基于选择集成的山区高速事故预测模型

孟祥海[1]，王国锐[1]，张明扬[1]，田毕江[1,2]

（1.哈尔滨工业大学 交通科学与工程学院，哈尔滨 150090；2.云南省交通规划设计研究院有限公司 陆地交通气象灾害防治技术国家工程实验室，昆明 650200）

摘　要：为提升交通事故预测模型的精度并减少鲁棒性，利用Stacking集成策略构建事故预测模型。首先，构建基于决策树、极端随机树等8种机器学习模型的单一事故预测模型，利用MIC检验与图着色法度量各事故预测模型的相似度，选取相似度低、多样性强的模型参与集成；其次，对单一事故预测模型结果进行Box-Cox变换，并利用特征加权法为各单一模型分别赋予不同的权重；最后，选用BP神经网络、Logistic回归等模型作为元学习器进行Stacking集成。研究结果表明：元学习器选用BP神经网络的集成模型预测精度高于其他集成模型，相较于预测精度最高的单一事故预测模型，集成模型的MAE、RMSE分别降低24%和14%，R^2提高6%。

关键词：交通运输规划与管理；交通事故预测；山区高速公路；机器学习；集成学习

中图分类号：U491.31　　**文献标志码**：A　　**文章编号**：1671-5497(2025)04-1298-09

DOI：10.13229/j.cnki.jdxbgxb.20230725

Traffic accident prediction model of mountain highways based on selection integration

MENG Xiang-hai[1], WANG Guo-rui[1], ZHANG Ming-yang[1], TIAN Bi-jiang[1,2]

(1. *School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin* 150090, *China*; 2. *National Engineering Laboratory for Prevention and Control Technology of Land Transport Meteorological Disasters, Yunnan Provincial Transportation Planning and Design Research Institute Co., Ltd., Kunming* 650200, *China*)

Abstract: To improve the prediction accuracy and reduce the robustness of the traffic accident prediction model, this paper uses the Stacking integration strategy to construct an integrated traffic accident prediction model. Firstly, single traffic accident prediction models based on eight machine learning models, such as Decision Tree and Extra Tree, were constructed and the MIC test was used to measure the similarity of each traffic prediction model with the graph coloring method, and the models with low similarity and high diversity were selected to participate in the integration. Secondly, Box-Cox transformations were applied to the results of the single accident prediction models and different weights were assigned to each single model separately using feature weighting method. Finally, models such as BP neural network and Logistic regression were selected as meta-learners for Stacking integration. The results of the study show that the

收稿日期：2023-07-11.

基金项目：云南省交通运输厅科技创新及示范项目(2021-90-2)；中央引导地方科技发展资金项目(2023ZYZX2009).

作者简介：孟祥海(1969-)，男，教授，博士.研究方向：道路交通安全.E-mail：mengxianghai100@126.com

prediction accuracy of the integrated model with BP neural network selected for the meta-learner is higher than other integrated models, and the MAE and RMSE of the integrated model have been respectively reduced by 24% and 14% and the R^2 has been improved by 6% compared to the single accident prediction model with the highest prediction accuracy.

Key words: transportation planning and management; traffic accident prediction; mountain highways; machine learning; integrated learning

0　引　言

本文研究的山区高速公路位于我国云南省，同时叠加了低纬高原双重特性，多存在连续陡下坡、急弯等不良路段，且气象环境复杂，相较于平原高速公路交通事故风险更高[1]，因此亟须构建预测精度高、抗噪性强的交通事故预测模型。

交通事故预测模型主要有参数模型[2-6]和非参数模型两种类型，本文主要研究后者。非参数模型又可以细分为传统机器学习模型和深度学习模型。在传统机器学习模型研究方面，主要包括基于树的模型和非基于树的模型。其中，基于树的模型主要包括决策树、随机森林、XGBoost[7]、LightGBM[8]等。树模型不需要特定形式的表达式，规则易于理解，结果具有一定的可解释性，同时对噪声和异常值相对不敏感，具有良好的鲁棒性。然而，它们具有一些缺点，如在处理不平衡数据时泛化能力弱，对少数类别（即小样本）的预测准确性差[9]。与树模型相比，非基于树的模型具有不同的优缺点。支持向量机在处理非线性和小样本数据集方面具有更好的性能，但对缺失数据很敏感，并难以选择核函数[10]。朴素贝叶斯具有对缺失数据不敏感、分类效率稳定等优点；然而，在属性条件独立的假设下，分类错误率很大[11]。

在深度学习模型方面，Zeng等[12]提出一种循环神经网络（Recurrent neural network, RNN），对将要发生而未发生的交通事故进行预测，提前告知交通事故的到来；宁静等[13]提出一种融合尺度缩减注意力机制和图卷积网络的城市交通事故预测模型，可以很好地捕捉时空相关性，并解决数据稀疏性和空间异质的问题。Lin等[14]借助深度神经网络、深度信念网络和卷积神经网络构建交通事故风险预测模型，识别出高风险交叉口路段的交通事故关键影响因素。此外，量子神经网络[15]、Elman神经网络[16]和灰色BP神经网络[17]等也常被用作开发交通事故预测模型。从交通事故预测的结果来看，深度学习模型具有学习能力强、覆盖范围广和适应性好等优点，但在处理小样本问题时，模型可能表现不佳。此外，深度学习模型还存在易过拟合、超参数选择困难等问题。

国内外诸多针对事故预测建模研究中，大多是利用单一模型进行交通事故预测，少有学者采用集成模型，但单一模型的解释能力有限，很难通过优化单一模型来提高模型的性能，在某一方面可能存在缺陷，导致常出现对某一路段预测精度较低的现象。通过结合各种算法的优势，集成学习可以获得更准确、更稳定的结果，是提高模型性能的有效途径。因此，本文结合山区高速公路特征，首先构建多个单一事故预测模型，之后在度量各模型相似度的基础上，选取预测精度高、多样性强的事故预测模型赋予权重后参与后续集成，最后利用改进的Stacking集成策略对各单一事故预测模型进行集成，以期进一步提升事故预测模型的精度并减少鲁棒性。

1　数据描述与处理

本文共采集到云南省境内三段山区高速公路的交通事故信息，共计11 739起。由于云南省地处云贵高原，地形复杂，地势起伏大，高速公路长大下坡、急弯等不良路段占比高。此外，云南地处亚热带，气候条件表现出高度的动态性和多样性。复杂的地貌叠加多变的气候导致交通事故频发，因此本文主要从道路线形、气候条件两方面筛选潜在的事故致因变量，主要包括交通运行状况（年平均日交通量）、平面线形（平曲线半径、直线长度等）、纵断面线形（纵坡坡度、竖曲线半径等）、气候条件（平均温度、平均风速、季节降雨量等），3条高速公路的里程、事故详细信息如表1所示。

为更好探究交通事故与各事故致因变量之间的关系，以路段属性为原则，取平纵线形指标作为划分依据，采用同质法结合道路线形几何条件划分路段单元，其中平面线形可以分为直线路段和

表1 交通事故数据基本信息

Table 1 Basic information about traffic accident data

高速公路名称	道路长度/km	起终点桩号	统计年限	伤亡事故	财产损失事故
高速公路一	104	K2457-K2561	3	56	5 113
高速公路二	48	K2579-K2627	5	30	1 443
高速公路三	59	K1959-K2018	5	28	5 069

平曲线路段,纵断面线形可以分为纵坡路段和竖曲线路段,划分后共得到1 718组路段单元。通过对所得的路段单元进行汇总,共收集到12种线形路段,具体如表2所示。

表2 预测单元路段类型划分

Table 2 Classification of road section types for forecasting units

路段单元类型	路段单元个数	公路一	公路二	公路三
直线-上坡路段	229	102	61	66
直线-下坡路段	229	102	61	66
直线-凸型竖曲线	82	26	16	40
直线-凹型竖曲线	114	44	16	54
右转-上坡路段	163	64	67	32
右转-下坡路段	165	64	73	28
右转-凸型竖曲线	96	41	33	22
右转-凹型竖曲线	108	45	45	18
左转-上坡路段	165	64	73	28
左转-下坡路段	163	64	67	32
左转-凸型竖曲线	96	41	33	22
左转-凹型竖曲线	108	45	45	18

利用全局距离法计算样本与近邻的两两间距,累加每个对象与其他对象的距离得到该对象的全局距离,同时引入箱线图法,将全局距离超过箱上限或低于箱下限的值认定为异常样本。由于无法进一步补充对应事故的可靠信息,本文选择直接剔除异常样本,剔除异常样本后剩余1 315组路段单元。为避免事故样本出现严重零堆积现象,预测模型的因变量选为年平均交通事故次数。

结合数据特点,引入方差膨胀因子法(VIF)对所收集到的各类事故致因变量进行多重共线性检验并判断其严重程度,由于气象变量、长陡下坡相关坡度变量各自间可相互解释,因此两者中分别选择季节降雨量、当前累计坡度参加多重共线性检验。检验结果显示,年平均日交通量与大型车比例、降雨量之间存在多重共线性。针对存在多重共线性的变量,有选择地保留其中之一,以提高模型预测精度。经筛选后纳入模型变量的统计性描述如表3、表4所示,共计17个变量,分为12个连续性变量、5个离散型变量。

表3 连续型变量描述性统计

Table 3 Continuous variable descriptive statistics

变量标识	变量名称	单位	最大值	最小值	平均值	标准差
NA	年平均事故次数	起	15.50	0	1.22	1.89
DT	日交通量	pcu/day	14 797	6 390	1 0246	2792
L	路段长度	m	598.36	50.18	244.00	135.97
LL	直线段长度	m	3 796.49	0	395.82	772.36
AH	平曲线偏角	(°)	185.68	0	23.60	29.05
CH	平曲线曲率	1/km	7.69	0	1.20	1.50
LH	平曲线长度	m	2 299.86	0	326.87	383.78
LS	纵坡长度	m	4 199.84	10	616.97	616.54
SD	竖曲线坡度差	(°)	8.15	0	1.30	1.70
ASC	当前累积坡度	%	0	−4.98	−0.64	1.23
LSC	当前累积坡长	m	26 917.28	0	2 029.34	4 881.48
AS	纵坡坡度	(°)	6.00	−6.00	0	2.25

表4 离散型变量描述性统计

Table 4 Discrete variable descriptive statistics

变量标识/赋值	变量名称	百分比/%	变量表示	变量名称	百分比/%
HC	平曲线		VT	竖曲线类型	
0	直线路段	41.34	0	纵坡路段	61.41
−1	右偏	29.33	−1	凸型竖曲线	17.93
1	左偏	29.33	1	凹型竖曲线	20.66
TC	缓和曲线		RT	路段类型	
0	直线路段	41.34	1	基本路段	83.21
1	有	53.35	2	收费站	10.33
−1	无	5.31	3	服务区	5.31
CS	连续下坡		4	桥梁	1.15
1	有	25.91			
0	无	74.09			

2 改进的Stacking策略

在集成时为体现各单一模型的差异性,首先利用机器学习方法构建单一的事故预测模型,然后利用最大信息系数(MIC)对各单一事故预测模型的相似性进行判别,在此基础上,借助特征加权法对Stacking集成策略加以改进,为后续建立交通事故预测集成模型提供理论基础。

2.1　单一事故预测模型构建与结果分析

按照机器学习数据划分原则：对于小规模样本集，常用的分配比例是80%训练集、20%测试集。因此，在1 315组预测样本中，训练集包含1 052条数据，占总样本的80%，剩余20%作为测试集，有263条数据。

采用决策树（Decision tree）、极端随机树（Extra-trees，ET）、随机森林模型（Random forest，RF）、梯度提升决策树模型（GBDT）、K近邻（K-nearest neighbor，KNN）、XGBoost模型、LightGBM模型、CatBoost模型8种机器学习模型分别构建单一事故预测模型，为避免过拟合并提升预测精度，通过采用5折交叉检验，将数据集划分为5个样本量相同且互不交叉的子集，依次遍历这5个子集，每次把当前子集作为验证集，其余所有样本作为训练集，进行模型的训练与评估，最后把5次评估指标E_i的平均值E作为最终的评估指标，其交叉检验的具体流程如图1所示。借助遗传算法对各单一事故预测模型进行超参数的寻优，以使预测效果最优。采用平均绝对误差（MAE）和均方根误差（RMSE）分析模型的准确性和稳定性，并借助R^2来检验模型的拟合优度，各单一事故预测模型的检验结果如表5所示。

表5　单一事故预测模型检验结果

Table 5　Single traffic accident prediction model test results

单一模型	MAE	RMSE	R^2
决策树	0.34	0.59	0.74
ET	0.32	0.52	0.8
RF	0.35	0.61	0.74
GBDT	0.35	0.51	0.81
KNN	0.43	0.69	0.64
XGBoost	0.29	0.43	0.84
LightGBM	0.33	0.51	0.81
CatBoost	0.3	0.46	0.83

通过对拟合优度进行综合比选，XGBoost模型拟合优度最高，其次是Catboost模型。LightGBM模型和GBDT模型的R^2均为0.81、决策树模型和随机森林（RF）模型的R^2同为0.74，这表明其对交通事故频次的拟合程度大致相同，但在后续研究中需探究预测模型的差异性；KNN模型拟合优度相对较差，但为保证异质集成模型的多样性，仍保留KNN模型，在后续的单一事故预测模型的选择中探究其作用。

图1　5折交叉检验

Fig. 1　5-fold cross-check

2.2　Stacking集成策略

单一事故预测模型常因模型自身的缺陷导致预测精度降低或出现模型陷入局部最小值等情况。借助Stacking集成将多个单一事故预测模型进行组合建模的思路：首先，利用初始数据训练单一模型；其次，利用单一模型输出的结果作为输入，进行二次预测；最后得到最终的预测结果。

Stacking作为一种分层结构，将上一层多个单一预测模型的预测结果作为下一层单一模型的输入，在模型训练的过程中，继续采用5折交叉验证来提高模型的预测精度，避免模型过拟合。基于交通事故数据集的Stacking模型的具体流程如下：

假设事故数据集$D=\{(x_i,y_i)\}$，$i=1,2,\cdots,n$，其中，x_i表示第i个样本的事故致因变量，y_i表示第i个样本的真实事故频次。

步骤1：将交通事故数据集D划分为k个大小相同且互不相交的子集$D_1,D_2,\cdots,D_k$，选取其中一个子集D_j作为验证集，$D_{-j}=D-D_j$作为训练集。

步骤2：利用D_{-j}训练得到T个单一事故预测模型$\eta_1,\eta_1,\cdots,\eta_T$，每一个单一事故预测模型$\eta_T$在验证集$D_j$可得到一个预测结果$R_{jt}$，$t=1,2,\cdots,T$。

步骤3：将所有单一事故预测模型在验证集上的输出结果R_{jt}以及事故频次真实值构成一个新的训练集，用于训练元学习器。

步骤4：将上一层多个单一预测模型的预测结果作为输入变量输入下一层元学习器中进行二次预测，进而得到最终的年平均事故预测次数。

2.3　Stacking模型的改进策略

每个单一事故预测模型都会输出一个事故频次预测值，但利用传统Stacking策略进行集成时没有考虑到各单一模型之间在预测机理、预测精度、模型的适用性等方面的差异，若将单一模型的所有预测结果赋予相同的权重，则最终的预测结

果难免会受到影响。

因此,采用特征加权法对Stacking集成策略的不足进行改进,基本思路:根据各单一模型的预测精度,为每一个单一模型的预测结果赋予一个权重,将赋予权重后的预测结果作为元学习器的输入,从而构建基于改进Stacking的集成模型。

计算各单一模型在训练集的绝对误差,并根据绝对误差确定相应的权重,公式为:

$$e_i=\frac{1}{n}\sum_{i=1}^{n}\left|y_i-\hat{y}_i\right| \tag{1}$$

$$a_i=\ln\frac{1+e_i}{e_i+0.1} \tag{2}$$

式中:e_i为第i个单一模型的平均绝对误差;y_i、$\hat{y}_i$分别为第i个样本点的真实值和预测值;n为训练集样本总数;a_i为第i个单一模型根据其绝对误差确定的权重。

将各单一模型的权重进行归一化处理并作为其输出预测结果的权值,公式为:

$$w_i=\frac{a_i}{\sum_{i=1}^{n}a_i} \tag{3}$$

式中:w_i为第i个单一模型归一化处理后的权值。

为提高Stacking模型的可预测性和正态性,对单一模型结果进行Box-Cox变换,以进一步减少预测误差。其Box-Cox的变换形式如下:

$$y(\lambda)=\begin{cases}\dfrac{y^{\lambda}-1}{\lambda},\lambda\neq 0\\ \exp(y(\lambda)),\lambda=0\end{cases} \tag{4}$$

式中:λ为变化参数;y、$y(\lambda)$分别为原始因变量和新变量。

Box-Cox的逆变换为:

$$y(\lambda)=\begin{cases}(\lambda y(\lambda)+1)^{\frac{1}{\lambda}},\lambda\neq 0\\ \exp(y(\lambda)),\lambda=0\end{cases} \tag{5}$$

参数λ可利用最大似然估计进行计算,构建似然函数L^*如下:

$$L^*(\lambda)=-\frac{n}{2}\lg(e^2)+(\lambda-1)\sum_{i=1}^{n}\lg(y_i) \tag{6}$$

式中:n为采样次数;e^2为$y(\lambda)$方差的极大似然估计值。

3 事故预测集成模型

3.1 结合多样性度量的单一事故预测模型选择

大量单一事故预测模型参与集成,会导致因各单一模型的预测结果相似而产生冗余的问题,降低模型的预测效率。因此,在保证Stacking集成模型中单一事故预测模型多样性的同时,应减少各单一模型的数量,所以需对各单一事故预测模型的相似性进行判别。

选用最大信息系数(Maximal information coefficient,MIC)进行单一事故预测模型的预测结果相似度度量。对所选取的8种单一模型进行最大信息系数检验,其结果如表6和图2所示。通过检验结果可以看出,不同单一模型的相似性存在较大的差异,ET模型和GBDT模型之间的相似度接近1,说明两者的预测结果相似,可只选择一种模型参与后续模型集成;LightGBM模型和CatBoost模型相较于其他单一事故预测模型存在较大的差异。

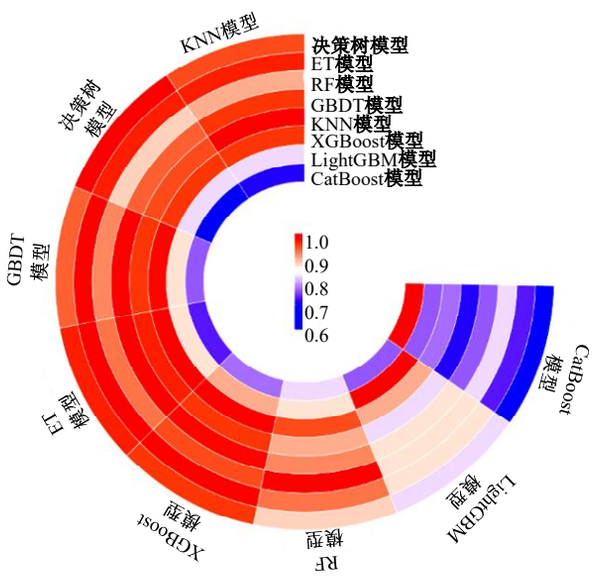

图2 最大信息系数环形热力图

Fig. 2 Maximum information coefficient circular heat map

为进一步探究各单一事故预测模型之间的互补性,利用互补指数方法进行邻接矩阵的转换。将8个单一事故预测模型的互补指数TF定义为:

$$\mathrm{TF}=\frac{\sum_{i,j=1,i\neq j}^{N}F_{ij}}{2C_N^2} \tag{7}$$

式中:F_{ij}为各单一事故预测模型之间的MIC值;N为单一事故预测模型数量;C_N^2为从N个单一事故预测模型中选取2个模型的组合数目。

通过计算得到互补指数为0.88,作为邻接矩阵转换时判别式的阈值。通过MIC定义可知,

表 6　单一事故预测模型 MIC 值

Table 6　MIC value of single traffic accident prediction model

单一事故预测模型	决策树模型	ET模型	RF模型	GBDT模型	KNN模型	XGBoost模型	LightGBM模型	CatBoost模型
决策树模型	1	0.97	0.88	0.94	0.95	0.96	0.82	0.69
ET模型	0.97	1	0.93	0.99	0.97	0.99	0.87	0.73
RF模型	0.88	0.93	1	0.92	0.9	0.95	0.87	0.82
GBDT模型	0.94	0.99	0.92	1	0.96	0.98	0.87	0.76
KNN模型	0.95	0.97	0.9	0.96	1	0.96	0.82	0.71
XGBoost模型	0.96	0.99	0.95	0.98	0.96	1	0.9	0.77
LightGBM模型	0.82	0.87	0.87	0.87	0.82	0.9	1	0.76
CatBoost模型	0.69	0.73	0.82	0.76	0.71	0.77	0.76	1

MIC 值越大代表单一事故预测模型之间的相似性越显著,而集成模型要求单一事故预测模型之间的相似性尽可能小且模型数量尽可能少,因此,提出以下判别式:

$$\mathrm{PB}=\begin{cases}0, F_{ij}>TF\\ 1, F_{ij}<TF\end{cases} \tag{8}$$

式中:PB 为单一事故预测模型的 MIC 值矩阵转变为邻接矩阵的转换值。

根据判别式(8)将单一事故预测模型的 MIC 值矩阵转变为邻接矩阵,如表 7 所示。

表 7　单一事故预测模型邻接矩阵

Table 7　Single accident prediction model adjacency matrix

单一事故预测模型	决策树模型	ET模型	RF模型	GBDT模型	KNN模型	XGBoost模型	LightGBM模型	CatBoost模型
决策树模型	0	0	0	0	0	0	1	1
ET模型	0	0	0	0	0	0	1	1
RF模型	0	0	0	0	0	0	1	1
GBDT模型	0	0	0	0	0	0	1	1
KNN模型	0	0	0	0	0	0	1	1
XGBoost模型	0	0	0	0	0	0	0	1
LightGBM模型	1	1	1	1	1	0	0	1
CatBoost模型	1	1	1	1	1	1	1	0

通过邻接矩阵绘制各单一事故预测模型的无向图(见图 3),图中模型相连代表模型之间相似度较小,采用图着色算法对相似度较高的单一事故预测模型赋予相同的颜色,着色结果如图 4 所示。LightGBM 模型和 CatBoost 模型显著区别于其他单一模型,各自独立成组;XGBoost 模型、RF、GBDT、决策树、KNN、ET 6类单一事故预测模型可归为一类,彼此之间有较大的相似性。选取该组中预测精度较高的 XGBoost 模型和 GBDT 模型参与后续的集成。最终选用 XGBoost、GBDT、LightGBM、CatBoost 4 种单一事故预测模型参与最终集成。

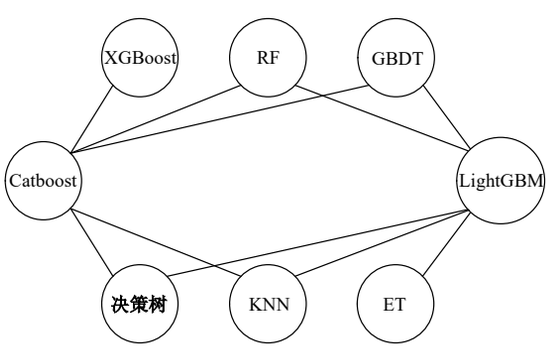

图 3　单一事故预测模型的无向图

Fig. 3　Undirected graph for single traffic accident prediction model

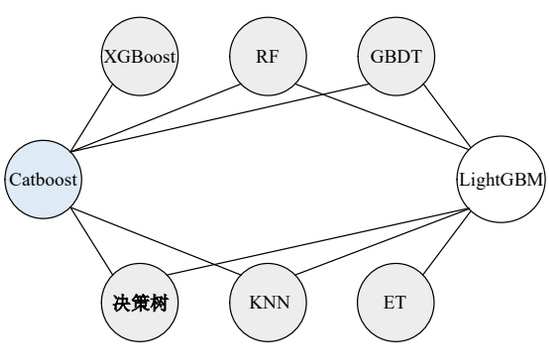

图 4　无向图的着色方案

Fig. 4　Coloring scheme for undirected graphs

3.2 事故异质集成模型构建

选择GBDT模型、Catboost模型、XGBoost模型、LightGBM模型构建单一事故预测模型，利用式(1)计算各模型的平均绝对误差(MAE)，并代入式(2)(3)中赋予各单一模型权重，4个单一模型权重依次为0.24、0.26、0.25、0.25，同时对各模型的预测结果分别进行Box-Cox转换。

目前较为常见的元学习器模型有线性回归模型、Logistic模型以及各种机器学习模型等，本文选择多重线性回归模型、Logistic模型、岭回归模型及BP神经网络模型作为元学习器。

将各单一事故预测模型的预测结果作为变量，结合各自的权重，将其输入元学习器中做进一步预测，进而得到最终的年平均事故预测次数，基于改进Stacking集成的事故预测模型流程如图5所示。

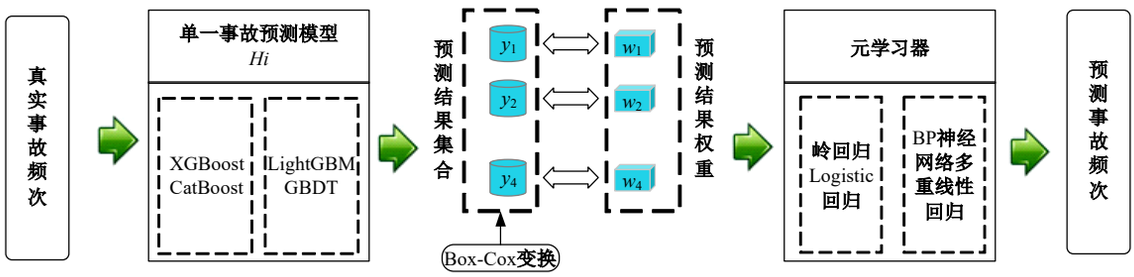

图5 基于改进Stacking集成的事故预测模型流程

Fig. 5 Incident prediction modeling process based on improved Stacking integration

3.3 事故异质集成模型检验及分析

为清晰直观表示预测事故数与实际事故数之间的对应关系，绘制能反映事故频次真实值与预测值之间变化规律的折线图，如图6所示。结果表明，改进后的Stacking集成模型的事故频次预测结果可以很好拟合真实事故频次，但个别样本仍存在较大的误差，这主要是由交通事故的随机性和偶然性的本质所决定的。

将测试集263条数据输入已建立的集成模型中进行预测，其预测结果如表8所示。可以看出，利用BP神经网络作为元学习器的集成模型预测精度优于以多重线形回归模型、Logistic回归模型

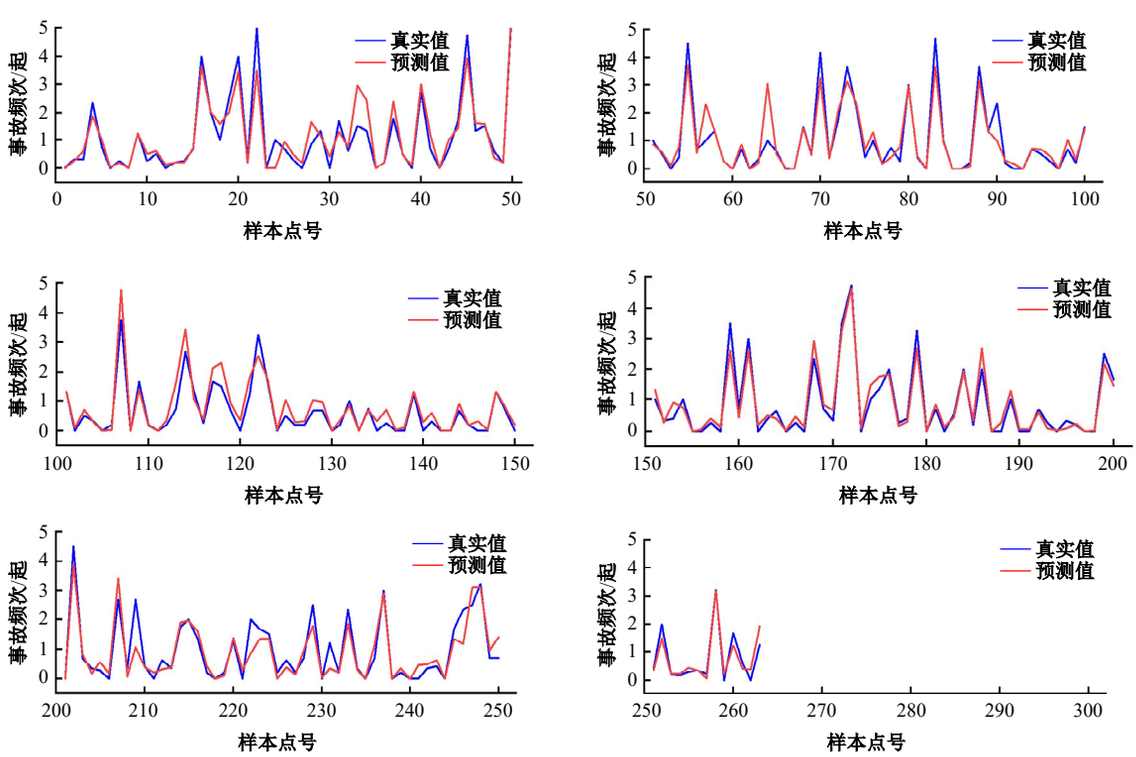

图6 基于改进Stacking的交通频次预测结果

Fig. 6 Traffic frequency prediction results based on improved Stacking

和岭回归模型作为元学习器的集成模型,且以上4种集成模型的预测精度均高于单一模型,相较于预测精度最高的单一事故预测模型XGBoost,以BP神经网络作为元学习器的集成模型预测准确度MAE、RMSE分别降低了24%和14%,R^2提高了6%。

表8　不同元学习器的Stacking集成模型综合比选

Table 8　Comprehensive comparison of stacking integration models with different meta-learners

元学习选取/模型精度	BP神经网络	多重线形回归	Logistic回归	岭回归
MAE	0.22	0.27	0.27	0.25
RMSE	0.37	0.45	0.43	0.41
R^2	0.89	0.82	0.85	0.86

4　结　论

(1)利用MIC检验、邻接矩阵、图着色算法和特征加权法等方法构建了基于改进Stacking集成的交通事故预测模型。

(2)在以多重线性回归模型、Logistic模型、岭回归模型及BP神经网络模型作为元学习器构建的4种集成模型中,BP神经网络模型展现出最佳的预测效果。

(3)4种集成模型不仅具有较高的预测精度,而且均优于单一事故预测模型。相较于预测精度最高的单一模型XGBoost,以BP神经网络作为元学习器的集成模型预测精度MAE、RMSE分别降低了24%和14%,R^2提高了6%。

(4)本文主要的贡献在于利用改进的Stacking集成策略进行事故预测,并与多种模型进行对比分析,丰富了道路交通事故频次预测的理论体系;同时,针对山区高速公路系统地梳理多方面的潜在事故影响因素,对现有事故影响因素体系进行了有效的补充。当然,建立更加完备的数据库、考虑更多方面潜在事故影响因素,进而有效地缓解交通事故随机波动问题,是开展下一步工作的重点。

参考文献:

[1] 张显强, 贺中华, 梁永娜, 等. 贵州省道路分形特征及其对交通事故影响机制[J]. 公路, 2017, 62(6): 197-203.
Zhang Xian-qiang, He Zhong-hua, Liang Yong-na, et al. Fractal characteristics of road and its impact mechanism on traffic accidents in Guizhou Province [J]. Highway, 2017, 62(6): 197-203.

[2] Macedo M R, Maia M L A, Rabbani E R K, et al. Traffic accident prediction model for rural highways in Pernambuco[J]. Case Studies on Transport Policy, 2022, 10(1): 278-286.

[3] 马壮林, 邵春福, 李霞. 基于Logistic模型的公路隧道交通事故严重程度的影响因素[J]. 吉林大学学报:工学版,2010, 40(2): 423-426.
Ma Zhuang-lin, Shao Chun-fu, Li Xia. Analysis of factors affecting accident severity in highway tunnels based on Logistic model[J]. Journal of Jilin University (Engineering and Technology Edition), 2010, 40(2): 423-426.

[4] 陈英, 袁华智, 黄中祥, 等. 零截尾负二项模型在交叉口事故预测中的应用[J]. 中国公路学报, 2020, 33(4): 146-154.
Chen Ying, Yuan Hua-zhi, Huang Zhong-xiang, et al. Modeling intersection traffic crashes using a zero-truncated negative binomial model[J]. China Journal of Highway and Transport, 2020, 33(4): 146-154.

[5] Roland J, Way P D, Firat C, et al. Modeling and predicting vehicle accident occurrence in Chattanooga, Tennessee[J]. Accident Analysis & Prevention, 2021(149): 105-117.

[6] Ihueze C C, Onwurah U O. Road traffic accidents prediction modelling: an analysis of Anambra State, Nigeria[J]. Accident Analysis & Prevention, 2018(7), 112: 21-29.

[7] 谢学斌, 孔令燕. 基于ARIMA和XGBoost组合模型的交通事故预测[J]. 安全与环境学报, 2021, 21(1):277-284.
Xie Xue-bin, Kong Ling-yan. On the ways to the traffic accident prediction based on the ARIMA and XGBoost combined model[J]. Journal of Safety and Environment, 2021, 21(1): 277-284.

[8] 纪俊红, 昌润琪, 温廷新. 基于GSK-AdaBoost-LightGBM的交通事故死亡人数预测研究[J]. 安全与环境工程, 2021, 28(1): 24-28.
Ji Jun-hong, Chang Run-qi, Wen Ting-xin. Prediction of traffic accident death toll based on GSK-AdaBoost-LightGBM[J]. Safety and Environmental Engineering, 2021, 28(1): 24-28.

[9] Vilaa M, Macedo E, Coelho M C. A rare event modelling approach to assess injury severity risk of vulnerable road users[J]. Safety, 2019, 5(2): 29-38.

[10] Xing L, He J, Li Y, et al. Comparison of different

models for evaluating vehicle collision risks at upstream diverging area of toll plaza[J]. Accident Analysis and Prevention, 2020(135): 86-97.

[11] Kwon O H, Rhee W, Yoon Y, et al. Application of classification algorithms for analysis of road safety risk factor dependencies[J]. Accident Analysis and Prevention, 2015(75): 1-15.

[12] Zeng K H, Chou S H, Chan F H, et al. Agent-centric risk assessment: accident anticipation and risky region localization[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 2222-2230.

[13] 宁静, 佘红艳, 赵东, 等. 一种路网级交通事故风险预测方法[J]. 北京邮电大学学报, 2022, 45(2): 72-78.
Ning Jing, She Hong-yan, Zhao Dong, et al. A road-level traffic accident risk prediction method[J]. Journal of Beijing University of Posts and Telecommunications, 2022, 45(2): 72-78.

[14] Lin L, Wang Q, Sadek A W. A novel variable selection method based on frequent pattern tree for real-time traffic accident risk prediction[J]. Transportation Research Part C: Emerging Technologies, 2015(55): 444-459.

[15] 孙棣华, 唐亮, 付青松, 等. 基于量子神经网络的道路交通事故预测[J]. 交通运输系统工程与信息, 2010, 10(5): 104-109.
Sun Di-hua, Tang Liang, Fu Qing-song, et al. Road traffic accidents forecasting based on quantum neural network[J]. Journal of Transportation Systems Engineering and Information Technology, 2010, 10(5): 104-109.

[16] 覃薇. 基于负二项回归分析的高速公路神经网络事故预测模型[D]. 哈尔滨: 哈尔滨工业大学交通科学与工程学院, 2017.
Qin Wei. Neural network crash prediction model of freeway based on negative binomial regression analysis [D]. Harbin: School of Transportation Science and Engineering of Harbin Institute of Technology, 2017.

[17] 范中洲, 赵羿, 周宁, 等. 基于灰色BP神经网络组合模型的水上交通事故数预测[J]. 安全与环境学报, 2020, 20(3): 857-861.
Fan Zhong-zhou, Zhao Yi, Zhou Ning, et al. Integrated model for forecasting waterway traffic accidents based on the Gray-BP neural network[J]. Journal of Safety and Environment, 2020, 20(3): 857-861.