区间为95%时,以下7个因素对交通事故的发生具有显著影响:季节、是否为工作日、道

路类型、事故类型、违法变道、未按规定让行以及不当制动。其中,普通道路、车辆与非机动车事故、夏季对轻微事故的影响更为显著;普通道路、三岔路口、车辆间运动碰撞对一般事故的影响更为显著;车辆与非机动车事故、普通道路、其他事故形态和违法变更车道对于重大事故的影响更为显著。

参考文献:

- [1] 世界卫生组织.2023 年全球道路安全状况报告[EB/OL]. (2023-12-13)[2024-02-15].<https://www.who.int/zh/publications/b/72400>.
- [2] 刘海珠. 道路交通事故严重程度影响因素分析及预测模型建立[D]. 长春: 吉林大学, 2014.
LIU H Z. The analysis of influencing factors for crash severity and the establishment of predication model[D]. Changchun: Jilin University, 2014.
- [3] SHAON M, QIN X, AFGHARI A, et al. Incorporating behavioral variables into crash count prediction by severity. A multivariate multiple risk source approach[J]. Accident Analysis and Prevention, 2019, 129: 277-288.
- [4] SUN J, SUN J. A dynamic Bayesian network model for real-time crash prediction using traffic speed conditions data[J]. Transportation Research Part C: Emerging Technologies, 2015, 54: 176-186.
- [5] 孙轶轩, 邵春福, 赵丹, 等. 交通事故严重程度 C5.0 决策树预测模型[J]. 长安大学学报(自然科学版), 2014, 34(5): 109-116.
SUN Y X, SHAO C F, ZHAO D, et al. Traffic accident severity prediction model based on C5.0 decision tree[J]. Journal of Chang'an University (Natural Science Edition), 2014, 34(5): 109-116.
- [6] PARSA A, TAGHIPOUR H, DERNIBLE S, et al. Real-time accident detection: coping with imbalanced data[J]. Accident Analysis and Prevention, 2019, 129:202-210.
- [7] VAIYAPURI T, GUPTA M. Traffic accident severity prediction and cognitive analysis using deep learning[J]. Soft Computing, 2021:1-13.
- [8] 李英帅, 张旭, 王卫杰, 等. 基于随机森林的电动自行车骑行者事故伤害程度影响因素分析[J]. 交通运输系统工程与信息, 2021, 21(1): 196-200.
LI Y S, ZHANG X, WANG W J, et al. Factors affecting electric bicycle rider injury in accident based on random forest model[J]. Transportation Systems Engineering and Information, 2021, 21(1): 196-200.
- [9] 石小林, 李涵, 黄子富. 基于二元 Logit 模型的直角事故受伤严重程度分析[J]. 交通运输工程与信息学报, 2020, 18(1): 106-110.
SHI X L, LI H, HUANG Z F. Analysis of the severity of right-angle accident injury based on a dual Logit model [J]. Journal of Traffic and Transportation Engineering and Information Technology, 2020, 18(1): 106-110.
- [10] JALAYER M, SHABANPOUR R, POUR-ROUHOLAMIN M, et al. Wrong-way driving crashes: a random-parameters ordered probit analysis of injury severity[J]. Accident Analysis & Prevention, 2018(117): 128-135.
- [11] 韩天园, 田顺, 吕凯光, 等. 基于文本挖掘的重特大交通事故成因网络分析[J]. 中国安全科学学报, 2021, 31(9): 150-156.
HAN T Y, TIAN S, LYU K G, et al. Network analysis on causes for serious traffic accidents based on text mining [J]. China Safety Science Journal, 2021, 31(9): 150-156.
- [12] SHUO J, FEI H, SHINING L, et al. Long short-term memory and convolutional neural network for abnormal driving behavior recognition[J]. Intelligent Transport Systems, 2020, 14(5): 306-312.
- [13] 赵建东, 赵志敏, 屈云超, 等. 轨迹数据驱动的车辆换道意图识别研究[J]. 交通运输系统工程与信息, 2022, 22(4): 63-71.
ZHAO J D, ZHAO Z M, QU Y C, et al. Vehicle lane change intention recognition driven by trajectory data[J]. Transportation Systems Engineering and Information, 2022, 22(4): 63-71.
- [14] 马吉权, 赵淑敏, 孔凡辉. 多尺度条形池化与通道注意力的图像语义分割[J]. 中国图象图形学报, 2022, 27(12): 3530-3541.
MA J Q, ZHAO S M, KONG F H. Semantic image segmentation by using multi-scale strip pooling and channel attention[J]. Chinese Journal of Image and Graphics, 2022, 27(12): 3530-3541.
- [15] XIE G, SHANGGUAN A, FEI R, et al. Unmanned system safety decision-making support: Analysis and assessment of road traffic accidents[J]. IEEE/ASME Transactions on Mechatronics, 2021, 26(2): 633-644.



通信作者:严利鑫(1988—),男,博士,副教授,博士生导师,研究方向为智能网联汽车关键技术,交通安全及事故致因分析。E-mail:yanlixinits@163.com。

(责任编辑:李根)