

# 基于 FHSPSO-RF 的交通事故严重程度预测

王淑椒, 胡学龙, 唐磊

安徽理工大学矿业工程学院, 安徽 淮南 232001

**摘要:**为提高交通事故严重程度的预测准确性,构建一种基于觅食生境选择粒子群优化(foraging habitat selection particle swarm optimization, FHSPSO)算法优化随机森林(random forest, RF)算法关键超参数的 FHSPSO-RF 模型,以美国华盛顿州西雅图市 2022 年 1 月至 2023 年 2 月的交通事故数据为基础,选取 12 项特征指标,采用合成少数类过采样技术,增加重伤和死亡事故的样本数,改善事故类别分布均衡性;将 FHSPSO-RF 模型与支持向量机(support vector machine, SVM)、K 近邻(K-nearest neighbors, KNN)和逻辑回归(logistic regression, LR)等模型进行性能对比,并通过 SHAP(Shapley additive explanations)分析法,解析各特征对交通事故严重程度的影响机制。结果表明:过采样后重伤、死亡事故的召回率显著增大, FHSPSO-RF 模型的整体性能更均衡; FHSPSO-RF 模型的准确率、精确率、召回率、F1 分数均高于其他 3 个模型,对交通事故严重程度的预测效果最好;在所有交通事故类型中,受伤人数和车辆数量均为最关键的驱动因素,对交通事故严重程度具有显著正向影响,行人数量、高冲击碰撞类型(如正面撞击)与复杂道路环境(如交叉口、匝道)是重伤及死亡事故的关键风险组合;财产损失事故与是否碰撞路边停放车辆密切相关。FHSPSO-RF 模型在交通事故严重程度预测中表现出良好的性能与可解释性,可为交通事故风险预测与防控决策提供依据。

**关键词:**交通事故;严重程度预测;FHSPSO;RF

中图分类号:U491.3

文献标志码:A

文章编号:1672-0032(2026)01-0025-09

引用格式:王淑椒,胡学龙,唐磊.基于 FHSPSO-RF 的交通事故严重程度预测[J].山东交通学院学报,2026,34(1):25-33.

WANG Shujiao, HU Xuelong, TANG Lei. Traffic accident severity prediction based on FHSPSO-RF[J]. Journal of Shandong Jiaotong University, 2026, 34(1): 25-33.

## 0 引言

随着我国机动车保有量的持续增长,交通事故导致的财产损失与人员伤亡问题日益严重。道路交通事故长期位居全球十大死亡原因之一,对公众健康构成严重威胁,也对社会稳定与经济发展带来负面影响。有效降低交通事故发生率及其伤害的严重程度,已成为交通管理领域亟待解决的关键问题。准确预测交通事故的严重程度,有助于交管部门实施及时、有针对性的干预措施,为事故预防与损失控制提供重要依据,进一步提高道路交通安全水平。

传统的交通事故严重程度预测方法主要基于 Logistic 回归模型与 Probit 模型等统计回归模型,何雅琴等<sup>[1]</sup>采用累积 Logistic 回归模型分析行人交通事故严重程度的影响因素,发现年龄、性别、天气条件和事故发生时间等因素具有显著影响;胡郁葱等<sup>[2]</sup>通过空间广义有序 Probit 模型分析高速公路交通事故的严重程度,指出天气条件、车速和交通量是关键因素。此类传统模型通常假设变量间存在线性关系,而实际交通事故中的影响因素一般呈复杂的非线性特征,传统模型的拟合效果往往存在一定局限。近年来,机器学习方法为交通事故严重程度的预测提供新的分析途径,张萌萌等<sup>[3]</sup>采用优化的分类与回归树

收稿日期:2024-11-13

第一作者简介:王淑椒(2000—),女,安徽铜陵人,硕士研究生,主要研究方向为深度学习与机器学习的智能交通系统优化、自动驾驶感知决策算法研发,E-mail:954573643@qq.com。

(classification and regression tree, CART)模型分析交通事故严重程度影响因素,但在小样本情况下模型精度可能受限;成卫等<sup>[4]</sup>基于树增广型贝叶斯网络对高速公路交通事故严重程度进行建模分析,但该方法仅依赖1 939起事故数据,可能影响模型的泛化能力。针对交通事故数据的不平衡问题,部分学者引入相应的优化策略,潘义勇等<sup>[5]</sup>提出一种基于 MobileViT 网络的交通事故严重程度预测模型,该模型结合焦点损失函数,能显著提高预测精度,但训练过程计算量较大;Vadhwani 等<sup>[6]</sup>采用改进的 XGBoost 模型预测道路碰撞伤害严重程度,模型的可解释性较弱,一定程度上影响结果的可信度;Kumar 等<sup>[7]</sup>对比分析多种机器学习与深度学习模型,识别影响交通事故的关键因素,但未优化模型重要参数。现有研究在交通事故严重程度预测的数据处理、模型构建与优化等方面已取得一定进展,但仍需进一步完善兼顾性能、效率与可解释性的模型优化方法。

本文构建一种基于觅食生境选择粒子群优化(foraging habitat selection particle swarm optimization, FHPSO)算法优化随机森林(random forest, RF)算法关键参数的 FHPSO-RF 模型,用于预测交通事故严重程度。在该模型中, FHPSO 将每个粒子位置映射为一组 RF 超参数组合,包括决策树的数量、分裂准则、树的最大深度和叶节点最小样本分裂数。通过引入生境选择策略更新粒子位置,增强种群多样性,避免陷入局部最优解。采用多项性能指标评估 FHPSO-RF 模型的预测性能,并与常见的机器学习分类模型进行对比分析,以期对交通事故的主动预防与精准管控提供参考。

## 1 模型构建

### 1.1 随机森林算法

RF 算法是一种基于 Bootstrap 重采样与决策树构建的集成学习算法。该算法通过 Bootstrap 方法从原始数据集中有放回地抽取多个子样本,每个子样本用于独立训练一棵决策树;每棵树的节点分裂过程中,未采用全部特征,只是随机选取一个特征子集作为候选分裂变量,增强模型的多样性与泛化能力。采用 RF 算法集成多棵决策树的预测结果形成最终输出:对于分类任务,采用投票机制,将得票数最多的类别作为预测结果;对于回归任务,取各棵决策树输出结果的平均值作为最终预测结果<sup>[8]</sup>。RF 算法投票机制公式为:

$$R(X) = \operatorname{argmax}_Y \sum_{i=1}^n I[r_{it}(X) = Y],$$

式中: $R(X)$ 为票数最多的决策结果类签; $X$ 为决策树中特征自变量; $Y$ 为决策结果的类标; $n$ 为决策树数量; $I(\cdot)$ 为指示函数; $r_{it}(X)$ 为决策树 $i$ 中节点 $t$ 的决策路径函数, $t \in [1, T_i]$ ,其中, $T_i$ 为决策树 $i$ 中的节点数。

RF 算法建模示意图如图 1 所示。

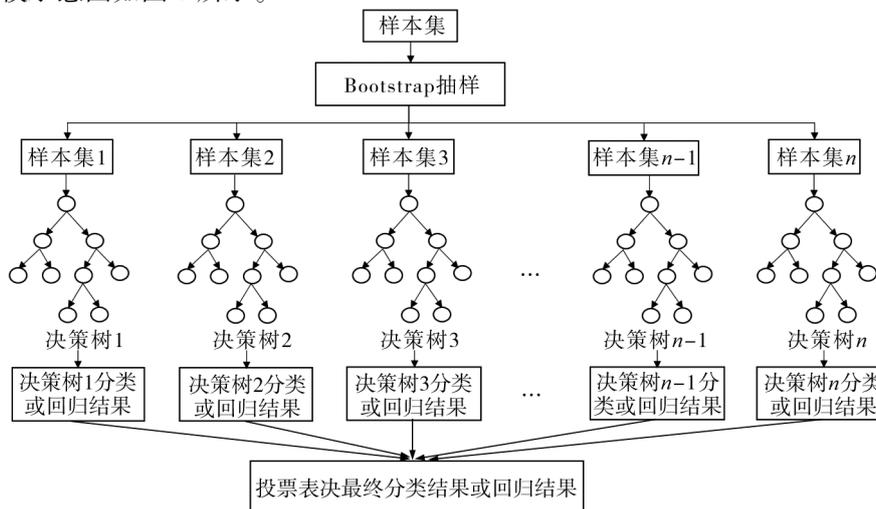


图 1 RF 算法建模示意图

RF算法的表现与其关键参数设置密切相关,选择合理参数可提高模型的性能,同时避免过拟合或欠拟合问题。其关键参数包括<sup>[9-11]</sup>:1)决策树数量,增加决策树的数量通常能提高模型的稳定性与预测准确率,直到达到某个数值后增益开始减少,更多的决策树意味着更强的集成效应,可进一步降低方差;但当决策树足够多时,继续增加决策树数量对性能改善影响不大,但增加计算量。2)最大深度,用于限制单棵决策树生长的最大深度,防止过拟合,若不限制树深,可能导致模型过复杂而捕捉到训练数据中的噪声。3)最小样本分裂数,控制节点在分裂前必须包含的最少样本数,防止过拟合,如果某节点中只有少量样本就进行分裂,那么这种分裂很可能是基于偶然性而非真实模式。4)分裂准则,由算法内部决定分裂准则,对每棵决策树,依据某个选定准则进行节点分裂,常见的分裂准则包括基尼不纯度和信息熵。基尼不纯度衡量随机抽取一个样本被错误分类的概率,基尼不纯度

$$G(D) = 1 - \sum_{m=1}^M p_m^2,$$

式中: $D$ 为数据集, $M$ 为类别总数, $p_m$ 为类别 $m$ 样本数在总样本数中的占比。

信息熵衡量数据集中类别分布混乱程度,信息熵

$$E(D) = - \sum_{m=1}^M p_m \log_2(p_m),$$

$\log_2(p_m)$ 选择2为底数是因为信息通常以比特(bits)衡量。

## 1.2 觅食生境选择粒子群优化算法

FHSPSO算法是一种受生物觅食行为启发的智能优化方法<sup>[12]</sup>。该算法在经典粒子群优化(particle swarm optimization, PSO)算法框架基础上,通过模拟动物在自然环境中觅食和选择栖息地行为<sup>[13]</sup>,引入生境评价与选择机制,增强全局搜索能力,有效避免早熟收敛。在FHSPSO算法中,粒子的特征与PSO算法类似,具有位置、速度和适应度3个属性,其核心改进在于迭代更新过程中,粒子依据个体历史最优与群体历史最优信息,结合当前生境质量动态调整运动策略。FHSPSO算法的位置与速度更新公式仍基于PSO公式,并引入生境相关参数和规则调控行为。

第 $i$ 个粒子在第 $k+1$ 次迭代的速度向量<sup>[14]</sup>

$$\mathbf{v}_i^{k+1} = w\mathbf{v}_i^k + c_1r_1(\mathbf{P}_i^k - \mathbf{X}_i^k) + c_2r_2(\mathbf{P}_g^k - \mathbf{X}_i^k), \quad (1)$$

式中: $w$ 为惯性权重,用于控制上一时刻速度的影响程度,保持粒子的动量, $w = w_s - (w_s - w_e) i_{\text{iter}} / \max i_{\text{iter}}$ ,其中, $w_s$ 为初始惯性权重, $w_e$ 为最终惯性权重, $i_{\text{iter}}$ 为当前迭代次数, $\max i_{\text{iter}}$ 为最大迭代次数; $\mathbf{v}_i^k$ 为第 $i$ 个粒子在第 $k$ 次迭代的速度向量; $c_1$ 、 $c_2$ 为学习因子; $r_1$ 、 $r_2$ 为在 $[0,1]$ 内均匀分布的随机数; $\mathbf{P}_i^k$ 为第 $i$ 个粒子的个体历史最优位置; $\mathbf{X}_i^k$ 为第 $i$ 个粒子在第 $k$ 次迭代的位置向量; $\mathbf{P}_g^k$ 为整个粒子群的全局历史最优位置。

第 $i$ 个粒子在第 $k+1$ 次迭代的位置向量

$$\mathbf{X}_i^{k+1} = \mathbf{X}_i^k + \mathbf{v}_i^{k+1}. \quad (2)$$

根据速度更新粒子位置后,需计算粒子适应度,评估当前位置的优劣。通过计算生境内所有粒子的平均适应度或最优适应度评估生境的整体质量,生境 $h$ 的适应度

$$F_h = \frac{1}{N_h} \sum_{h_i \in H_h} f(x_i), \quad (3)$$

式中: $N_h$ 为生境 $h$ 中的粒子数, $f(x_i)$ 为第 $i$ 个粒子的适应度, $H_h$ 为属于生境 $h$ 的所有粒子集合。

为模拟生物对更优生境的选择行为,定义迁移概率函数用于决定粒子是否从当前生境迁移至其他生境。该迁移概率主要依据各生境的适应度进行动态调整,粒子可自主决定是否向适应度更高的生境迁移。该机制有助于增强算法在不同搜索区域间的探索能力,提高全局寻优效果。粒子从生境 $h_i$ 迁移到生境 $h_j$ 的概率

$$P(h_j | h_i) = \frac{\exp(\beta F_{h_j})}{\sum_{h_j \in H, j \neq i} \exp(\beta F_{h_j})}, \quad (4)$$

式中: $\beta$ 为温度参数,用于控制选择过程的严格程度, $\beta$ 越大,选择过程更严格,粒子倾向于选择适应度更高的生境, $\beta$ 越小,粒子选择过程更随机; $F_{h_j}$ 、 $F_{h_i}$ 分别为生境 $h_j$ 、 $h_i$ 的适应度; $H$ 为所有生境的集合。

### 1.3 FHSPSO-RF 模型构建

引入FHSPSO算法优化RF算法,构建FHSPSO-RF模型预测交通事故严重程度。FHSPSO算法能有效调整RF算法的参数,使其在训练过程中迅速找到最佳参数组合。FHSPSO算法中粒子位置代表RF算法分类器的超参数组合,如粒子的位置边界为 $\{50,0,10,5\}$ 时,对应RF算法中决策树的数量为50,分裂准则(0表示基尼不纯度,1表示信息熵)为基尼不纯度,树的最大深度为10,内部节点最小样本分裂数为5;粒子的速度决定其在搜索空间中的移动方向和距离,速度向量的各分量为粒子提供多维搜索空间中的移动方向,速度的大小决定粒子每次迭代的移动距离;适应度用于评价每个粒子当前解的优劣。FHSPSO-RF模型的预测流程如图2所示。

采用FHSPSO-RF模型预测交通事故严重程度包括以下9个步骤。

1)初始化参数。初始化粒子群位置和速度;定义惯性权重 $w$ ,学习因子 $c_1$ 、 $c_2$ ,最大迭代次数 $\max i_{\text{ter}}$ ;建立初始生境,将粒子分配到不同生境中,并计算每个生境的适应度。

2)评估粒子适应度。计算每个粒子的适应度,根据适应度更新个体历史最优位置以及全局历史最优位置。

3)更新粒子速度和位置。根据式(1)(2)调整粒子的速度及位置,同时确保粒子位置在边界内。

4)动态调整惯性权重。根据当前迭代次数线性调整惯性权重 $w$ ,平衡模型全局探索和局部开发能力。

5)生境适应度评估。根据式(3)计算生境适应度,重新评估各生境的质量。

6)生境选择与迁移。根据式(4)将粒子迁移至更优生境。

7)记录历史最优解。更新并记录每次迭代后的个体最优和全优最佳解。

8)终止条件检查。若达到最大迭代次数或全局最优位置满足的最小界限,则停止迭代;否则返回步骤3)继续迭代。

9)输出最优参数并建模。将优化得到的最优超参数组合输入RF算法,预测交通事故严重程度。

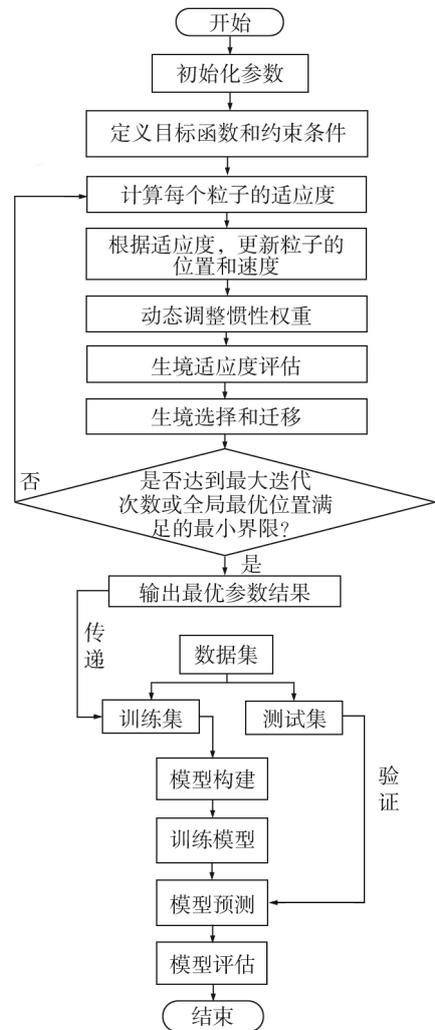


图2 FHSPSO-RF模型的预测流程

## 2 实证分析

### 2.1 特征指标

以美国华盛顿州西雅图市2022年1月至2023年2月的交通事故数据为样本,对原始样本数据进行清洗,剔除缺失、异常及重复记录的数据后,共得到5158条有效样本数据。将交通事故严重程度划分为财产损失事故、轻伤事故、重伤事故和死亡事故4类<sup>[15]</sup>。有效样本数据中共包括3165条财产损失事故、1744条轻伤事故、223条重伤事故、26条死亡事故。结合数据字段及文献[16-18],选取12项特征指标。针对特征指标文本类型的变量,为便于模型计算,对其进行统一编码处理。各特征指标及其赋值如表1所示。

表 1 特征指标及其赋值

特征指标	符号	赋值
人员数量	$X_1$	连续变量,取值范围为[1,9]
行人数量	$X_2$	连续变量,取值范围为[0,2]
非机动车数量	$X_3$	连续变量,取值范围为[0,2]
车辆数量	$X_4$	连续变量,取值范围为[0,8]
受伤人数	$X_5$	连续变量,取值范围为[0,5]
道路条件	$X_6$	干燥, $X_6=1$ ;湿滑, $X_6=2$
天气	$X_7$	晴天, $X_7=1$ ;雨天, $X_7=2$ ;雪天, $X_7=3$ ;阴天, $X_7=4$
光照条件	$X_8$	日光, $X_8=1$ ;昏暗, $X_8=2$ ;黑暗, $X_8=3$
碰撞类型	$X_9$	追尾, $X_9=1$ ;侧面碰撞, $X_9=2$ ;左转碰撞, $X_9=3$ ;右转碰撞, $X_9=4$ ;正面碰撞, $X_9=5$ ;角度碰撞, $X_9=6$ ; 与非机动车碰撞, $X_9=7$ ;与行人碰撞, $X_9=8$ ;与路边障碍物碰撞, $X_9=9$
地址类型	$X_{10}$	十字路口, $X_{10}=1$ ;街区道路, $X_{10}=2$ ;小巷, $X_{10}=3$
交叉口类型	$X_{11}$	在路口处,但事故与路口无关, $X_{11}=1$ ;在路口处,且事故与路口相关, $X_{11}=2$ ;在车道交汇处, $X_{11}=3$ ; 在路段中间,事故与路口有关, $X_{11}=4$ ;在路段中间,事故与路口无关, $X_{11}=5$ ;在匝道连接处, $X_{11}=6$
是否碰撞路边停放车辆	$X_{12}$	是, $X_{12}=0$ ;否, $X_{12}=1$

## 2.2 数据预处理

为消除特征指标的量纲差异、提高模型训练稳定性与收敛效率,对所有连续型特征指标进行 Z-score 标准化处理。该方法将原始特征值转换为均值为 0、标准差为 1 的正态分布,使不同尺度的特征具有可比性<sup>[19]</sup>。标准化后的特征值

$$Z = (x - \mu) / \sigma,$$

式中: $x$  为原始特征值, $\mu$  为特征值的均值, $\sigma$  为特征值的标准差。

## 2.3 FHSPSO-RF 模型预测性能

FHSPSO 算法的参数设置为:粒子群规模为 50,最大迭代 100 次,学习因子  $c_1=c_2=2.0$ ,初始惯性权重  $w_s=0.9$ ,最终惯性权重  $w_e=0.4$ ,单个生境的最大粒子数为 5。RF 算法的关键超参数通过 FHSPSO 算法在以下范围进行优化:决策树数量为 10~100,分裂准则为基尼不纯度和信息熵,树的最大深度为 5~100,节点最小分裂样本数为 2~100。经 FHSPSO 优化得到的最优参数为:决策树数量为 76,分裂准则为信息熵,树的最大深度为 16,节点最小分裂样本数为 2。

采用准确率  $A$ 、精确率  $P$ 、召回率  $R$  和 F1 分数  $F$  评估 FHSPSO-RF 模型对交通事故严重程度的预测性能<sup>[20]</sup>。准确率

$$A = (T_p + T_n) / (T_p + T_n + F_p + F_n),$$

式中: $T_p$  为将正类预测为正类的数量, $T_n$  为将负类预测为负类的数量, $F_p$  为将负类预测为正类的数量, $F_n$  为将正类预测为负类的数量。

精确率

$$P = T_p / (T_p + F_p)。$$

召回率

$$R = T_p / (T_p + F_n)。$$

F1 分数

$$F = 2PR / (P + R)。$$

将样本数据按 7:3 划分为训练集和测试集,其中,训练集包含 3 610 条数据,测试集包含 1 548 条数

据。原始数据下 FHPSO-RF 模型的预测结果如表 2 所示。由表 2 可知: FHPSO-RF 模型对测试集中重伤事故召回率仅为 21.74%, 表明模型在训练过程中倾向于财产损失事故和轻伤事故类别, 该现象主要由训练样本严重不平衡导致<sup>[21]</sup>, 而本文重点关注重伤事故和死亡事故等少数类别的准确识别。为解决数据不平衡的问题, 采用合成少数类过采样技术 (synthetic minority oversampling technique, SMOTE)<sup>[22-23]</sup>, 增加重伤和死亡事故的样本数, 使死亡、重伤、轻伤和财产损失事故 4 类样本数均为 3 500 条, 改善样本分布均衡性。SMOTE 处理后 FHPSO-RF 模型的预测结果如表 3 所示。由表 3 可知: 过采样后重伤、死亡事故的召回率显著增大, FHPSO-RF 模型的整体性能更均衡, 说明通过 SMOTE 增加少数类样本数, 能为模型提供更充分的少数类特征信息, 有效提高对严重事故类别的识别能力。

表 2 原始数据下 FHPSO-RF 模型的预测结果

数据集	交通事故 严重程度类别	P/%	R/%	F/%	A/%
训练集	财产损失事故	96.53	95.30	95.91	95.72
	轻伤事故	95.36	96.67	96.01	
	重伤事故	88.64	75.00	81.25	
	死亡事故	87.78	92.94	90.29	
测试集	财产损失事故	95.14	94.70	94.92	94.06
	轻伤事故	94.28	95.46	94.87	
	重伤事故	62.50	21.74	32.26	
	死亡事故	51.85	70.00	59.57	

表 3 SMOTE 处理后 FHPSO-RF 模型的预测结果

数据集	交通事故 严重程度类别	P/%	R/%	F/%	A/%
训练集	财产损失事故	99.95	85.06	91.91	85.86
	轻伤事故	81.63	75.45	78.42	
	重伤事故	77.74	84.62	81.03	
	死亡事故	86.60	98.40	92.12	
测试集	财产损失事故	100.00	85.15	91.98	85.05
	轻伤事故	80.19	72.89	76.37	
	重伤事故	75.42	83.20	79.12	
	死亡事故	86.84	98.68	92.38	

基于 SMOTE 过采样后的数据, 将 FHPSO-RF 模型与支持向量机 (support vector machine, SVM)、K 近邻 (K-nearest neighbors, KNN) 和逻辑回归 (logistic regression, LR) 模型进行预测性能对比, 结果如表 4 所示。

由表 4 可知: FHPSO-RF 模型对交通事故严重程度的预测效果最好, 准确率、精确率、召回率、F1 分数均高于其他 3 个模型; KNN 模型和 SVM 模型预测性能相对较好, 各指标均大于 82.00%; LR 模型的预测性能最差, 各指标仅约为 72.00%。原因是 LR 属于线性分类器, 其决策边界为线性形状, 默认每个特征对结果的影响呈线性且相互独立, 无法捕捉特征间的相互作用或非线性影响。FHPSO-RF 模型基于 RF 模型构建, RF 作为一种集成学习方法, 通过构建多棵决策树并集成输出结果, 能对非线性关系进行建模, 决策树本身具有递归划分特征空间的能力, 不依赖线性假设, 能更灵活地学习数据中的复杂模式, 提高对交通事故严重程度这类非线性问题的预测性能。

#### 2.4 SHAP 结果分析

SHAP (Shapley additive explanations)<sup>[24-25]</sup> 分析是一种基于博弈论沙普利值的机器学习模型解释方法。该方法为每个特征对模型预测的贡献提供统一且可比的度量, 适用于线性模型、树模型及深度神经网络等多种机器学习模型。

不同交通事故严重程度类别的 SHAP 摘要图如图 3 所示。图 3 中每个点代表样本特征的 SHAP 值, 其水平位置表示该特征 SHAP 值的正负与大小: 大于 0 表示特征对预测该类事故有正向贡献, 小于 0 表示有负向贡献。点的颜色代表该样本中此特征值的相对大小: 红色表示特征值较大, 蓝色表示特征值较小。通过 SHAP 摘要图可直观识别哪些特征在驱动模型做出特定分类决策时起关键作用, 以及特征值的影响方向。

表 4 不同模型预测性能对比

模型	A/%	P/%	R/%	F/%
FHPSO-RF	85.05	85.68	85.05	85.03
SVM	82.14	82.81	82.14	82.12
KNN	83.05	83.29	83.05	82.97
LR	72.14	72.88	72.14	72.13

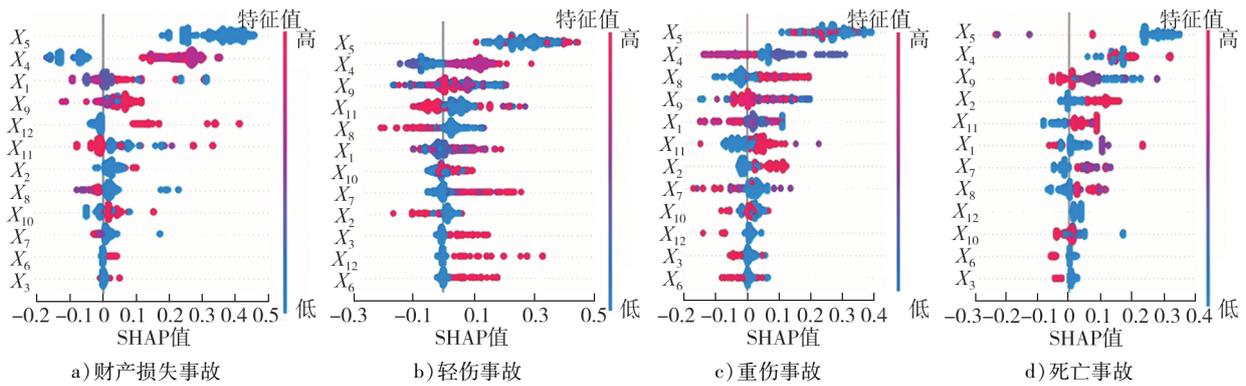


图3 不同交通事故严重程度类别的 SHAP 摘要图

由图3可知:1)在所有交通事故类型中,受伤人数 $X_5$ 和车辆数量 $X_4$ 均为最关键驱动因素,SHAP值较高,表明二者对交通事故严重程度具有显著正向影响,符合交通工程的基本规律,交通事故涉及的人员与车辆规模越大,潜在危害越高。2)在财产损失事故中,除车辆数量 $X_4$ 和受伤人数 $X_5$ 外,是否碰撞路边停放车辆 $X_{12}$ 的低值(即未碰撞)对降低经济损失具有明显作用,这说明此类交通事故多发生在动态交通环境中,而静态障碍物介入可能进一步增加维修成本。3)在轻伤事故中,交叉口类型 $X_{11}$ 的高值(如匝道连接处)显著增加交通事故风险,反映复杂交通节点易引发轻微冲突;光照条件 $X_8$ 的低值(日光)也呈较强影响,这可能与白天交通流量大、出行活跃导致事故暴露机会增加有关,而非光照本身直接提高轻伤概率。4)在重伤及死亡事故中,行人数量 $X_2$ 、碰撞类型 $X_9$ 及受伤人数 $X_5$ 共同构成核心风险组合,行人参与度较高的场景中,由于行人缺乏有效保护,一旦发生碰撞极易导致严重后果,正面撞击等高冲击碰撞类型进一步加剧伤害程度;道路交汇类型 $X_{11}$ 的高值也呈正向影响,说明复杂交通环境是重大交通事故的重要影响因素。

### 3 结论

本文通过构建基于觅食生境选择粒子群算法优化随机森林算法的FHPSO-RF模型,选取美国华盛顿州西雅图市2022年1月至2023年2月期间的交通事故数据为研究对象,对交通事故严重程度进行预测与分析。

1)原始数据中重伤、死亡事故类别数较少,针对样本数据不均衡问题,采用合成少数类过采样技术扩充少数类样本,重伤、死亡事故的召回率显著增大,显著提高了FHPSO-RF模型对严重事故类别的识别能力,整体性能更均衡。

2)通过将FHPSO-RF模型与支持向量机、K近邻和逻辑回归模型进行预测性能对比可知,FHPSO-RF模型对交通事故严重程度的预测效果最好,准确率、精确率、召回率、F1分数均高于其他3个模型;K近邻和支持向量机模型的预测性能次之,逻辑回归模型的预测性能最差。FHPSO-RF模型通过构建多棵决策树并集成输出结果,能更灵活地学习数据中的复杂模型,提高了对交通事故严重程度的预测准确性。

3)通过SHAP摘要图识别交通事故严重程度的关键驱动因素及其情境依赖特征可知:受伤人数与车辆数量在所有事故类型中均为最显著的正向驱动因素;行人数量、高冲击碰撞类型(如正面撞击)与复杂道路环境(如路口、匝道)是重伤及死亡事故的关键风险组合;财产损失事故与是否碰撞路边停放车辆密切相关。交通事故严重程度由“人—车—路—环境”多维因素耦合驱动,研究结果为精准识别高风险场景、优化交通设计与制定差异化安全干预策略提供了理论依据与实践支撑,为交通事故主动预防与精准防控提供方法参考。

未来研究可引入车速、交通流量等实时动态特征,并探究多影响因素间的交互效应,拓展该模型在更

多交通安全分析场景中的应用。

### 参考文献:

- [1] 何雅琴,段雨阳,王晨. 基于累积 Logistic 模型的行人交通事故严重程度分析及对策研究[J]. 安全与环境学报, 2021, 21(3):1165-1172.
- [2] 胡郁葱,韦湖,曾强. 基于空间广义有序 Probit 模型的高速公路事故严重程度分析[J]. 华南理工大学学报(自然科学版), 2023, 51(1):114-122.
- [3] 张萌萌,李泽文,徐云帆,等. 基于优化 CART 的交通事故严重程度影响因素分析[J]. 重庆交通大学学报(自然科学版), 2023, 42(9):93-98.
- [4] 成卫,马铭炜,张小龙. 基于贝叶斯网络的高速公路交通事故严重程度预测及致因分析[J]. 重庆交通大学学报(自然科学版), 2023, 42(7):87-95.
- [5] 潘义勇,徐翔宇. 数据不平衡的 MobileViT 网络交通事故严重程度预测模型[J]. 吉林大学学报(工学版), 2025, 55(3):947-953.
- [6] VADHWANI D, THAKOR D. An improved XGBoost model to predict the injury severity of person in road crash[J]. International Journal of Crashworthiness, 2025, 30(2):115-124.
- [7] KUMAR GANGADHARI R, KUMAR TAREI P. Qualitative investigation of the influential factors behind unsafe trucking behaviors in India[J]. Transportation Research Record:Journal of the Transportation Research Board, 2021, 2675(1):67-78.
- [8] 乔建刚,范颖蓉,陶瑞,等. 山区高速隧道交通事故严重程度预测及特重大事故决策规则提取[J]. 中国安全生产科学技术, 2024, 20(7):186-192.
- [9] 熊向阳,杨小周,赵银超,等. 基于超参数优化随机森林算法的森林生物量遥感反演[J]. 中南林业科技大学学报, 2024, 44(5):102-111.
- [10] 严格齐,赵婉莹,于镇伟,等. 基于超参数优化算法的随机森林模型预测奶牛呼吸频率[J]. 农业工程学报, 2024, 40(11):195-203.
- [11] 佟敏,崔亚茹,党乐,等. 带参数优化的随机森林变压器油中溶解气体浓度预测[J]. 河北师范大学学报(自然科学版), 2022, 46(5):468-473.
- [12] 陈子廓,史宪睿. 基于觅食生境选择的改进粒子群算法[J]. 辽宁工业大学学报(自然科学版), 2022, 42(1):19-21.
- [13] 陆雅雯,李正权,谭立容,等. 基于遗传粒子群算法的超混沌 S 盒设计[J]. 江苏大学学报(自然科学版), 2024, 45(6):701-708.
- [14] SHI Y, EBERHART R. A modified particle swarm optimizer[C]//IEEE World Congress on Computational Intelligence. Proceedings of IEEE International Conference on Evolutionary Computation Proceedings. Anchorage, AK, USA. IEEE, 1998:69-73.
- [15] 中华人民共和国公安部. 道路交通管理信息采集规范 第3部分:道路交通事故处理信息采集:GA/T 946.3—2011[S]. 北京:中国标准出版社, 2011.
- [16] 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.
- [17] 于琪,马路. 固有匹配对下不同动力类型车辆交通事故乘员伤害严重程度影响因素研究[J]. 综合运输, 2024, 46(9):114-119.
- [18] 司贺兵,赵丹. 低等级道路交通事故严重程度影响因素分析[J]. 中国人民公安大学学报(自然科学版), 2024, 30(3):42-48.
- [19] 秦海波,赵哲源,刘兵,等. 基于皮尔逊系数的防窃电系统建设方案[J]. 电气技术与经济, 2024(6):1-3.
- [20] 张恒,陈焕明,党步伟,等. 基于 Attention 机制和递归思想的 LSTM 车辆轨迹预测[J]. 青岛大学学报(工程技术版), 2024, 39(2):74-82.
- [21] 潘比,余靖华,黄译贤,等. 基于 SMOTE 算法和机器学习模型建立原发性肝癌术后的预后预测模型[J]. 陆军军医大学学报, 2024, 46(19):2236-2240.
- [22] 陈丽萍,王洪海,何舒平. 一种基于数据分布的不平衡数据过采样方法[J]. 安徽大学学报(自然科学版), 2024, 48(5):26-36.

- [23] 李爱华,刘婉昕,陈思帆,等. 面向不平衡数据的 SMOTE-BO-XGBoost 集成信用评分模型研究[PP]. 中国管理科学. (2024-10-11)[2024-11-5]. <https://link.cnki.net/doi/10.16381/j.cnki.issn1003-207x.2023.0635>.
- [24] 王健宇,陈献天,焦朋朋,等. 考虑建成环境的交通事故严重程度致因交互效应研究[J]. 交通运输系统工程与信息, 2024, 24(2):272-280.
- [25] 孙国锋,景云,李和壁,等. 基于可解释机器学习框架的列车乘车区段客流分布预测方法[J]. 交通运输系统工程与信息, 2024, 24(2):249-262.

## Traffic accident severity prediction based on FHSPSO-RF

WANG Shujiao, HU Xuelong, TANG Lei

School of Mining Engineering, Anhui University of Science & Technology, Huainan 232001, China

**Abstract:** To enhance the accuracy of traffic accident severity prediction, an FHSPSO-RF model is developed, where foraging habitat selection particle swarm optimization (FHSPSO) is used to optimize key hyperparameters of a random forest (RF). Using traffic accident data from Seattle, Washington, from January 2022 to February 2023, twelve features are selected. The synthetic minority oversampling technique is applied to increase the samples of severe-injury and fatal crashes and improve class balance. The FHSPSO-RF model is compared with support vector machine (SVM), K-nearest neighbors (KNN), and logistic regression (LR). Shapley additive explanations (SHAP) are used to interpret how each feature influences severity. Results indicate that after oversampling, the recall of severe-injury and fatal crashes increases significantly, and the FHSPSO-RF model achieves more balanced overall performance. The model attains higher accuracy, precision, recall, and F1 score than the three benchmarks, yielding the best severity prediction. Across all crash types, the numbers of injuries and vehicles are the most influential drivers with significant positive effects on severity. The number of pedestrians, high-impact collision types such as head-on crashes, and complex road environments such as intersections and ramps form key risk combinations for severe-injury and fatal crashes. Property-damage-only crashes are closely related to whether a parked roadside vehicle is struck. The FHSPSO-RF model demonstrates strong predictive performance and interpretability, providing support for crash risk prediction and prevention-control decision making.

**Keywords:** traffic accident; severity prediction; FHSPSO; RF

(责任编辑:赵玉真)