

DOI:10.11918/202311093

S-FBLS 驱动的城市道路交叉口事故严重程度预测

王铭欢,王睿涵,李巧茹,陈亮

(河北工业大学 土木与交通学院,天津 300401)

摘要:为全面提升交通事故严重程度预测效果,针对现阶段传统机器学习与深度学习方法预测精度有限、网络收敛缓慢等问题,提出一种改进模糊宽度学习系统(fuzzy broad learning system, FBLS)的城市道路交叉口事故严重程度预测方法。模型采用 Takagi-Sugeno 模糊系统取代宽度学习系统(broad learning system, BLS)的特征节点层,在保留 BLS 快速收敛特性的前提下,更为广泛地提取高维事故数据所隐藏的特征;同时在模型的输入层融合 SMOTE 过采样算法,平衡事故数据类别,增强预测结果的可靠性。通过英国大曼彻斯特地区交通事故的历史数据,在横向维度上选择原始 FBLS,在纵向维度上选择交通事故严重程度预测常用的随机森林(RF)、支持向量机(SVM)、BP 神经网络(BPNN)、长短期记忆网络(LSTM)、卷积神经网络(CNN),分别与 S-FBLS 进行预测性能对比,结果表明:S-FBLS 在横向上相较原始 FBLS 将严重事故的预测准确率提升 52.87%,在纵向上相较 5 种对比模型提升网络训练速度 97% 以上,整体准确率分别提升 2.2%、8.95%、8.68%、6.47%、5.64%,特异度平均提升 6.49%,灵敏度平均提升 6.31%,精确度平均提升 5.66%。基于 S-FBLS 的事故严重程度预测方法可为城市道路交叉口事故发生预警提供可靠的理论支撑。

关键词:交通安全;事故预测;FBLS;SMOTE 算法;城市道路交叉口

中图分类号: TP301.6

文献标志码: A

文章编号: 0367-6234(2025)07-0162-09

S-FBLS-driven accident severity prediction at urban road intersection

WANG Minghuan, WANG Ruihan, LI Qiaoru, CHEN Liang

(School of Civil and transportation, Hebei University of Technology, Tianjin 300401, China)

Abstract: In order to comprehensively improve the effect of traffic accident severity prediction, for the current stage of traditional machine learning and deep learning methods with limited prediction accuracy and slow convergence of the network, proposing an improved FBLS method for predicting accident severity at urban road intersection. The model replaces the feature node layer of BLS with Takagi-Sugeno fuzzy system to extract the hidden features of high-dimensional accident data more extensively and still retains the fast convergence characteristics of BLS; the SMOTE algorithm is also fused in the input layer of the FBLS to balance the accident data categories and enhance the reliability of the prediction results. Through the historical data of traffic accidents in Greater Manchester, UK, the original FBLS was selected in the horizontal dimension, and RF, SVM, BPNN, LSTM, CNN, which are commonly used for traffic accident severity prediction, were chosen in the vertical dimension, to compare the model performance with the S-FBLS. The results show that comparison with the original FBLS, S-FBLS improves the accuracy of severe accidents by 52.87%, comparison with five comparative models, S-FBLS improves the network training speed by more than 97%, improves the overall accuracy by 2.2%, 8.95%, 8.68%, 6.47%, 5.64%, improves the specificity by an average of 6.49%, improves the sensitivity by an average of 6.31%, and improves the precision by an average of 5.66%. The S-FBLS-driven accident severity prediction method can provide a reliable theoretical support for the early warning of the occurrence of accident at urban road intersection.

Keywords: traffic safety; accident prediction; FBLS; SMOTE algorithm; urban road intersection

在城市道路交通网络中,平面交叉口作为各交通参与者发生方向转换的节点,也是各种交通组成集中发生交通冲突的地点,其潜在的事故风险对城市交通安全品质造成了较大的冲击。十字路口作

为平面交叉口的主要类型之一,具有更多的潜在冲突点,事故隐患与安全风险更高。因此,本文以十字路口为研究对象,构建可靠的城市道路平面交叉口事故严重程度预测模型,对于为城市交通事故预

收稿日期: 2023-11-30;录用日期: 2024-01-06;网络首发日期: 2025-02-05

网络首发地址: <https://link.cnki.net/urlid/23.1235.T.20250127.1527.002>

基金项目: 国家自然科学基金(52372302)

作者简介: 王铭欢(1996—),男,硕士研究生;陈亮(1978—),男,副教授,硕士生导师

通信作者: 陈亮,chenliang@hebut.edu.cn

防与决策提供支持,以及城市道路平面交叉口交通安全品质提升具有重要的现实意义。

长期以来,国内外学者在交通事故严重程度预测的研究中多使用基于统计回归的模型^[1-2]与基于传统机器学习的模型^[3-4],如随机森林^[5]、极限梯度提升与支持向量机^[6]、故障树^[7]等。此外,随着人工智能技术近年来的迅猛发展,在机器学习领域除传统的机器学习模型,人工神经网络^[8-9]以及深度学习模型^[10-12]在交通事故严重程度的预测中也取得了较好的效果。在传统机器学习模型的研究基础上,文献[2]建立了累积 Logistic 回归预测模型,探究影响高原地区交通事故严重程度的显著性因素,预测准确率为 77.41%;文献[5]将 k -means 算法与随机森林模型相结合,并通过 SHAP 算法对预测结果进行可视化输出,预测准确率为 83.20%;文献[6]基于 XGBoost 和 SVM 分别建立十字路口事故严重程度预测模型,预测准确率分别为 79.60% 和 79.0%;文献[8]采用遗传算法对 BP 神经网络进行改进,预测准确率为 75.90%。在深度学习模型的研究基础上,文献[11]借助 Keras 框架搭建 RNN 模型,对美国内华达州 2015—2017 年交通事故进行预测,预测准确率为 77.35%;文献[12]将图卷积网络与长短期记忆网络相结合,通过时空图卷积网络对交通事故严重程度进行预测,在 3 个不同数据集上的平均预测准确率为 86.57%。以上研究均取得了良好的模型表现,但仍存在不同程度的局限性,如大多数研究并未考虑事故类别的不平衡问题、特征变量较少等。同时,上述模型准确率区间为 [75.90%, 86.57%],当预测准确率达到此区间水平,即高于 75% 时,说明模型达到现有研究先进水平,具有有效性。

尽管深度学习模型在事故严重程度预测上的表现优于多数传统机器学习模型,但由于涉及到庞大的超参数和复杂的“深度”结构,并采用逐层求梯度的方式通过反向传播更新权重,使得网络对于高维特征的数据训练过程缓慢,且易陷入局部最优。为了能够在保证预测精度的基础上提升网络收敛速度,文献[13]提出了宽度学习系统(broad learning system, BLS),该方法舍弃了深度学习模型的深层属性,在横向的逻辑范畴上构建网络,能够在保证预测准确率的同时大幅度减少训练时间,并且避免了梯度消失或爆炸的问题。在保留 BLS 结构的基础上,文献[14]介绍了一种改进的神经网络算法,即模糊宽度学习系统(fuzzy broad learning system, FBLS),并与包括 BLS 在内的非模糊以及神经模糊系统进行了预测性能的比较,结果表明较其他神经模糊系

统,FBLS 在预测精度和训练速度上均表现出了显著的优势。目前,基于 BLS 与 FBLS 的方法在图像分类^[13]、时间序列预测^[15]等方面得到了逐步的应用,并取得了较好的成果,但该方法在交通安全领域的研究中仍处于空白状态,基于 FBLS 的方法在交通事故严重程度预测中的有效性与优越性尚需验证。

本研究为广泛提取高维事故数据更细粒度的特征,解决深度学习模型面对此类复杂非线性数据时网络收敛缓慢的问题,同时针对事故类别分布显著不均衡的特点,在 FBLS 的输入层引入 SMOTE 过采样算法改进原始模型,提出一种基于 S-FBLS 的城市道路交叉口事故严重程度预测方法,并通过训练耗时、准确率(ACC)、灵敏度(SEN)、特异度(SPE)、精确度(PRE)对该方法进行评估。研究成果对城市道路交叉口交通事故的准确预测具有重要意义。

1 数据处理

1.1 数据来源与描述

样本数据来源于 2016—2018 年英国道路交通事故数据集 UK Road Safety Date,原始数据主要涉及事故发生时的道路、天气、环境以及涉事车辆的车辆类型、碰撞部位、驾驶员特征等因素。通过对原始数据进行筛选与清洗,选取最终获得的 2016—2018 年大曼彻斯特(Greater Manchester)地区交通事故数据为样本数据。

1.2 特征工程与数据清洗

1.2.1 数据筛选

1)受城市内部交叉口设计理念的影响,不同城市之间的地域性差异对交通事故特征存在潜在的影响,若研究对象包含多个不同城市的数据样本则可能对模型的预测精度产生干扰。同时,考虑到大曼彻斯特地区数据整体缺失较少,因此选取英国规模最大的都会郡之一大曼彻斯特(Greater Manchester)为研究样本,以此增强样本数据的同质性。

2)在确定研究样本的基础上,为得到面向城市道路的数据样本,根据道路类别将涉及公路的数据样本进行剔除。最后,结合事故发生地点与交叉口类型对无关数据进一步过滤,得到 2016—2018 年大曼彻斯特地区城市道路十字路口交通事故数据。

1.2.2 变量选取

在删除样本数据中无用属性的前提下,如不涉及事故形成诱因的因素,基于事故基本信息、人员因素、车辆因素、道路因素、环境因素确定特征变量。另外,针对现有研究普遍受限于数据维度较低从而导致数据的特征表达能力不足,影响模型对信息学习的效果,因此结合事故要素构造部分 0-1 变量以

提升数据维度,增强样本数据的特征表达能力,使模型可以更好地学习有效信息,在更细粒度的层面上提取样本特征。在以上工作的基础上,采用斯皮尔曼方法对全部特征变量与预测结果进行相关性分析,结果表明,在 0.05 级别下,每个特征变量与预测结果之间的 Sig. (双尾)值均小于 0.05,表明各特征变量与事故严重程度之间呈显著性关联。

因变量为事故严重程度,原数据集中将事故严

重程度分为亡人事故、重大事故以及轻微事故,考虑到现阶段中国道路交通事故普遍较为严重且致死率较高这一特点,并结合各国对于“交通事故导致的死亡是不可接受的”这一交通安全愿景的追求,根据是否有交通参与者在事故中受到导致死亡或需住院治疗的伤害,在样本数据中将亡人事故和重大事故划分为严重事故,轻微事故划分为非严重事故,并对选取的变量进行量化,最终自变量见表 1。

表 1 自变量列表

Tab. 1 Argument list

类别	特征变量	变量说明	变量取值
事故信息	number_of_vehicles	涉事车辆数	连续变量
	number_of_casualties	伤亡人数	连续变量
	season	事故发生的季节	1:3—5 月份
			2:6—8 月份
			3:9—11 月份
			4:12—(次年)2 月份
	day_of_week	事故发生在星期几	1:工作日
			2:休息日
			1:0:00—6:00 时
	time	事故发生的时间段	2:6:00—12:00 时
3:12:00—18:00 时			
4:18:00—24:00 时			
人员因素	male_driver	是否涉及男性驾驶员	0:否;1:是
	female_driver	是否涉及女性驾驶员	0:否;1:是
	driver_age_max	涉事驾驶员最大年龄	1:20 岁以下
			2:21 ~ 30 岁
		3:31 ~ 50 岁	
		4:50 岁以上	
车辆因素	pedal_cycle	是否涉及自行车	0:否;1:是
	motorcycle	是否涉及摩托车	0:否;1:是
	van	是否涉及货车	0:否;1:是
	towing_and_articulation	是否涉及拖挂车	0:否;1:是
	waiting_to_go_or_moving_off	是否涉及待驶或正在驶离车辆	0:否;1:是
	slowing_or_stopping	是否涉及减速或正在停止车辆	0:否;1:是
	turning_right	是否涉及右转车辆	0:否;1:是
	approaching_junction	是否涉及驶向交叉口进口处车辆	0:否;1:是
	cleared_junction	是否涉及通过交叉口进口处车辆	0:否;1:是
	leaving_main_road	是否涉及驶离主路的车辆	0:否;1:是
	entering_main_road	是否涉及驶入主路的车辆	0:否;1:是
	skidding_and_overturning	涉事车辆有无打滑或翻车	0:无;1:有
	offside_impact	涉事车辆有无驾驶员侧碰撞	0:无;1:有
	nearside_impact	涉事车辆有无乘客侧碰撞	0:无;1:有
	front_impact	涉事车辆有无正面碰撞	0:无;1:有
rear_impact	涉事车辆有无追尾碰撞	0:无;1:有	
道路因素	road_type	道路类型	1:单向车道
			2:双车道
			3:单车道
speed_limit	道路限速	1:≤30 km/h	
		2:>30 km/h	
road_surface_conditions	事故发生时的路面环境	1:干燥	
		2:潮湿	
		3:其他	

表1(续)

类别	特征变量	变量说明	变量取值
环境因素	light_conditions	事故发生时的照明情况	1:日光 2:黑暗但有路灯照明 3:其他
	weather_conditions	事故发生时的天气情况	1:晴朗无大风 2:下雨无大风 3:晴朗有大风 4:下雨有大风 5:其他
	junction_control	事故发生时的交叉口控制情况	1:交通信号灯 2:其他控制方式 3:无控制
	special_conditions_at_site	事故发生时的特殊情况	1:无特殊情况 2:道路标志有缺陷或模糊 3:道路施工 4:其他
	pedestrian_crossing_human_control	行人过街人为管控情况	1:无行人过街控制 2:存在人为管控
	pedestrian_crossing_physical_facilities	行人过街设施	1:无行人过街设施 2:非交叉口处信号灯控制的人行横道 3:交叉口处行人过街信号相位 4:路中安全岛 5:其他

1.2.3 缺失值处理

经分析,整体数据中变量 driver_age_max 存在10%的数据缺失,若直接删除缺失样本可能对数据整体结构造成影响,因此根据该变量的正态分布特

性对缺失值随机做插补,使得填补前后数据均沿正态分布趋势线两侧分布,数据分布形式保持一致,如图1所示。其他变量数据缺失的现象较少,对研究精度影响较小,可将该条数据样本直接删除。

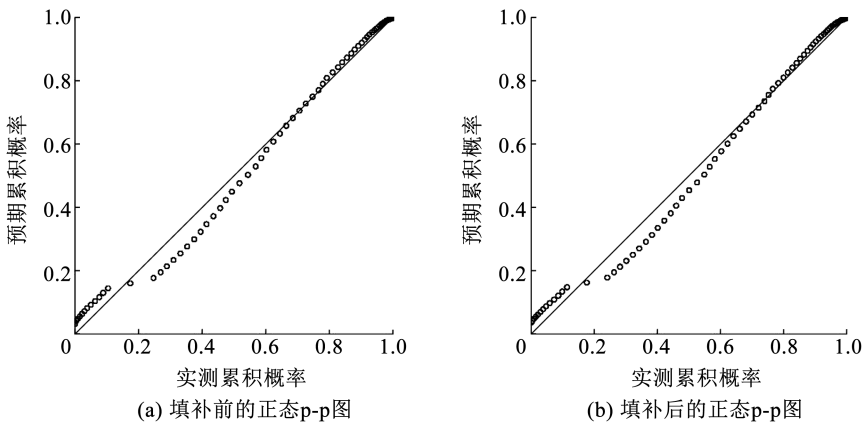


图1 填补前后变量分布形式

Fig. 1 Fill before and after variable distribution form

经缺失值处理后,得到有效样本2 022条,其中非严重事故编码为1,包含样本1 713条,占比84.7%;严重事故编码为2,包含样本309条,占比15.3%。

2 改进的模糊宽度学习系统

2.1 BLS 简介

宽度学习系统(broad learning system, BLS)是一种单层增量式神经网络,网络构建基于随机向量函

数链接神经网络(RVFLNN)和单层前馈神经网络(SLFNN)。其整体为3层网络结构,隐藏层由特征节点和增强节点两部分组成,与深度学习模型相比,BLS在对特征提取的过程中不依赖于网络层数的增加,且摒弃了梯度更新的方法,而是通过伪逆求取隐藏层到输出层的权重,避免了可能会遇到的陷入局部最优以及梯度消失或爆炸等问题,以上特点极大地提升了网络收敛速度。BLS的结构如图2所示。

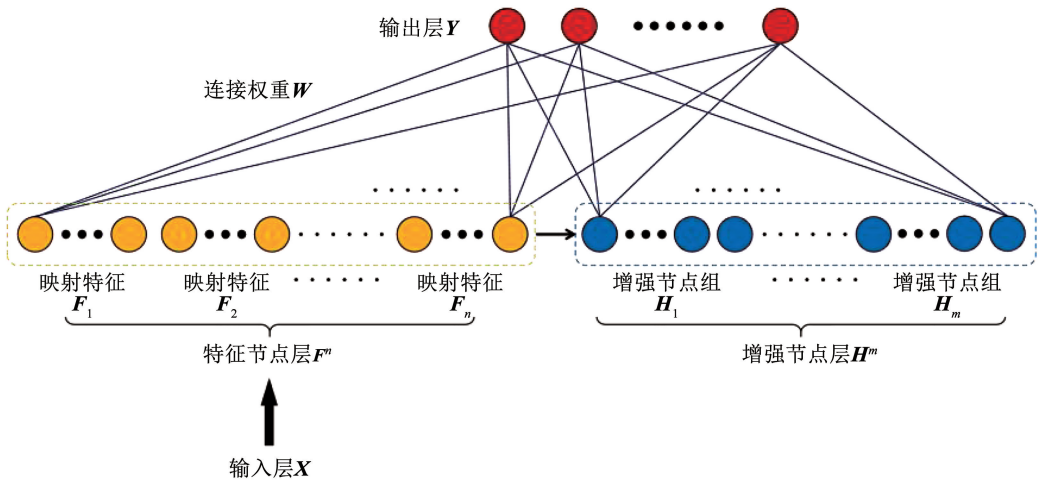


图 2 BLS 结构

Fig. 2 The structure of the BLS

2.2 S-FBLS 总体架构

模糊宽度学习系统 (fuzzy broad learning system, FBLS) 是一种改进的神经网络算法, 在保留 BLS 结构的前提下, 用多组 Takagi-Sugeno 模糊子系统替代 BLS 的特征节点层以增强网络的特征提取能力, 通过 k -means 算法确定各模糊子系统中模糊规则的个数以及高斯隶属函数的初始中心, 并将模糊子系统产生的输出传递至增强节点层, 最终的输出由模糊子系统的输出与增强节点组的输出共同构成。在 FBLS 整体架构的基础上, 针对交通事故严重程度类别不平衡的特点, 即非严重事故的数量远多于严重

事故, 在本研究的样本数据中严重事故占样本总体的 15.3%, 非严重事故占样本总体的 84.7%, 呈现出明显的不平衡性, 为避免原始 FBLS 对多数类样本可能产生的偏好倾向, 降低数据不平衡性对预测结果可靠性的影响, 在 FBLS 的输入层引入 SMOTE 过采样算法^[16-18]改进原始 FBLS, 将耦合 SMOTE 过采样算法的 FBLS 命名为 S-FBLS。模型整体结构如图 3 所示, $X \in \mathbf{R}^{N \times M}$ 为输入矩阵, $Y \in \mathbf{R}^{N \times C}$ 为输出层, 其中 N 为输入矩阵的行, M 为输入矩阵的列, C 为样本数据标签的类别。

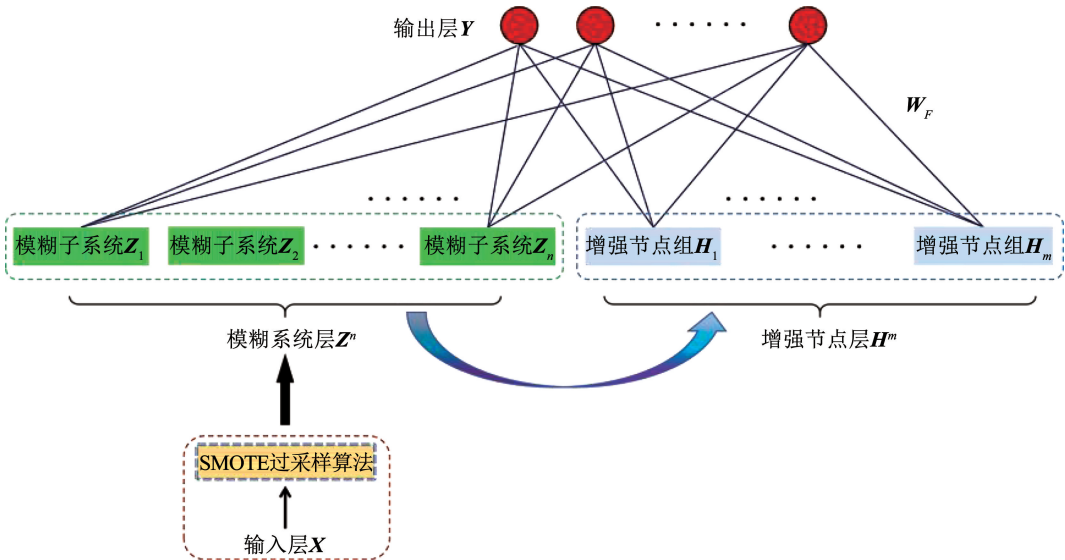


图 3 S-FBLS 结构图

Fig. 3 The structure of the S-FBLS

SMOTE 算法首先选择一个少数类样本 x 作为主样本, 并基于剩余的少数类样本选择主样本的 K 近邻样本, 在 K 近邻样本中随机产生主近邻样本 x' , 在主样本 x 与主近邻样本 x' 的连线上生成新的

少数类样本, 算法示意如图 4 所示。算法流程如下:
步骤 1 确定少数类样本的目标规模 Q 。样本数据中有 1 713 条非严重事故, 309 条严重事故, 严重事故经 SMOTE 过采样后的样本规模为 $Q = 1\ 713$ 。

步骤2 计算样本距离。在由 m 个特征变量所组成的空间中,从严重事故样本中随机选取某一个样本 x , x 与剩余 n 个严重事故样本的距离 $d(x, x_j)$ 可表示为

$$d(x, x_j) = \sqrt{\sum_{j=1}^m (f_i - f_{ji})^2} \quad (1)$$

式中 $j \in n$, f_i 与 f_{ji} 分别表示事故样本 x 与事故样本 x_j 第 i 个特征变量的数值。

步骤3 合成少数类样本。在步骤2的基础上,通过 x 的 K 近邻样本随机产生主近邻样本 x' , 并根据如下公式合成新的少数类样本 x_{new} , 即

$$x_{new} = x + \text{rand}(0, 1) \times |x - x'| \quad (2)$$

步骤4 重复进行步骤2、3,直至 $Q=1\ 713$ 。

经 SMOTE 过采样,使严重事故和非严重事故具有相同的样本数量,样本总量为 3 426 条。

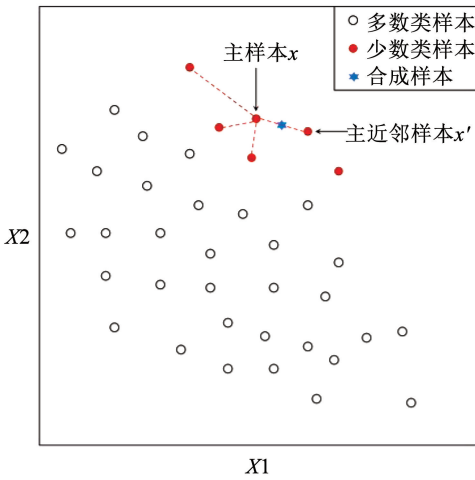


图4 SMOTE 算法示意图

Fig. 4 The illustration of SMOTE algorithm

为提高 S-FBLS 的网络鲁棒性和泛化能力,在 S-FBLS 中采用模糊子系统取代 BLS 特征节点部分用以微调权重的稀疏自动编码(sparse auto-encoder, SAE)。模糊子系统的结构如图5所示。

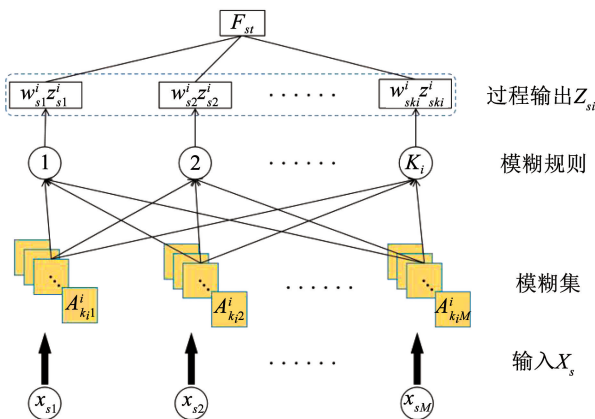


图5 模糊子系统结构

Fig. 5 The structure of the fuzzy subsystem

在图5中,假设一个 S-FBLS 有 n 个模糊子系统和 m 个增强节点组,输入数据为 $\mathbf{X} = (x_1, x_2, \dots, x_N)^T \in \mathbf{R}^{N \times M}$, 样本特征 $x_s = (x_{s1}, x_{s2}, \dots, x_{sM})$, $s = 1, 2, \dots, N$ 。假设第 i 个模糊子系统中有 K_i 条模糊规则,如果 x_{s1} 为 A_{k1}^i , x_{s2} 为 A_{k2}^i 且 x_{sM} 为 A_{kM}^i , 则 $z_{sk}^i = f_{sk}^i(x_{s1}, x_{s2}, \dots, x_{sM})$, $k = 1, 2, \dots, K_i$, 其中 z_{sk}^i 为一阶 TS 模糊系统,可表示为

$$z_{sk}^i = \sum_{t=1}^M \alpha_{kt}^i x_{st} \quad (3)$$

式中 α_{kt}^i 为由伪逆确定的系数。

τ_{sk}^i 为第 i 个模糊子系统中第 k 个模糊规则的激活强度,可表示为

$$\tau_{sk}^i = \prod_{t=1}^M \mu_{kt}^i(x_{st}) \quad (4)$$

$$\mu_{kt}^i(x) = e^{-\left(\frac{x - c_{kt}^i}{\sigma_{kt}^i}\right)^2} \quad (5)$$

式中 μ_{kt}^i 为模糊集 A_{kt}^i 的高斯隶属函数; c_{kt}^i 和 σ_{kt}^i 分别为高斯隶属函数的宽度和中心,通过 k -means 算法将 c_{kt}^i 应用于模糊子系统的训练集,并根据聚类中心完成对高斯隶属函数中心的初始化。

为了广泛提取输入数据所隐藏的特征,定义了一个由第 i 个模糊子系统中所有模糊规则的输出所构成的向量,并将其汇总为一个值作为该模糊子系统的去模糊化输出,然后将所有模糊子系统的去模糊化输出传递至增强节点层以进行非线性变换。用于第 s 个训练样本的第 i 个模糊子系统的输出向量为

$$\mathbf{Z}_{si} = (w_{s1}^i z_{s1}^i, w_{s2}^i z_{s2}^i, \dots, w_{sK_i}^i z_{sK_i}^i) \quad (6)$$

式中 w_{sk}^i 为每个模糊规则对应的权重,可表示为

$$w_{sk}^i = \frac{\tau_{sk}^i}{K_i} = \frac{\tau_{sk}^i}{\sum_{K=1}^{K_i} \tau_{sK}^i} \quad (7)$$

在样本数据标签为 C 个类别的情况下,即 $\mathbf{Y} \in \mathbf{R}^{N \times C}$, 用于第 s 个训练样本的第 i 个模糊子系统的去模糊化输出为

$$\mathbf{F}_{si} = \sum_{i=1}^M \alpha_{kt}^i x_{st} (w_{s1}^i, \dots, w_{sK_i}^i) \begin{pmatrix} \delta_{11}^i & \dots & \delta_{1C}^i \\ \vdots & \ddots & \vdots \\ \delta_{K_i1}^i & \dots & \delta_{K_iC}^i \end{pmatrix} \quad (8)$$

整体训练样本的第 i 个模糊子系统的输出为

$$\mathbf{Z}_i = (Z_{1i}, Z_{2i}, \dots, Z_{Ni})^T \in \mathbf{R}^{N \times K_i}, i = 1, 2, \dots, n \quad (9)$$

将 n 个模糊子系统的中间输出矩阵表示为 \mathbf{Z}^n , 并将 \mathbf{Z}^n 传递至增强节点层进行非线性变换。

$$\mathbf{Z}^n = (Z_1, Z_2, \dots, Z_n) \in \mathbf{R}^{N \times (K_1 + K_2 + \dots + K_n)} \quad (10)$$

假设第 j 个增强节点组中有 L_j 个神经元,增强节点层的输出矩阵可表示为

$$\mathbf{H}^m = (H_1, H_2, \dots, H_m) \in \mathbf{R}^{N \times (L_1 + L_2 + \dots + L_m)} \quad (11)$$

式中第 j 个增强节点组的输出矩阵为 $\mathbf{H}_j = \xi_j(\mathbf{Z}^n \mathbf{W}_{hj} + \beta_{hj}) \in \mathbf{R}^{N \times L_j}$, ξ_j 为非线性激活函数, 模糊子系统的输出 \mathbf{Z}^n 由权重 \mathbf{W}_{hj} 和偏置项 β_{hj} 连接至 $[0, 1]$ 内随机生成的相应增强节点组。

S-FBLS 的输出可表示为以下形式:

$$\mathbf{Y} = [\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n | \mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_m]^+ \mathbf{W}_F = [\mathbf{Z}^n, \mathbf{H}^m]^+ \mathbf{W}_F \quad (12)$$

与 BLS 相同, 式中 \mathbf{W}_F 同样为通过伪逆求取的将隐藏层连接至输出层的权重。

3 模型构建与比较

3.1 基于 S-FBLS 的十字交叉口事故严重程度预测

3.1.1 评价指标

对于二分类预测, 混淆矩阵被广泛用于度量模型的预测效果, 如表 2 所示。矩阵中包含 4 个参数, 分别为真阳性 (true positives, TP)、假阳性 (false positives, FP)、真阴性 (true negatives, TN)、假阴性 (false negatives, FN)。

表 2 事故严重程度预测二分类混淆矩阵

Tab.2 Confusion matrix of accident severity prediction

标签类型	预测为严重事故(2)	预测为非严重事故(1)
实际为严重事故(2)	TP	FN
实际为非严重事故(1)	FP	TN

选取准确率 (accuracy, ACC)、灵敏度 (sensitivity, SEN)、特异度 (specificity, SPE) 和精确度 (precision, PRE) 对模型的预测性能进行评价, 具体的计算方法定义为

$$ACC = \frac{TP + TN}{TP + FP + FN + TN} \quad (13)$$

$$SEN = \frac{TP}{TP + FN} \quad (14)$$

$$SPE = \frac{TN}{FP + TN} \quad (15)$$

$$PRE = \frac{TP}{FP + TP} \quad (16)$$

其中准确率表征模型的整体预测性能, 特异度表征模型对于非严重事故预测的准确率, 灵敏度表征模型对于严重事故预测的准确率, 精确度表征被正确预测的严重事故在所有被预测为严重事故样本中的比例。

3.1.2 S-FBLS 预测结果

本研究在 Windows10 下 MATLAB2022a 环境中进行编程分析, 随机选取样本数据的 80% 和 20% 作为训练集和测试集, 建立基于 S-FBLS 的城市道路十字路口事故严重程度预测模型。模型构建主要步

骤如下:

步骤 1 加载样本数据至输入层, 将经过 SMOTE 算法处理后的数据按比例划分训练集与测试集, 并作为模糊系统层的输入。

步骤 2 对于训练样本 $(\mathbf{X}, \mathbf{Y}) \in \mathbf{R}^{N \times (M+C)}$, 确定模糊规则 K_i 、模糊子系统 n 和增强节点组 m 的数量。

步骤 3 初始化函数 f_k 中的系数 α_{kt}^i 。

步骤 4 对于 $i = 1, \dots, n$, 将 k -means 算法应用于训练样本 \mathbf{X} 以获得 K_i 聚类中心, 并根据所得值完成高斯隶属函数中心的初始化。对于 $s = 1, \dots, N$, 计算用于训练样本 \mathbf{X}_s 的第 i 个模糊子系统的输出向量 \mathbf{Z}_{si} 与每个模糊子系统的去模糊输出。

步骤 5 获得训练样本 \mathbf{X} 的第 i 个模糊子系统的输出 \mathbf{Z}_i 。

步骤 6 获得 n 个模糊子系统的中间输出矩阵 \mathbf{Z}^n , 并计算增强节点层的输出矩阵 \mathbf{H}^m 与将 \mathbf{Z}^n 和 \mathbf{H}^m 连接到输出层的权重 \mathbf{W}_F 。

步骤 7 通过 \mathbf{W}_F 将 \mathbf{Z}^n 与 \mathbf{H}^m 输入至 S-FBLS 的顶层, 得到 S-FBLS 的最终输出。

步骤 8 输入预测数据至训练完成的 S-FBLS 中, 得到预测值。

模型的主要参数为模糊规则 N_r 、模糊子系统 N_f 以及增强节点组 N_e , 在保证模型不出现过拟合的前提下, 当 $N_r = 512, N_f = 1, N_e = 93$ 时模型在实验环境下取得最优的整体表现, 训练时间为 0.4 s, 且测试集与训练集的 SPE、SEN、PRE、ACC 均处于同一水平, 模型表现出了理想的拟合效果与泛化能力, 训练结果如表 3 所示。

表 3 模型训练结果

Tab.3 Model training results

	特异度	灵敏度	精确度	准确率
训练集	86.73	91.14	85.61	88.78
测试集	84.66	89.47	85.22	87.05

此时, 模型在测试集的整体准确率为 87.05%, 精确度为 85.22%, 对严重事故的预测准确率为 89.47%, 对非严重事故的预测准确率为 84.66%, 整体准确率高于现有研究先进水平, 表明了 S-FBLS 对交叉口事故严重程度预测的有效性。测试集混淆矩阵见表 4 所示, 输出结果如图 6 所示。

表 4 S-FBLS 测试集混淆矩阵

Tab.4 Confusion matrix of S-FBLS prediction set

标签类型	预测为严重事故(2)	预测为非严重事故(1)
实际为严重事故(2)	323	38
实际为非严重事故(1)	56	309

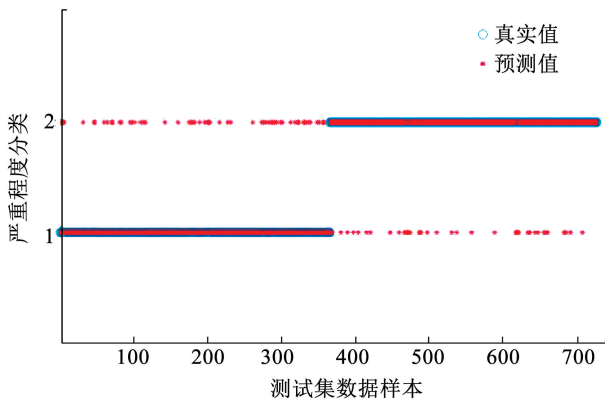


图 6 S-FBLS 预测结果

Fig. 6 S-FBLS prediction results

在此基础上,固定 $N_f = 1$,以 $N_r = 512$ 且 $N_e = 93$ 为中心评估网络对 N_r 与 N_e 数量变化的敏感性,结果如图 7 所示。当 N_r 与 N_e 分别满足 $N_r \in [507, 517]$ 与 $N_e \in [88, 98]$ 时,整体准确率的波动区间为 $[80.03\%, 87.05\%]$,且达到 82% 以上整体准确率的样本占总体的 84%,网络显示出了良好的稳定性。

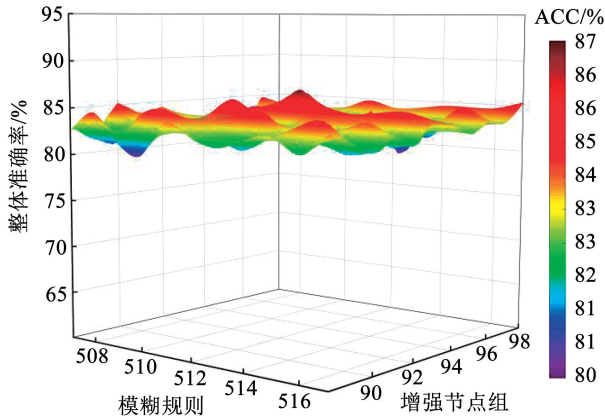


图 7 S-FBLS 稳定性测试

Fig. 7 Stability test of S-FBLS

3.2 模型对比分析

在验证 S-FBLS 对交叉口事故严重程度预测具有有效性的基础上,分别在横向和纵向的维度上综合对比 S-FBLS 与不同模型之间的预测效果,进一步验证 S-FBLS 的优越性。在模型对比分析的过程中,同样随机选取样本数据的 80% 与 20% 作为训练集与测试集,并控制实验环境与 S-FBLS 保持一致。

在横向上,选取原始 FBLS 作为对比模型,并将预测结果与 S-FBLS 进行综合对比,见表 5。结果表明,原始 FBLS 虽然取得了 78.91% 整体准确率,达到了现有研究先进水平,但对于少数类样本即严重事故的预测准确率仅为 36.60%,被正确预测为严重事故的样本仅占有被预测为严重事故样本总体的 30%,使模型 78.91% 的整体准确率表现出了强

不可靠性。S-FBLS 在原始 FBLS 的基础上极大地提升了灵敏度与精确度,提升幅度分别为 52.87% 与 55.22%,将模型对严重事故的预测精度提升至与非严重事故同一水平,消除了原始模型在预测过程中对非严重事故的偏好倾向,极大增强了模型输出的可靠性。

表 5 横向对比

Tab. 5 Horizontal comparison

预测模型	训练耗时/s	特异度/%	灵敏度/%	精确度/%	准确率/%
FBLS	0.3	85.90	36.60	30.00	78.91
S-FBLS	0.4	84.66	89.47	85.22	87.05

在纵向上,选用随机森林 (RF)、支持向量机 (SVM)、BP 神经网络 (BPNN)、长短期记忆网络 (LSTM)、卷积神经网络 (CNN) 5 种传统机器学习与深度学习领域较为成熟的模型对样本数据进行预测,并将预测结果与 S-FBLS 进行综合对比,见表 6。结果表明,5 种对比模型的整体准确率均高于 75% 现有研究先进水平,但基于 S-FBLS 的方法仍全面提升了预测性能,并大幅度降低了训练耗时,相较 LSTM 与 CNN 两种深度学习模型,S-FBLS 提升网络训练速度 99% 以上,在对城市道路交叉口事故严重程度的预测中表现出了较为明显的优越性。

表 6 纵向对比

Tab. 6 Vertical comparison

预测模型	训练耗时/s	特异度/%	灵敏度/%	精确度/%	准确率/%
RF	9.0	82.10	88.00	81.30	84.85
SVM	12.0	75.73	80.69	75.27	78.10
BPNN	105.0	75.29	81.43	78.12	78.37
LSTM	160.0	77.74	82.91	81.89	80.58
CNN	394.0	80.00	82.75	81.22	81.41
S-FBLS	0.4	84.66	89.47	85.22	87.05

4 结 论

本文首次在交通安全领域以 FBLS 为框架进行交通事故严重程度预测,同时通过 SMOTE 算法对模型整体架构进行改进,提出了一种基于 S-FBLS 的城市道路交叉口事故严重程度预测方法,并在横向和纵向两个维度上分别验证该方法的性能,主要结论如下:

1) 通过在原始 FBLS 的输入层添加 SMOTE 算法,使模型对严重事故的预测准确率由 36.60% 提升至 89.47%,回归至与非严重事故相同的水平,保证了模型输出结果的可靠性。

2) S-FBLS 的整体预测准确率达到 87.05%,

高于现有研究先进水平,并相较 5 种传统机器学习与深度学习模型分别提升 2.2%、8.95%、8.68%、6.47%、5.64%,验证了该方法在交叉口事故严重程度预测中的有效性和优越性。

3) S-FBLS 大幅度降低了训练耗时,克服了传统机器学习与深度学习模型潜在的网络收敛缓慢缺陷。

4) 相较于特异度, S-FBLS 取得了更高的灵敏度,并显著高于 5 种传统机器学习与深度学习模型,表明 S-FBLS 对严重事故具有更强的特征提取能力。考虑到在实际的事故预防与决策支持系统的构建中人们对非严重事故的预测结果往往较为包容,而对严重事故的预测结果更为看重,因此, S-FBLS 所表现出的这一特性在工程实际中具有广泛的实用空间。

参考文献

- [1] SHAHEED M S B, GKRIKZA K, ZHANG Wei, et al. A mixed logit analysis of two-vehicle crash severities involving a motorcycle[J]. *Accident Analysis & Prevention*, 2013, 61: 119
- [2] 孙小端, 侯洋洋, 贺玉龙, 等. 青藏高原双车道事故严重程度预测模型的建立[J]. *重庆交通大学学报(自然科学版)*, 2017, 36(7): 106
SUN Xiaoduan, HOU Yangyang, HE Yulong, et al. Establishment of accident severity prediction model of two-lane highway on Qinghai-Tibet Plateau[J]. *Journal of Chongqing Jiaotong University(Natural Science)*, 2017, 36(7): 106
- [3] SANTOS K, DIAS J P, AMADO C. A literature review of machine learning algorithms for crash injury severity prediction[J]. *Journal of Safety Research*, 2022, 80: 254
- [4] OBASI I C, BENSON C. Evaluating the effectiveness of machine learning techniques in forecasting the severity of traffic accidents[J]. *Heliyon*, 2023, 9(8): e18812
- [5] 陈治. 考虑天气和道路条件影响的道路交通事故严重程度分析及预测[D]. 北京: 北京交通大学, 2022
CHEN Zhi. Analysis and prediction of the severity of road traffic accidents considering the influence of weather and road conditions [D]. Beijing: Beijing Jiaotong University, 2022
- [6] 王利. 基于 XGBoost 和 SVM 的十字路口事故严重程度分析研究[D]. 北京: 北京交通大学, 2022
WANG Li. Analysis and research of intersection accident severity based on XGBoost and SVM [D]. Beijing: Beijing Jiaotong University, 2022
- [7] 刘佳雨, 冷军强, 尚平, 等. 冰雪路面下高速公路事故及严重程度影响因素分析[J]. *哈尔滨工业大学学报*, 2022, 54(3): 57
LIU Jiayu, LENG Junqiang, SHANG Ping, et al. Analysis of traffic crashes and injury severity influence factors for ice-snow covered freeway roads[J]. *Journal of Harbin Institute of Technology*, 2022, 54(3): 57
- [8] 吴芮. 基于 BP 神经网络的道路交通事故严重程度分析预测[D]. 成都: 西华大学, 2020
WU Rui. Analysis and prediction of road traffic accident severity based on BP neural network [D]. Chengdu: Xihua University, 2020
- [9] DELEN D, SHARDA R, BESSONOV M. Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks [J]. *Accident Analysis & Prevention*, 2006, 38(3): 434
- [10] RAHIM M A, HASSAN H M. A deep learning based traffic crash severity prediction framework[J]. *Accident Analysis & Prevention*, 2021, 154: 106090
- [11] 钱程. 数据驱动的交通事故严重程度预测[D]. 武汉: 华中科技大学, 2022
QIAN Cheng. Date-driven traffic accident severity prediction [D]. Wuhan: Huazhong University of Science and Technology, 2022
- [12] 刘志, 王锦梦, 孔祥杰. 基于时空图卷积网络的交通事故预测研究[J]. *浙江工业大学学报*, 2022, 50(2): 128
LIU Zhi, WANG Jinmeng, KONG Xiangjie. Research on traffic accident prediction based on spatio-temporal graph convolutional network[J]. *Journal of Zhejiang University of Technology*, 2022, 50(2): 128
- [13] CHEN C L P, LIU Zhulin. Broad learning system: an effective and efficient incremental learning system without the need for deep architecture [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, 29(1): 10
- [14] FENG Shuang, CHEN C L P. Fuzzy broad learning system: a novel neuro-fuzzy model for regression and classification [J]. *IEEE Transactions on Cybernetics*, 2018, 50(2): 414
- [15] ZHONG Dexiang, LIU Fagui. RF-OSFBLS: an RFID reader-fault-adaptive localization system based on online sequential fuzzy broad learning system[J]. *Neurocomputing*, 2020, 390: 28
- [16] ELREEDY D, ATIYA A F, KAMALOV F. A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning[J]. *Machine Learning*, 2024, 113(7): 4903
- [17] 熊馨, 徐礼胜, 王春武, 等. 不平衡数据集的 CT 结肠镜息肉检测方法[J]. *哈尔滨工业大学学报*, 2013, 45(11): 112
XIONG Xin, XU Lisheng, WANG Chunwu, et al. Polyp detection in CT colonography based on imbalanced date sets[J]. *Journal of Harbin Institute of Technology*, 2013, 45(11): 112
- [18] 石洪波, 陈雨文, 陈鑫. SMOTE 过采样及其改进算法研究综述[J]. *智能系统学报*, 2019, 14(6): 1073
SHI Hongbo, CHEN Yuwen, CHEN Xin. Summary of research on SMOTE oversampling and its improved algorithms [J]. *CAAI Transactions on Intelligent Systems*, 2019, 14(6): 1073