

doi: 10.3969/j.issn.1671-7775.2023.02.002

开放科学(资源服务)标识码(OSID):



考虑严重程度的城市道路交叉口 事故密度空间分布

张文会¹, 王美娜², 强添纲¹

(1. 东北林业大学 交通学院, 黑龙江 哈尔滨 150040; 2. 华设计集团股份有限公司, 江苏 南京 210014)

摘要: 为探究城市道路交叉口交通事故的空间分布特征, 找出影响事故严重程度的关键因素, 基于哈尔滨市城市道路交通事故数据库, 获得有效交叉口事故样本 1 758 起, 并提取 9 类事故特征. 采用密度分析方法, 分别获得考虑路网密度、交叉口密度、严重程度的事故密度空间分布特征, 并将其可视化显示. 将事故严重程度分为死亡事故和非死亡事故 2 类, 选择事故严重程度预测效果最好的随机森林算法分别对整个城市区域、低密度区域和中高密度区域影响交叉口事故严重程度的因素排序. 结果表明: 考虑路网密度、交叉口密度和事故严重程度时, 交叉口事故空间分布呈现一定的差异性; 对于整个城市区域、低密度区域和中高密度区域, 季节和天气情况均为事故严重程度的主要影响因素, 交叉口类型、事故形态和时段分别为 3 类区域的主要影响因素.

关键词: 交通事故; 交叉口; 空间分布; 随机森林; 路网密度

中图分类号: U491.31 **文献标志码:** A **文章编号:** 1671-7775(2023)02-0133-09

引文格式: 张文会, 王美娜, 强添纲. 考虑严重程度的城市道路交叉口事故密度空间分布[J]. 江苏大学学报(自然科学版), 2023, 44(2): 133-141.

Spatial distribution of traffic accident density at urban road intersection considering severity

ZHANG Wenhui¹, WANG Meina², QIANG Tiangang¹

(1. School of Traffic and Transportation, Northeast Forestry University, Harbin, Heilongjiang 150040, China; 2. China Design Group Co., Ltd., Nanjing, Jiangsu 210014, China)

Abstract: To explore the spatial distribution patterns of traffic accidents at urban road intersections and find out the key factors affecting the severity of accidents, 1 758 valid intersection accident samples and 9 categories features were extracted from the traffic accident database of Harbin. The spatial distribution of intersection accidents was visualized, and the density analysis algorithm was used to obtain the spatial distribution characteristics of intersection accidents considering the road network density, intersection density and severity respectively. The accident severity was divided into fatal accidents and non-fatal accidents. The random forest was used to rank the significant factors affecting the accident severity in the whole urban area, low density area and medium-high density area. The results show that the spatial distribution of intersection accidents differs considering road network density, intersection density and severity respectively. Season and weather are the significant factors for the whole urban, low density and

收稿日期: 2021-06-05

基金项目: 国家重点研发计划项目(2017YFC0803901); 中央高校基本科研业务费专项资金资助项目(2572021DT09); 黑龙江省重点研发计划项目(JD22A014)

作者简介: 张文会(1978—), 男, 黑龙江哈尔滨人, 博士, 副教授(rayear@163.com), 主要从事交通安全研究.

王美娜(1995—), 女, 黑龙江宾县人, 硕士研究生(mn55177@163.com), 主要从事交通信息工程及控制研究.

medium-high density area, while intersection type, accident type and time of day are the respectively significant factors for the three types of areas.

Key words: traffic accident; intersection; spatial distribution; random forest; road network density

城市道路交叉口的交通行为较复杂,是交通事故的多发点,但城市不同区域的交叉口事故频度、事故类型和严重程度也呈现不同的特点.基于交通事故统计数据,深入分析城市行政区域交叉口事故空间分布特征、探究影响事故严重程度的突出因素,有助于交通管理部门直观了解城市交通事故分布规律,以采取有针对性的交通安全管理措施.

国内外学者对于交通事故空间分布特征的研究较多,部分学者利用地理信息系统(geographic information system, GIS)等空间分析技术确定事故热点区域,预测可能发生事故的区域,寻找特定集群特征^[1].也有学者利用热点分析、密度分析和聚类分析等方法^[2],从行政区划^[3]和道路网络^[4]层面研究城市道路交通事故时空分布特性.还有学者利用数理统计和机器学习算法来辨识事故多发路段,首先将交通事故划分事故类型,采用统计方法挖掘违法行为所导致的交通事故高发区域^[5]或利用机器学习和深度学习算法预测事故黑点等^[6].

对于交通事故严重程度影响因素辨识方面,多采用集成学习算法^[7]或机器学习算法^[8]进行特征提取,通过将交通事故数据与气象、人口、土地等数据库结合,利用广义线性混合模型^[9]、二元逻辑回归、决策树^[10]、极端梯度提升(extreme gradient boosting, XGBoost)^[11]等方法,分析影响城市道路交通事故、追尾交通事故^[9]、交织区交通事故^[10]和高速公路交通事故^[11]等严重程度的突出因素.

上述文献多对城市道路交通事故展开研究,对交叉口事故空间分布研究较少,结合交叉口密度分析事故空间分布特征的更少.笔者应用GIS分析城市道路交叉口事故空间分布特征,并将其与事故严重程度相关联,利用随机森林(random forest, RF)、极端梯度提升和轻量化梯度促进机(light gradient boosting machine, LightGBM)3种决策树集成算法,深度挖掘影响交叉口事故严重程度的突出因素,从不同角度直观显示城市道路交叉口事故分布特征.

1 交叉口事故数据处理

1.1 交通事故数据来源

利用哈尔滨市2014—2018年城市道路交

故数据库,其中,主城区事故样本13 284起,去除非交叉口数据、不完善数据和明显错误的的数据,共选取交叉口事故样本1 758起,占比13.23%.

为进一步探究城市道路交叉口交通事故空间分布特征,根据数据库记录的地理位置,利用地图工具定位每起事故经纬度坐标,结合GIS技术在ARCMAP中完成交叉口事故的空间分布描述.定位后的哈尔滨市主城区道路交叉口事故空间分布如图1所示.根据哈尔滨市行政区划进行道路交叉口事故数据统计如下:平房区36起、南岗区343起、道里区538起、道外区439起、松北区170起、香坊区232起.可见,发生在道里区、道外区和南岗区交叉口的交通事故占比较高,分别为30.60%、24.97%和19.51%.

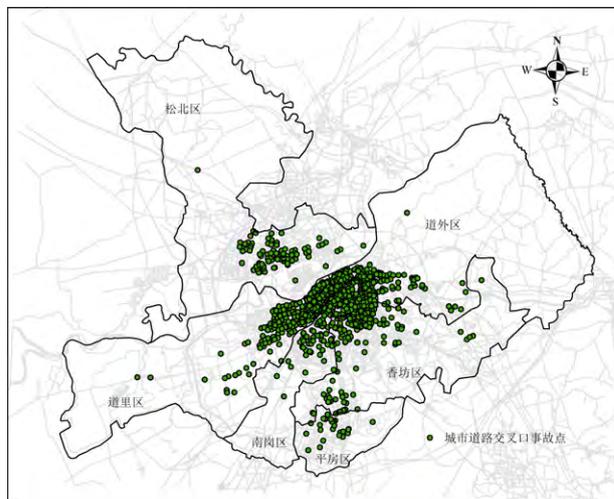


图1 哈尔滨市道路交叉口事故空间分布

1.2 事故特征变量提取

基于哈尔滨市道路交通事故数据库的属性,提取的特征变量包括发生时段、发生星期、发生季节、事故形态、交叉口类型、信号控制方法、发生位置、涉及货车和天气情况9个,将特征变量数字化,结果如表1所示.

表1 事故数据特征变量

特征变量	特征变量说明	频数/ 次	占比/ %
发生时段	夜间非高峰时段 (19:00—次日7:00) = 1	601	34.19
	白天非高峰时段 (9:00—17:00) = 2	709	40.33
	高峰时段 (7:00—9:00 和 17:00—19:00) = 3	448	25.48

续表

特征变量	特征变量说明	频数/ 次	占比/ %
发生星期	工作日 = 0	1 287	73.21
	周末 = 1	471	26.79
发生季节	冬季 = 1	413	23.49
	春季 = 2	433	24.63
	夏季 = 3	556	31.63
	秋季 = 4	356	20.25
事故形态	单车事故 = 1	6	0.34
	车辆间碰撞 = 2	664	37.77
	人车事故 = 3	693	39.42
	摩托车/电动车/三轮车 = 4	395	22.47
交叉口类型	十字路口 = 1	968	55.06
	T 形交叉口 = 2	441	25.09
	X 形交叉口 = 3	186	10.58
	Y 形交叉口 = 4	82	4.66
	其他 = 5	81	4.61
信号控制方法	无信号控制 = 0	595	33.85
	有信号控制 = 1	1 163	66.15
发生位置	非人行横道 = 0	1 097	62.40
	人行横道事故 = 1	661	37.60
涉及货车	其他事故 = 0	1 476	83.96
	货车/挂车/牵引车事故 = 1	282	16.04
天气情况	晴天 = 1	574	32.65
	阴天 = 2	633	36.01
	雪天、雨天、雾天 = 3	551	31.34

2 研究方法

2.1 技术路线

采用 GIS 密度分析法划分城市道路交叉口事故密度区域,再结合交叉口事故数据集提取的事故特征变量,利用 RF、XGBoost 和 LightGBM 对事故严重程度进行预测。经过模型对比,选择预测效果较好的模型识别影响事故严重程度的关键因素,技术路线如图 2 所示。

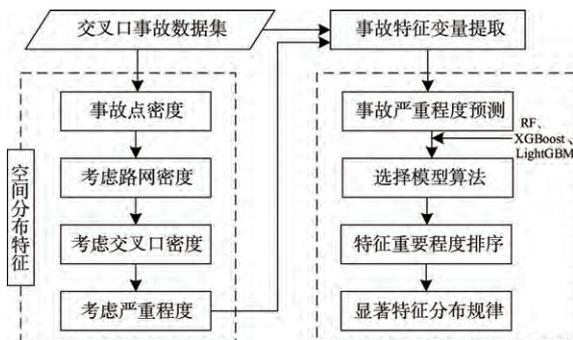


图 2 技术路线

2.2 密度分析法

为精确划分城市道路交叉口事故密度区域,基于 GIS 中的点密度和线密度确定交叉口事故点密度

和路网密度.点密度分析原理是计算单位面积范围内数据点的数量,线密度分析原理则是计算单位面积范围内线段的长度。

GIS 软件通常采用的密度计算方法是邻域法,将城市划分为若干个边长为 d 的小型正方形元胞(对应于最终 GIS 地图上的像素单元).元胞 c 所代表的区域内交叉口事故密度为 D_{ca} 、路网密度为 D_{cr} 、交叉口密度为 D_{ci} .设定邻域半径为 r , $N_c(r)$ 是以元胞 c 的中心为圆点 r 为半径的邻域范围内的事故数量, $L_c(r)$ 为相同邻域范围内的道路长度, $I_c(r)$ 为相同邻域范围内的交叉口数量,计算公式如下:

$$D_{ca} = N_c(r) / (\pi r^2), \quad (1)$$

$$D_{cr} = L_c(r) / (\pi r^2), \quad (2)$$

$$D_{ci} = I_c(r) / (\pi r^2). \quad (3)$$

设元胞 c 在邻域范围内第 m 起事故的严重程度为 x_m , $m = 1, 2, \dots, N_c(r)$, 则元胞 c 的事故严重程度密度值为

$$D_{cs} = \sum_{m=1}^{N_c(r)} x_m / (\pi r^2). \quad (4)$$

2.3 模型算法

2.3.1 RF 算法

RF 算法是一种有监督的数据挖掘算法,其存在固有随机性.该模型每次给予特征变量不同的重要性权重,经过多次模型训练,即每次通过选取一定量的特征变量与上次特征变量中的交集进行保留,以此循环一定次数,从而可以得到对分类任务影响有重要贡献的特征。

RF 算法可应用于事故严重程度与交通流特征之间的关系辨识、交通事故持续时间预测和车辆跟驰行为风险因素提取等.采用随机森林算法的目的是通过了解每个特征在随机森林中的每棵树上做了多大的贡献,得到每个特征对事故严重程度的影响程度,并对特征重要程度排序。

利用 Bootstrap 重抽样方法,从样本数据集中抽取数据,每个样本未被抽取的概率为

$$P = \left(1 - \frac{1}{N}\right)^N, \quad (5)$$

式中: N 为原始数据样本个数。

当 $N \rightarrow \infty$ 时,可得

$$\lim_{N \rightarrow \infty} P = \lim_{N \rightarrow \infty} \left(1 - \frac{1}{N}\right)^N = e^{-1} \approx 0.368. \quad (6)$$

随机森林本身拥有 2 种特征选择方法:平均精度减少(mean decrease accuracy)和平均节点不纯度减少(mean decrease in node impurity).相比 2 种方

法,平均精度减少具有更好的非偏倚性能,所以选择该方法进行特征选择.根据式(6),约有36.8%的数据不会被抽取,这些数据称为袋外数据(out-of-bag, OOB).对于随机森林中的每一棵决策树,使用相应的OOB来计算袋外数据误差,记为 $errOOB1$.随机地对OOB所有样本特征 x 加入噪声干扰,再次计算它的袋外数据误差,记为 $errOOB2$.假设随机森林中有 j 棵树,那么对于特征变量 x 的重要性 $Importance$ 公式为

$$Importance = \sum (errOOB2 - errOOB1) / j. \quad (7)$$

将式(7)作为相应特征变量重要度值,若给某个特征变量随机加入噪声之后,OOB准确率大幅度降低,则说明这个特征变量对于样本的分类结果影响很大,即重要程度较高.

2.3.2 XGBoost 算法

XGBoost是一种集成学习算法,可以进行各个特征变量对模型输出量的重要性排序,便于模型结构选择,提高预测精度,对处理大规模的表格数据具有一定优势.XGBoost的原理是由多颗决策树进行迭代运算,最后对所有的决策树结果进行累计得到最终结果.

对基于树模型的XGBoost算法建模,可以简述为树模型结构和其叶子节点输出值的确定,包括定义树的复杂度、XGBoost中的Boosting Tree模型建立、改写目标函数、树结构的打分函数、正则化学习目标等内容.XGBoost的目标函数为

$$Obj = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k), \quad (8)$$

式中: n 为样本的数量; K 为决策树的数量; l 为误差函数,体现了模型预测值和真实值偏差的程度.

为了防止过拟合,定义了正则化项以惩罚 $\Omega(f_k)$ 复杂模型,其中 $\Omega(f_k)$ 的公式为

$$\Omega(f_k) = \gamma T_k + \frac{1}{2} \lambda \|w_k\|^2, \quad (9)$$

式中: γ 为惩罚系数,取值范围为 $[0, 1]$; T_k 为树中的叶子节点数; λ 为正则项系数; w_k 为第 k 棵树的叶子权重值.

XGBoost算法基于树模型,可以认为是树模型结构和其叶子节点输出值的确定.XGBoost算法利用迭代方式,将弱学习器不断叠加成强学习器,每次迭代产生的残差用于矫正,矫正后的预测器优化指定的损失函数,在损失函数中加入正则项,建立XGBoost模型的目标函数.

2.3.3 LightGBM 算法

轻量级梯度提升机算法是梯度提升决策树(gradient boosting decision tree, GBDT)算法的另一种进化形式,该算法使用深度限制的叶子生长(leaf-wise)策略,从当前叶子节点中找到增益值最大的节点进行分裂,并对树的深度进行限制,防止过拟合,缩短寻找最优深度树的时间.同时保证分裂次数相同的情况下,能够降低误差,得到更高精度.为了节省计算时间,LightGBM使用直方算法、单边梯度抽样(gradient-based one-side sampling, GOSS)算法和互斥特征捆绑(exclusive feature building, EFB)算法来提升运行效率.

在特征维度较高和样本较多的情况下,传统的分类算法运算效率低下,且最终模型预测的准确率得不到保证;新型深度学习算法在处理大数据集时,运算速度缓慢,且模型预测准确率受其参数影响较大.LightGBM算法具有训练速度快、消耗内存少、预测精度高的特点,能够克服两者的缺点,在训练速度和准确率之间达到平衡,适合实时检测系统的需要.

2.4 评价指标

为了评估预测方法的精度,使用如下的评价方法和评价指标:首先,采用预测方法对测试数据集进行训练,进而对交叉口事故严重程度进行预测;再将属于测试数据集的预测结果与实际数据比较;根据比较结果,计算真阳性(TP)、假阳性(FP)、真阴性(TN)和假阴性(FN)4个指标.其中:TP是将正类预测为正类的样本数;FP是将负类预测为正类的样本数;TN是将负类预测为负类的样本数;FN是将正类预测为负类的样本数.

交通事故严重程度预测性能有不同的度量标准,使用精确率(Precision)、召回率(Recall)、准确率(Accuracy)和F1分数(F1-Score)来评估所提出的方法.计算公式如下:

$$Precision = \frac{TP}{TP + FP}, \quad (10)$$

$$Recall = \frac{TP}{TP + FN}, \quad (11)$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}, \quad (12)$$

$$F1-Score = \frac{2 \times Precision \times Recall}{Precision + Recall} = \frac{2TP}{2TP + FP + FN}. \quad (13)$$

3 交叉口事故密度分布特征

3.1 交叉口事故点密度分布

衡量城市道路交通安全水平的一个重要指标是单位面积发生事故的起数,即事故点密度,利用 GIS 中的点密度功能即可计算该值. 为方便事故密度区间分类,采用最大值量纲一化方法,对所得密度值进行量纲一化处理,在密度分布图中选取 0 ~ 0.25 为低密度区域、0.25 ~ 0.50 为中密度区域、0.50 ~ 1.00 为高密度区域. 根据经纬度计算,元胞长度 d 和邻域半径 r 分别为 130、1 100 m,对城市道路交叉口事故频度进行密度分析(不考虑路网/交叉口密度),所得事故点密度分布如图 3 所示.

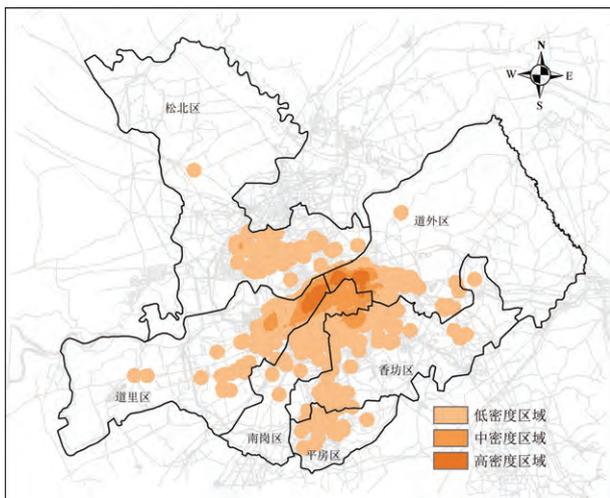


图 3 交叉口事故点密度分布

结合 GIS 分析工具,中密度区域和高密度区域总面积为 77.76 km²,其中:道里区 24.40 km²,占比 31.38%;道外区 20.83 km²,占比 26.79%;南岗区 23.06 km²,占比 29.65%;香坊区 7.07 km²,占比 9.09%;松北区 2.40 km²,占比 3.09%;平房区为低密度区域.

3.2 考虑路网密度的事故点密度分布

以上计算事故点密度分布时没有考虑路网密度,结果不能完全展示区域交通事故空间分布特征.根据文献[4]的分析方法,利用事故点密度与路网密度的比值(即 D_{ca} 与 D_{cr} 之比),深入分析考虑路网密度时交叉口事故分布特征(可理解为单位道路长度发生事故的起数).考虑路网密度的事故点密度分布如图 4 所示,中密度区域和高密度区域的总面积减少为 45.01 km²,其中:道里区 10.37 km²,占比 23.04%;道外区 14.89 km²,占比 33.08%;南岗

区 12.22 km²,占比 27.15%;香坊区 6.51 km²,占比 14.46%;松北区 1.02 km²,占比 2.27%;平房区为低密度区域.考虑路网密度后,无论是整体区域还是各个行政区内,中密度区域和高密度区域的面积均有减少,位置也发生了变化.这是因为城市路网密度高的区域,交叉口数量往往相对较多.比如,哈尔滨市道里区路网密度较高,在考虑路网密度的情况下,中高密度区域占比减少了 8.34%;而道外区路网密度较低,考虑路网密度后,中高密度区域占比增加了 6.29%.

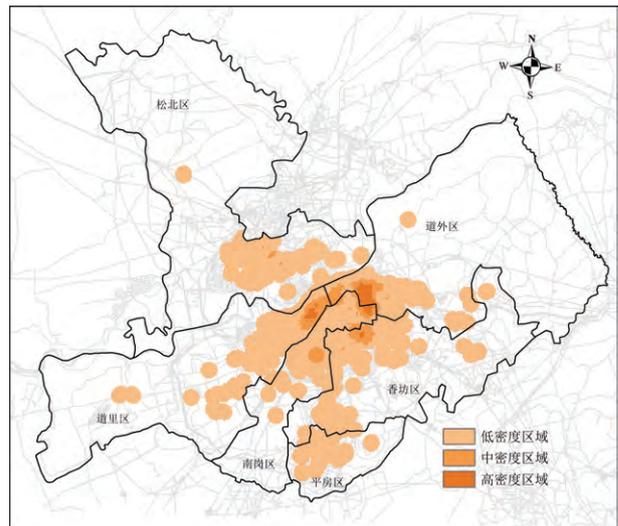


图 4 考虑路网密度的事故点密度分布

3.3 考虑交叉口密度的事故点密度分布

文献[4]采用考虑路网密度的方法分析事故空间分布,获得研究区域单位道路长度交通事故起数.本研究的对象为城市道路交叉口事故,单纯考虑路网密度还不能完全消除交叉口密度对事故空间分布的影响.因此,提出考虑交叉口密度的事故密度分析方法.利用交叉口事故点密度与交叉口密度的比值(即 D_{ca} 与 D_{ci} 之比)表示单位交叉口发生事故起数,消除城市交叉口密度对事故空间分布的影响,可视化结果如图 5 所示,中密度区域和高密度区域的总体面积为 54.66 km²,其中:道里区 11.93 km²,占比 21.83%;道外区 29.44 km²,占比 53.86%;南岗区 6.10 km²,占比 11.16%;香坊区 2.73 km²,占比 4.99%;松北区 4.44 km²,占比 8.12%;平房区 0.02 km²,占比 0.04%.考虑交叉口密度后,中高密度区域分布均有所变化,这是因为交叉口数量较多的区域,事故发生起数也相对较多.比如,哈尔滨市道里区和南岗区,在考虑交叉口密度的情况,中高密度区域占比要低于前两种情

况;而道外区交叉口密度较低,中高密度区域占比显著增加。

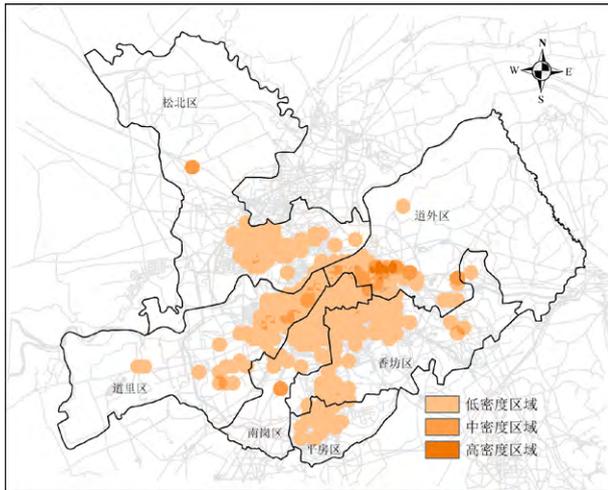


图5 考虑交叉口密度的事故点密度分布

3.4 考虑事故严重程度的事故点密度分布

以上分析交叉口事故空间分布只考虑了事故频度,并未考虑事故严重程度.对于交通管理者来说,交通事故带来的人员伤亡程度,往往更值得关注.因此参考文献[12],并结合交通事故数据库特征,根据权值 w 将事故严重程度分为2个等级,如表2所示.

表2 交通事故严重程度分级

等级	描述	频数/次	占比/%
1	非死亡事故	1 517	86.29
2	死亡事故	241	13.71

表3 哈尔滨各行政区事故中高密度分布数据统计

区域位置	仅考虑事故频度		考虑路网密度		考虑交叉口密度		考虑事故严重程度	
	面积/km ²	占比/%						
总体区域	77.76	100.00	45.01	100.00	54.66	100.00	47.37	100.00
道里区	24.40	31.38	10.37	23.04	11.93	21.83	15.55	32.83
道外区	20.83	26.79	14.89	33.08	29.44	53.86	26.25	55.41
南岗区	23.06	29.65	12.22	27.15	6.10	11.16	4.62	9.75
香坊区	7.07	9.09	6.51	14.46	2.73	4.99	0.95	2.01
松北区	2.40	3.09	1.02	2.27	4.44	8.12	0	0
平房区	0	0	0	0	0.02	0.04	0	0

表3从不同层面反映了哈尔滨市城市道路交叉口事故密度分布情况,考虑事故频度和路网密度的情况下,哈尔滨市道里区、道外区和南岗区中高密度区域占比较高,但综合考虑交叉口密度和事故严重程度后,道里区和道外区中高密度区域占比较高。

为了综合考虑交叉口密度和事故严重程度,将事故严重程度作为权重,计算交叉口事故的加权密度分布,即权值 w 乘以 D_{ca} 与 D_{ci} 之比,即可得到考虑事故严重程度的事故点密度分布,如图6所示。

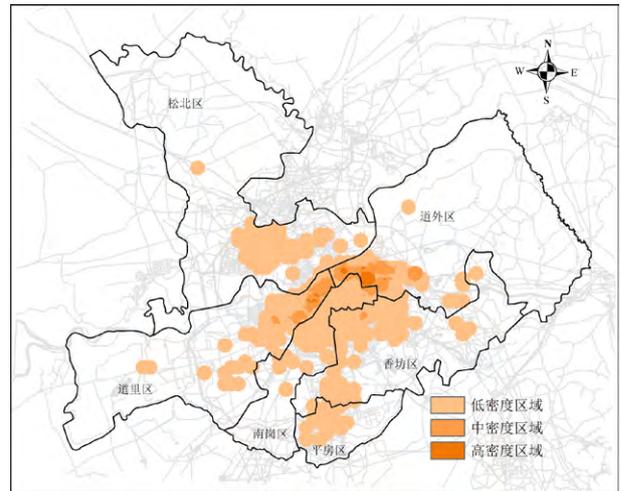


图6 考虑事故严重程度的事故点密度分布

从图6可以看出:考虑事故严重程度后,交叉口事故密度分布又有变化,中密度区域和高密度区域的总体面积为47.37 km²,其中:道里区15.55 km²,占比32.83%;道外区26.25 km²,占比55.41%;南岗区4.62 km²,占比9.75%;香坊区0.95 km²,占比2.01%;松北区和平房区均为低密度区域。

3.5 综合分析

综合以上分析方法,得到哈尔滨各行政区域事故中高密度分布数据统计,如表3所示。

4 交叉口事故严重程度预测和分析

4.1 事故严重程度预测

为探究城市道路交叉口事故严重程度的影响因素,以综合考虑交叉口密度和事故严重程度的分

析结果为基础(见图 6),统计低密度区域和中高密度区域事故数据,将事故密度区域最终划分成城市整体区域、低密度区域和中高密度区域(见表 4),分别进行事故严重程度预测和影响因素辨识。

表 4 事故密度区域划分

序号	城市区域划分	事故数/件
1	城市整体区域	1 758
2	低密度区域	990
3	中高密度区域	768

表 5 模型预测结果评价

城市区域划分	模型	评价指标			
		Precision	Recall	Accuracy	F1-Score
城市整体区域	RF	0.94	0.89	0.85	0.91
	XGBoost	0.93	0.87	0.84	0.90
	LightGBM	0.94	0.82	0.80	0.88
低密度区域	RF	0.94	0.86	0.84	0.90
	XGBoost	0.94	0.84	0.82	0.89
	LightGBM	0.93	0.81	0.79	0.87
中高密度区域	RF	0.95	0.92	0.89	0.94
	XGBoost	0.96	0.90	0.88	0.93
	LightGBM	0.96	0.86	0.84	0.91

4.2 显著性特征

为了辨识影响事故严重程度的突出因素,采用

RF 算法对事故特征变量进行重要程度排序,如图 7 所示。

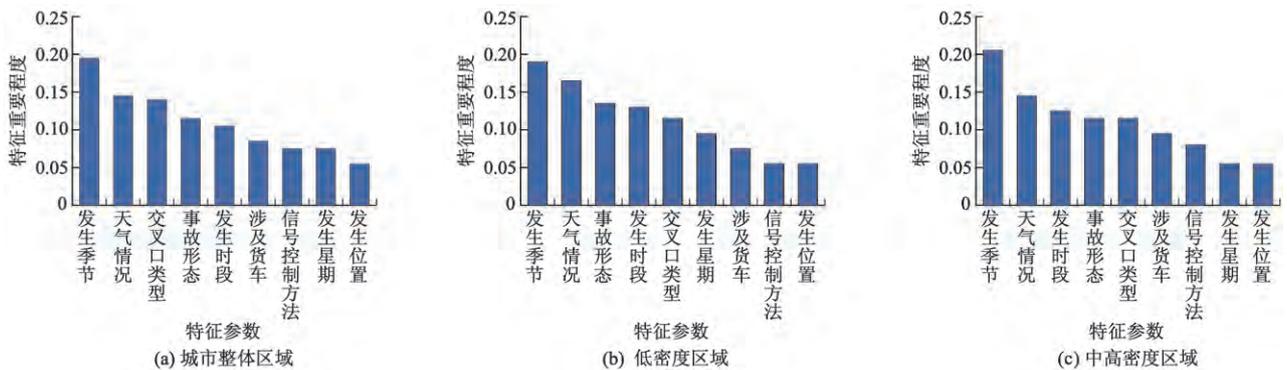


图 7 交叉口事故特征变量重要程度排序

从图 7 可以看出: ① 对于城市整体区域,影响交叉口事故严重程度的重要因素排序为发生季节、天气情况、交叉口类型、事故形态、发生时段; ② 对于低密度区域,影响交叉口事故严重程度的重要因素排序为发生季节、天气情况、事故形态、发生时段、交叉口类型; ③ 对于中高密度区域,影响交叉口事故严重程度的重要因素排序为发生季节、天气情况、发生时段、事故形态、交叉口类型。

4.3 显著性特征与事故严重程度关系

为深入分析事故特征变量对事故严重程度的影响,选择图 7 中 3 个最为重要的特征变量,采用小

提琴图进行分析,如图 8-10 所示, ① 对于城市整体区域的非死亡事故,发生季节主要集中在夏季,天气情况主要为阴天,交叉口类型主要为十字交叉口,其次是 T 型交叉口;对于有人员死亡事故,发生季节主要集中在冬季,天气情况主要为阴天,交叉口类型主要为十字交叉口,其次是 T 型交叉口。 ② 对于城市低密度区域的非死亡事故,发生季节主要集中在夏季,天气情况主要为阴天,事故形态多为车辆间碰撞;对于有人员死亡事故,发生季节主要集中在冬季,天气情况主要为晴天,事故形态多为人车事故。 ③ 对于城市中高密度区域的非死亡

事故,发生季节主要集中在夏季,天气情况主要为阴天,发生时段集中于白天非高峰时段。对于有人

员死亡的事故,发生季节主要集中在夏季,天气情况主要为阴天,主要发生时段为夜间非高峰时段。

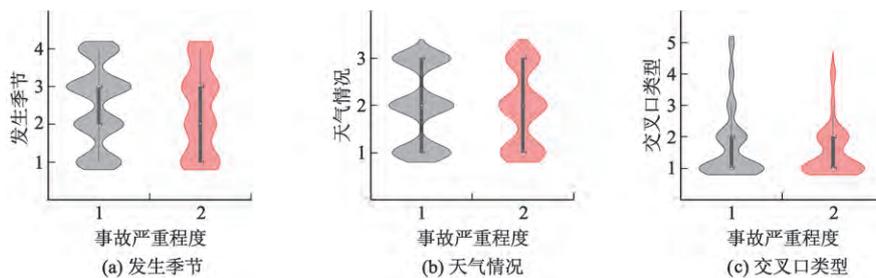


图8 城市整体区域,事故特征变量与事故严重程度关系

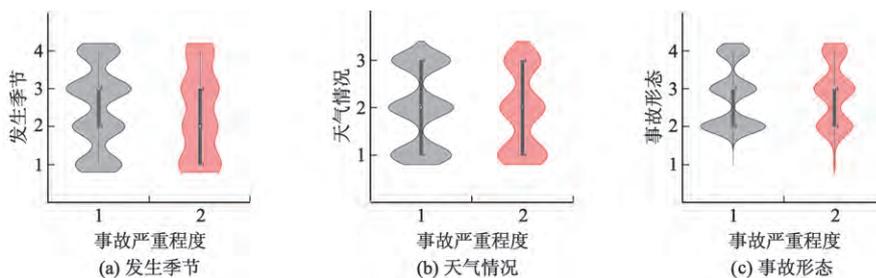


图9 低密度区域,事故特征变量与事故严重程度关系

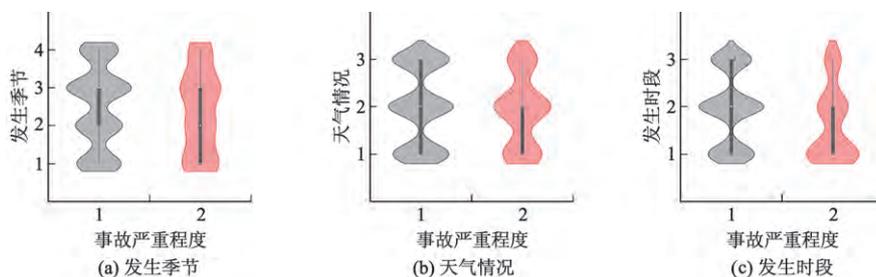


图10 中高密度区域,事故特征变量与事故严重程度关系

5 结论

1) 应用GIS中的密度分析方法,获得城市道路交叉口事故点密度、考虑路网密度和考虑交叉口密度的事故空间分布特征,从不同层面直观显示事故分布规律。与单纯考虑交叉口事故频数分布相比,考虑路网密度和交叉口密度后,事故分布特征存在一定的差异,为交通安全管理提供更为丰富的信息。

2) 将事故严重程度分为非死亡事故和死亡事故2类,分配不同的权重系数,获得综合考虑交叉口密度和严重程度的事故空间分布。哈尔滨市南岗区交叉口事故起数虽然较多,但非死亡事故较多,而道里区和道外区事故严重程度相对较高。考虑事故严重程度的情况下,道里区和道外区中高密度区域

占比较高。

3) 将城市划分为事故低密度区域和中高密度区域,再结合城市整体区域交叉口事故数据,分别采用RF、XGBoost和LightGBM算法进行事故严重程度预测,发现RF算法的 $F1-Score$ 和置信度较高。采用RF算法对影响事故严重程度的特征变量排序,分别得到城市整体区域、低密度区域和中高密度区域的突出因素。

4) 采用小提琴图深入分析影响交叉口事故严重程度的突出因素,获得城市整体区域、低密度区域和中高密度区域的重要特征变量。

5) 限于交通事故数据库记录的信息,提取的事故特征变量并不全面,采用的因素分析方法也有待进一步改进。下一步可结合视频图像、街景图片等多源数据,深入分析城市道路交叉口事故致因,为交通安全精细化管理提供更加有效的信息。

参考文献(References)

- [1] OUNI F , BELLOUMI M. Pattern of road traffic crash hot zones versus probable hot zones in Tunisia: a geospatial analysis [J]. *Accident Analysis & Prevention* , 2019 , 128: 185 - 196.
- [2] LE K G , LIU P , LIN L T. Determining the road traffic accident hotspots using GIS-based temporal-spatial statistical analytic techniques in Hanoi , Vietnam [J]. *Geospatial Information Science* , 2020 , 23(2) : 153 - 164.
- [3] FAN Z Y , LIU C , CAI D J , et al. Research on black spot identification of safety in urban traffic accidents based on machine learning method [J]. *Safety Science* , 2019 , 118: 607 - 616.
- [4] 陆化普, 罗圣西, 李瑞敏. 基于 GIS 分析的深圳市道路交通事故空间分布特征研究 [J]. *中国公路学报* , 2019 , 32(8) : 156 - 164.
LU H P , LUO S X , LI R M. GIS-based spatial patterns analysis of urban road traffic crashes in Shenzhen [J]. *China Journal of Highway and Transport* , 2019 , 32(8) : 156 - 164. (in Chinese)
- [5] 张光南, 钟俏婷, 杨清玄. 交通违法事故时空分布特征及其影响因素—以广州市为例 [J]. *交通运输系统工程与信息* 2019 , 19(3) : 208 - 214.
ZHANG G N , ZHONG Q T , YANG Q X. Temporal-spatial characteristics and influencing factors of at-fault traffic crashes: a case study in Guangzhou [J]. *Journal of Transportation Systems Engineering and Information Technology* , 2019 , 19(3) : 208 - 214. (in Chinese)
- [6] 田准, 张生瑞. 优化经验贝叶斯事故黑点识别与排序方法 [J]. *长安大学学报(自然科学版)* 2019 , 39(5) : 115 - 126.
TIAN Z , ZHANG S R. Identification and ranking of accident black spots using advanced empirical Bayes method [J]. *Journal of Chang'an University(Natural Science Edition)* , 2019 , 39(5) : 115 - 126. (in Chinese)
- [7] 白玉, 魏毅. 重大道路交通事故严重程度影响因素分析 [J]. *交通与运输* 2022 , 38(3) : 22 - 26.
BAI Y , WEI Y. Influencing factors on severity of serious road traffic accidents [J]. *Traffic & Transportation* , 2022 , 38(3) : 22 - 26. (in Chinese)
- [8] SUN Z Y , WANG J Y , CHEN Y Y , et al. Influence factors on injury severity of traffic accidents and differences in urban functional zones: the empirical analysis of Beijing [J]. *International Journal of Environmental Research and Public Health* , 2018 , 15(12) : 2722 - 2738.
- [9] 柳本民, 闫寒. 基于 SVM 事故分类的连环追尾事故影响因素分析 [J]. *交通信息与安全* 2020 , 38(1) : 43 - 51.
LIU B M , YAN H. An analysis of influencing factors of multi-vehicle rear-end accidents based on accident classification of SVM [J]. *Journal of Transport Information and Safety* , 2020 , 38(1) : 43 - 51. (in Chinese)
- [10] MAO X H , YUAN C W , GAN J H , et al. Risk factors affecting traffic accidents at urban weaving sections: evidence from China [J]. *International Journal of Environmental Research and Public Health* , 2019 , 16(9) : 1542 - 1559.
- [11] 张文会, 于秋影, 沈航先. 基于 SEM 的高速公路事故路段行车风险因素辨识 [J]. *森林工程* 2021 , 37(2) : 95 - 103.
ZHANG W H , YU Q Y , SHEN H X. Identification for driving risk factors of expressway accidents section based on SEM [J]. *Forest Engineering* , 2021 , 37(2) : 95 - 103. (in Chinese)
- [12] LEE J , CHAE J , YOON T , et al. Traffic accident severity analysis with rain-related factors using structural equation modeling: a case study of Seoul city [J]. *Accident Analysis & Prevention* , 2018 , 112: 1 - 10.

(责任编辑 贾国方)