

基于移动设备位置数据的油气管道第三方破坏行为识别研究

张行^{1*}, 凌嘉瞳¹, 刘思敏², 董绍华¹

1 中国石油大学(北京)管道技术与安全研究中心, 北京 102249

2 中油国际管道公司, 北京 102206

* 通信作者, zhanghang@cup.edu.cn.

收稿日期: 2021-06-07

中国石油天然气股份有限公司—中国石油大学(北京)战略合作科技专项(ZLZX2020-05)和中国石油科技创新基金项目(2018D-5007-0601)联合资助

摘要 管道第三方破坏是长输油气管道面临的主要风险之一, 第三方破坏活动极强的随机性与不确定性使得相应的防范工作变得非常困难。针对目前人工巡线, 光纤振动监测, 无人机巡线等安全预警技术存在预警不及时、误报、漏报等问题, 结合易获取的具有时空序列的手机位置数据, 本文建立了基于手机位置数据的管道第三方破坏行为识别模型。首先, 通过将手机位置信息进行预处理, 获取更准确的目标管线附近的第三方活动位置信息, 根据位置数据的密度在空间层上对轨迹点进行聚类分析, 提出一种基于时空聚类的停留点识别方法; 结合停留点位置关键特征对其进行语义标记, 并基于TF-IDF规则对停留点异常程度进行计算, 准确提取出管道监控范围内异常停留点; 然后通过对第三方轨迹进行提取与分段, 结合轨迹位置特征完成停留点所在轨迹的邻域搜索, 根据速度、加速度、转角等多个轨迹移动特征计算近邻轨迹分段的行为差异度; 最后基于管道风险特征与第三方人员的行为特征构建管道第三方破坏行为决策树模型, 深入分析各类特征与第三方破坏活动类型的相关性, 实现管道附近第三方破坏行为类别判断。通过搜集的第三方历史特征数据集进行训练测试, 本文建立的识别模型准确率为90.9%, 且对某长输管段附近30天内的移动设备信息进行处理, 依据获得的53994条有效数据对附近第三方活动异常行为进行监测, 结果表明该模型可准确识别出轨迹中的异常行为, 有助于及时发现第三方管道破坏行动, 为智能防范管道第三方破坏维护管道完整性提供了有效依据。

关键词 长输油气管道; 第三方破坏; 位置数据; 安全预警; 异常检测

Identification of oil and gas pipeline third-party damage based on mobile devices location

ZHANG Hang¹, LING Jiatong¹, LIU Simin², DONG Shaohua¹

1 Pipeline Technology and Safety Research Center, China University of Petroleum-Beijing, Beijing 102249, China

2 Sino-pipeline International Company Limited, Beijing 102206, China

Abstract For the long-distance oil and gas pipelines, the third-party damage (TPD) is a main risk, which is randomness and

引用格式: 张行, 凌嘉瞳, 刘思敏, 董绍华. 基于移动设备位置数据的油气管道第三方破坏行为识别研究. 石油科学通报, 2022, 02: 261-269
ZHANG Hang, LING Jiatong, LIU Simin, DONG Shaohua. Identification of oil and gas pipeline third-party damage based on mobile devices location. Petroleum Science Bulletin, 2022, 02: 261-269. doi: 10.3969/j.issn.2096-1693.2022.02.024

uncertainty, and difficult to prevent. At present, the safety early warning technologies such as line patrol, fiber-optical vibration and Unmanned Aerial Vehicle (UAV) line patrol are mainly methods adopted by the TPD. But that has many problems such as untimely warning, false alarm and missed report. Combined with the easily obtained mobile phone location data with time-space sequence, a third-party damage behavior identification model for pipelines based on mobile phone location data was established here. Firstly, the mobile phone location information was preprocessed to obtain more accurate third-party activity location information near the target pipeline. The trajectory points were clustered and analyzed based on the density of data, and a stop recognition method based on spatiotemporal clustering was proposed. The key features of the staying point were semantically marked, and the abnormality degree of the staying point is calculated based on the TF IDF rule to accurately extract the abnormal staying point within the pipeline monitoring range. Secondly, extracted and segmented the third-party trajectory, completed the neighborhood search of the trajectory where the stay point located in accordance with trajectory location characteristics, and calculated the behavior difference degree of the neighbor trajectory segment according to multiple trajectory movement characteristics such as velocity, acceleration and rotation angle. Finally, established the model of pipeline TPD decision tree based on the pipeline risk characteristics, and in depth analysis the correlation between various characteristics and the types of third-party sabotage activities. In the end, used the behavior characteristics of the third party to judge the type of TPD. Through the training and testing of the collected historical characteristic data set of the third-party, the accuracy of the identification model established in this paper is 90.9%, and the mobile equipment information in the vicinity of a long-distance pipeline section within 30 days was processed, and the abnormal activities of nearby third parties were monitored according to the 53994 valid data obtained. The results show that the model can accurately identify abnormal behaviors based on the trajectory, it is helpful for timely detection of TPD damage activities such as private excavation, engineering damage and oil theft by drilling, which provides an effective basis for intelligently PTD damage to the pipeline and maintaining the integrity of the pipeline.

Keywords long distance oil and gas pipeline; third-party damage; location data; safety warning; abnormal detection

doi: 10.3969/j.issn.2096-1693.2022.02.024

0 引言

第三方破坏是管道线路安全面临的主要风险之一,我国 2001 年至 2020 年由第三方破坏引起的管道事故占事故总量的 30% 至 40%。国内外管道安全研究领域学者多年来一直致力于油气管道第三方破坏事故风险的研究,重点开展了基于不确定性的第三方破坏事故可能性分析、第三方威胁事件监测预警、第三方破坏事故后果模拟等研究工作^[1-3]。由于油气管网线路点多线长、人文地质环境复杂,部分地区居民并未正确认识破坏管道的危险后果,仍在管道周边实施占压、开挖等活动,或受利益驱使开展盗油盗气、破坏管道重要设施等违法活动;加之各地区工业化发展需求使得地面施工日渐频繁,各施工企业与管道运营企业之间缺乏沟通,因非管道企业施工造成管道破坏的事故时有发生,故管道保护工作仍面临巨大挑战^[4-5]。

随着信息化技术的发展,庞大的手机用户群提供了大量表示其时空出行序列的手机位置数据,已经成为当前用来感知个人或群体活动规律的重要资源,在生活服务领域中得到了广泛的应用^[6-8]。例如根据居民位置信息可为广告投放筛选出最合适的位置,通过获取用户位置信息向其推荐附近热门商户和服务等。基于位置信息的异常轨迹作为表征用户异常行为的重要

因素之一,使轨迹异常检测成为了当前位置信息相关的研究热点。异常检测算法已经广泛应用于交通领域,例如在同一片海域中识别含异常轨迹的船只^[9];基于出租车行驶轨迹发现绕路欺诈行为,利用车辆 GPS 定位数据对行驶道路拥堵状态进行判断等^[10]。

为解决当前第三方破坏识别中存在的实时性不足、监测范围小、数据匮乏、活动预测难、不确定性强的问题,以位置数据异常轨迹研究为基础,将位置数据引入到管道第三方破坏防范领域,对打孔盗油、私人挖掘、工程建设类型第三方破坏识别展开研究。提出从位置数据中挖掘管道附近用户的行为模式,提取用户的活动规律,为管道第三方破坏活动预警提供新思路。

1 第三方人员活动异常停留点提取

1.1 数据的采集与预处理

用于识别管道第三方破坏行为的移动设备位置信息不涉及用户隐私,将采取隐藏用户真实标识信息的保密措施,通过与通信公司签订数据使用协议,说明数据使用具体用途并提交技术可行性方案,向当地公安机关备案,保证位置数据使用的合法性。位置数据采集方式可通过手机应用程序与运营商数据采集等多

种方式，多样的数据来源保障了管道附近位置信息获取的可行性。

本文手机定位数据来源于移动通信网络与手机终端的交互，该交互过程可分为非周期性和周期性位置更新^[11]。在数据的采集、存储过程中，由于外部自然、人为环境的干扰和移动通信网络自身存在的缺陷会导致定位数据中存在着大量“噪声”数据^[12]，会对定位数据的分析结果产生极大的影响。本文对收集的定位数据预处理操作如下：对字段缺失或取值不在正常范围内的无效数据进行删除；对漂移数据进行纠正或平滑处理；通过对时间间隔的设置进行时间分片，对定位数据进行等时间间隔化处理，使得每条位置数据在时间维度上代表的意义相同，为后续基于聚类算法识别定位数据中停留点提供数据基础。

1.2 基于时空聚类方法的停留点识别

对于油气管道领域，管道路由走向指出了管道位置的分布情况，管道附近用户位置数据标记了用户与管道位置的关系，轨迹停留点识别有助于挖掘位置数据中与用户日常行为特征有关的重要信息。对用户轨迹中的停留点进行提取、处理与分析，可在一定程度上对用户活动规律进行重新刻画^[13-14]。在与管道相关的私人挖掘、工程建设、打孔盗油等第三方破坏活动中，由于破坏活动的实施过程需花费一定时间，其部分移动轨迹一定会表现出停留或在一定区域内移动的状态，故管道附近用户的停留点提取是识别管道异常第三方活动的首要研究内容，本文将停留点定义为用户在管道或光纤两侧一定范围内停留时间超过给定时间阈值的位置。

在对手机位置数据进行预处理后，得到按时间排序的等时间间隔定位数据，手机用户在某个地点停留时间的长短可以根据定位轨迹点在空间上的密度计算得出，但如果仅以空间密度作为唯一依据进行停留点筛选，可能将处于不同时间段但距离相近点聚集在一起，造成停留点的误识别。因此本文利用一种基于点

排序的时空聚类算法识别停留点^[15-16]，首先根据位置数据的密度在空间层上对轨迹点进行聚类，再结合停留点判别时间阈值约束条件，初步实现管道附近区域手机用户停留点的获取。

在基于点排序的时空聚类算法进行停留点识别时，将所采集位置数据用于创建初始样本集D；创建有序队列Q用于保存核心对象及其对应的直接密度可达对象，队列中元素按可达距离顺序从小到大依次进行排列；同时创建结果队列O用于存储已完成访问处理的样本点。基于时空聚类方法的停留点识别步骤如下：首先，从样本集D中随机选取一个核心对象样本点作为研究对象存入结果队列O中，同时搜索该样本点给定邻域半径内所有直接密度可达对象，将所有对象按规则放入有序队列Q中，此时可达距离最小的元素排在队首。其次，从有序队列Q中取出样本点，将其标记为已访问样本点后保存至结果队列O中，并对该点进行核心对象判别，若该样本点为核心对象，则继续搜索其给定邻域半径内直接密度可达点并存储到有序队列Q中，每次插入新样本点到有序队列Q中都按可达距离排序进行位置更新；按照以上步骤对样本集D中所有数据进行处理。为进一步完成结果队列O中的样本点聚类，依次取出结果队列O中样本点p进行判别。首先进行样本点p可达距离与给定半径ε之间的比较，若样本点p可达距离在给定半径ε范围内，则将该点划分到当前点簇中，否则进入下一步判别；第二阶段判别以样本点p核心距离与给定半径ε之间的大小关系为判断依据，若样本点p核心距离大于给定半径ε，则将该点判断为噪声，反之，若样本点p核心距离不大于给定半径ε，则将该点划分到新的聚类中。按照上述判断流程，遍历结果队列O中所有样本点。最后，按位置数据中时间要素对样本数据进行排序更新，生成以位置数据时间戳为横轴，样本数据可达距离为纵轴的排序图。

提取管段附近某一时段手机定位数据进行停留点识别与分析，表1列出了预处理后某手机用户的定位

表1 定位数据
Table 1 List of location data

用户识别码	时间戳	经度	纬度	墨卡托经度	墨卡托纬度
d4f3b0deb34	1 860 478807	8 3.0 847888	4 1.7 271398	9248956.38	5091746.51
d4f3b0deb34	1 860 478927	8 3.0 851647	4 1.7 272255	9248956.38	5091759.24
d4f3b0deb34	1 860 479047	8 3.0 855086	4 1.7 273171	9248956.38	5091772.85
d4f3b0deb34	1 860 479167	8 3.0 857941	4 1.7 273885	9249068.292	5091783.46
d4f3b0deb34	1 860 479287	8 3.0 860677	4 1 727448	9249098.753	5091792.3
.....					

数据。其中,用户识别码是经脱敏处理后的用户标识码,表示用户身份信息,具有唯一性;时间戳指获取位置时的时间信息,已完成等时间间隔处理,同一用户两条位置数据间的时间间隔为2 min,即120 s;经度、纬度是位置数据中直接获取到的信息,为方便距离计算,将位置数据中经度、纬度分别转换为投影坐标下的墨卡托经纬度。

结合第三方破坏停留时间特征,在本停留点识别中约束条件设置如下:时间邻域设为1800 s,距离邻域设为3 m,最小邻域点数MinPts设为15,基于该参数取值进行停留点识别,可达距离排序图如图1所示。

从可达距离排序图可知,样本点最大可达距离10.2 m,大部分样本点可达距离集中在1 m附近,当

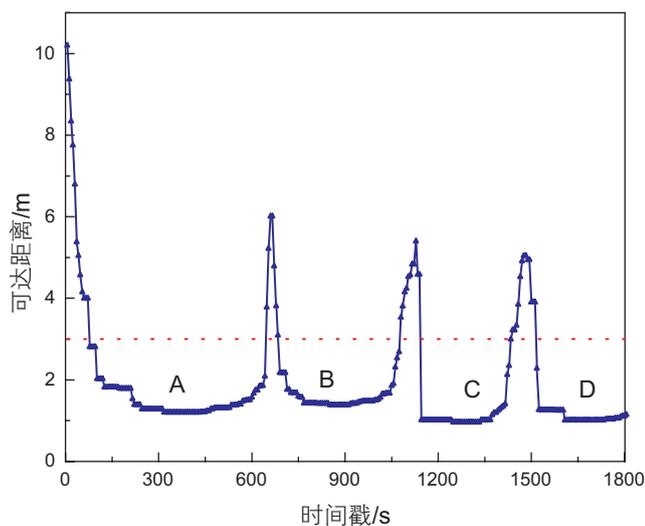


图1 可达距离排序图

Fig. 1 Reachable distance sorting graph

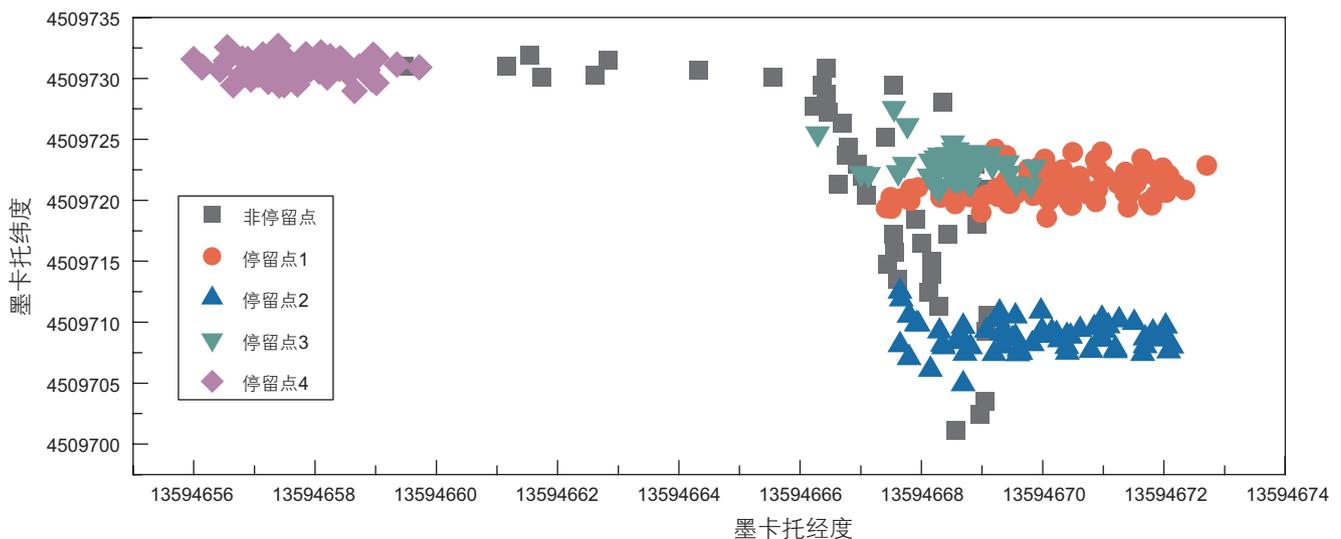


图2 聚类结果图

Fig. 2 Clustering result graph

距离邻域设定为3 m时,该数据集共识别出簇稠密区A、B、C、D共4个停留点。从时间戳信息中可以判断停留时间由长到短依次为A点、D点、B点、C点。可达距离排序图在原位置数据中的聚类结果如图2所示。

2 第三方人员活动轨迹差异度计算

2.1 寻找轨迹邻域

基于由于第三方破坏行为特点的多样性,难以对异常行为轨迹的移动特征进行具体的总结归纳,本文提出利用管道附近移动对象之间的关联性对异常轨迹进行识别。位置特征是轨迹的基本特征之一,包括轨迹分段起始点经纬度,轨迹数据反映了移动对象的位置活动规律,标记了第三方用户在监测范围内的空间位置变化情况,通过位置信息可获取轨迹分段在当前局部空间中的其他近邻轨迹^[17]。第三方人员在管道附近区域的正常活动普遍具有周期性与规律性,如果一个对象与邻域内的对象都发生了偏离,则该对象也一定偏离距离它更远的对象,换言之,一个基本单元的空间特征往往能被邻域内的基本单元所反映。各轨迹分段按位置特征查找邻域,轨迹分段 t_i 的局部空间近邻包括所有与 t_i 距离不超过给定距离阈值 d 的轨迹分段集合。

$$N(t_i) = \{Diff_s(t_i, t_j) < d\} \quad (1)$$

式中, $N(t_i)$ 表示轨迹分段 t_i 的近邻轨迹分段集合, $Diff_s(t_i, t_j)$ 表示轨迹分段 t_i 与 t_j 两条轨迹之间空

间距离,具体描述为轨迹段间垂直、水平、角度距离的综合加权。

2.2 邻域轨迹差异度

异常轨迹通常被描述为违反某类既定规则的事件,或是表现出不同于大多数对象的行为。在不同应用场景中,用户的异常轨迹通常被描述为轨迹异常、地点异常、行动异常等^[18-19]。通过对管道附近行人移动特征的研究与分析,要实现第三方破坏行为的早期预警,找出第三方破坏的迹象和正在实施的第三方破坏行为,准确识别第三方异常轨迹是关键。将相邻区域内移动轨迹进行比较,若某一轨迹与大多数第三方运动轨迹相似,则认为其为正常活动行为,反之判断为异常行为活动。

提取轨迹的速度、加速度、转角作为判断异常轨迹移动特征。速度特征作为移动对象的固有属性之一,表示移动对象运动的快慢程度,在第三方进行异常活动时,相应轨迹通常表现为停留或是以极小的速度移动,利用轨迹特征点中的地理位置标记和时间标记来计算手机用户速度,方向即沿特征点连线方向。加速度特征是移动对象的内在属性之一,表示移动对象速度的变化情况,因为异常行为的出现一般可表现为速度的突变,包括速率和方向,所以加速度是判断异常轨迹的重要因素。转角特征表示移动对象运动方向的变化量,由目标特征点与紧邻前、后时刻特征点连线所构成的角度,轨迹转角的异常变化一定程度上反映了受外界扰动或影响情况,第三方异常行为轨迹与正常行为轨迹存在的位置偏移现象可用转角特征表示。依据移动特征计算轨迹分段行为差异度,寻找出在轨迹邻域内发生移动偏移的轨迹分段。

根据不同移动特征对异常轨迹识别的重要程度分别赋予恰当的权重并进行加权处理。

$$Diff_D(tf_i, tf_j) = \sum_{i=1}^M \omega_i \cdot dis_i(tf_i, tf_j) \quad (2)$$

$$\sum_{i=1}^M \omega_i = 1 \quad (3)$$

$Diff_D(tf_i, tf_j)$ 表示轨迹分段 tf_i 和 tf_j 的行为差异度,以 $\omega_1, \dots, \omega_M$ 分别表示轨迹数据每个特征的权重, $dis_i(tf_i, tf_j)$ 表示任意两条轨迹分段 tf_i 和 tf_j 在特征 i 上的距离。

为了计算轨迹分段的异常程度,将轨迹异常因子(Trajectory Anomaly Factor, TAF)用于表示轨迹分段在其轨迹邻域内移动的异常程度。由于同一用户轨迹会根据不同特征点被划分为多个轨迹分段,所以在计算某一用户轨迹异常因子时,选取最大异常因子作为该

用户最终轨迹行为差异度。

$$TAF(tf_i) = \frac{\sum_{tf_j \in N_{TC}(tf_i)} Diff_D(tf_i, tf_j)}{|N_{TC}(tf_i)|} \quad (4)$$

式中, $TAF(tf_i)$ 为轨迹分段 tf_i 的轨迹异常因子, $Diff_D(tf_i, tf_j)$ 为轨迹分段 tf_i 和 tf_j 的行为差异度, $|N_{TC}(tf_i)|$ 为轨迹分段 tf_i 邻域内轨迹分段个数。

3 基于决策树的异常活动识别

决策树是基于有监督学习进行分类的方法,能够从给定的带有特征和属性标签的样本中分析特征与属性间的映射关系,并以树状图的结构形式呈现决策规则,实现对新样本的正确分类^[20]。基于历史破坏数据先验信息,建立异常活动识别决策树,将第三方人员的行为与管道风险特征结合,对轨迹行为差异度大的异常活动进行第三方破坏类型的判断。

3.1 特征提取

管线监测范围内待识别的破坏行为与历史破坏行为通常具有相似性,且第三方人员的破坏行为与管道风险因素之间存在相关性,本文采用基于数据驱动的方法挖掘第三方人员行为特征。通过第三方破坏风险因素识别与相关信息收集,将管道风险特征与第三方人员的行为特征作为管道第三方异常活动识别决策树模型输入特征^[21]。时间和人数特征按实际数值给出,其余各特征参数对应下表内容给出。

3.2 基于决策树的第三方异常活动识别模型

通过建立一种基于决策树的管道第三方异常活动识别方法,挖掘管道风险特征、第三方活动特征与第三方破坏类型的潜在关系。利用开源python语言sklearn机器学习工具包,调用tree模块,对模型对象进行实例化、训练、测试,完成一棵完整决策树的建立。模型建立步骤如下:

(1)确定输入与输出。将第三方活动时间、人数、位置、当地经济水平、公众宣传效果、巡线质量、安全标志、人员活动频率共8个因素作为输入特征,第三方破坏类型作为标签,根据历史数据建立模型,挖掘各类特征与第三方破坏的关系。

(2)确定最佳节点和最佳的分枝。纯度用于衡量节点处各类标签所占比例,在决策树模型中以不纯度作为衡量最佳节点与分枝的指标,较低的不纯度值表明决策树对训练集的拟合效果越好。在异常活动识别模

表2 特征值表示

Table 2 Characteristic value description

特征名称	特征值	特征描述
位置	1	农田
	2	地下基础设施密集
	3	附近有盗油阀门
	4	其它
当地经济水平	0~100	根据实际情况打分, 见表3
公众宣传效果	0~100	根据实际情况打分, 见表4
巡线质量	0~100	根据实际情况打分, 见表5
安全标志	0	管道附近无明显安全标志
	1	管道附近有清晰安全标志
人员活动频率	1	一级地区, 周围人员活动频率特别低
	2	二级地区, 周围人员活动频率较低
	3	三级地区, 周围人员活动频率较高
	4	四级地区, 周围人员活动频率高

表3 当地经济水平

Table 3 Economic level description

特征名称	特征值	特征描述
当地经济发展水平	70~100	经济发展水平较高, 经济发展速度快
	60~75	经济发展水平中等, 经济发展速度较快
	20~65	经济发展水平落后, 经济发展速度较慢

表4 公众宣传效果

Table 4 Publicity effect description

特征名称	特征值	特征描述
公众安全宣传	80	具有明确公众宣传计划, 定期进行公众宣传并走访
	60	无明确公众宣传计划, 随机进行公众宣传或走访
	40	无明确公众宣传计划, 偶尔进行公众宣传或走访
公众宣传反馈	20	积极
	10	无所谓
	0	抵触

表5 巡线质量

Table 5 Line inspection quality description

特征名称	特征值	特征描述
员工培训	30	具有巡线工定期培训, 并定期对巡线工进行
	20	具有相应巡线工考察, 但未建立培训和审查制度
	10	没有相应的培训和考核制度
巡检效果	40	优: 巡线便道通畅, 无受阻情况
	30	良: 巡线基本通畅, 由于自然原因或人工无法进入但能够方便的观察受阻区域管道情况
	20	中: 巡线受阻, 需要绕行较远或者观察该区域困难
	10	差: 人工受阻, 并禁止进入巡线
巡检频率	30	每日巡查2次及以上
	20	每日巡查1次
	10	每周巡查1次及以下

型中引入信息熵和基尼系数作为衡量指标分别对节点不纯度进行了计算, 结果表明两种不纯度指标下的模型准确率大小基本相同, 最终选择信息熵作为衡量指标。

$$Entropy(t) = -\sum_{i=0}^{c-1} p(i|t) \log_2 p(i|t) \quad (5)$$

$$Gini = 1 - \sum_{i=0}^{c-1} p(i|t)^2 \quad (6)$$

式中, $Entropy$ 为信息熵, c 为叶子节点上标签类别的个数, t 为决策树节点, i 为标签分类, $p(i|t)$ 为标签分类 i 在节点 t 中的占比, $Gini$ 为基尼系数, 该模型中的信息熵值是父节点信息熵与子节点信息熵之差。

(3)确定最大深度值。为使模型具有更好的泛化性, 减轻过拟合对结果的影响, 应对决策树进行剪枝操作。设置树的最大深度值(max_depth)是限制过拟合的有效方式, 通过计算不同深度下的模型拟合效果以确定最佳决策树深度值。结果表明, 当决策树最大深度为4时模型准确率达到最大值, 当深度小于4时决策树欠拟合且未能覆盖重要特征; 当深度大于4时, 多余的分枝使得模型过拟合, 不仅增大了模型计算负担, 而且降低了模型准确率。因此, 确定 max_depth 为4。

(4)决策树剪枝策略优化。设置最小叶子节点样本数($min_samples_leaf, msl$)与最小划分样本数($min_samples_split, mss$)值对决策树进行优化, 最小叶子节点样本数 msl 表示在分枝后的任一子节点都必须包含

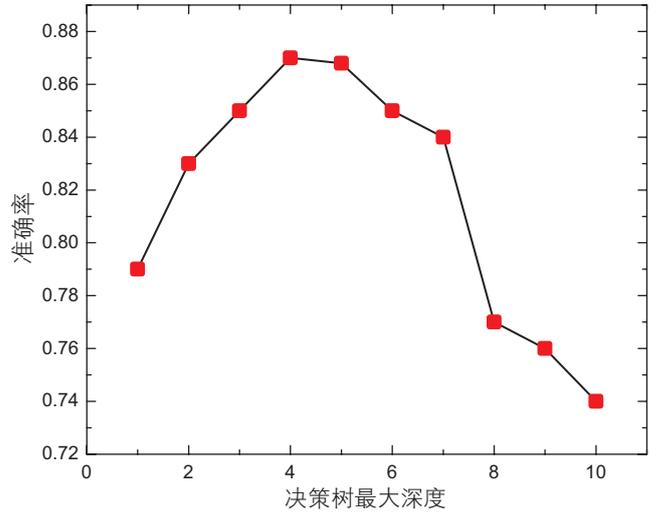


图3 不同深度决策树准确率

Fig. 3 Accuracy of decision trees with different depths

至少 msl 个训练样本; 最小划分样本数 mss 表示当节点包含至少 mss 个训练样本时才允许被分枝。为寻找基于决策树的异常活动识别模型中最小叶子节点样本数与最小划分样本数的最佳组合, 对 0~50 之间的数字组合进行遍历, 最终得出当最小叶子节点样本数为 2, 最小划分样本数为 4 时模型准确率最高。

3.3 模型训练与测试

将所收集到第三方破坏历史特征数据 7/10 划分为训练集, 3/10 划分为测试集, 按照决策树建立步骤, 由训练集数据所建立的异常活动识别决策树如图 5 所

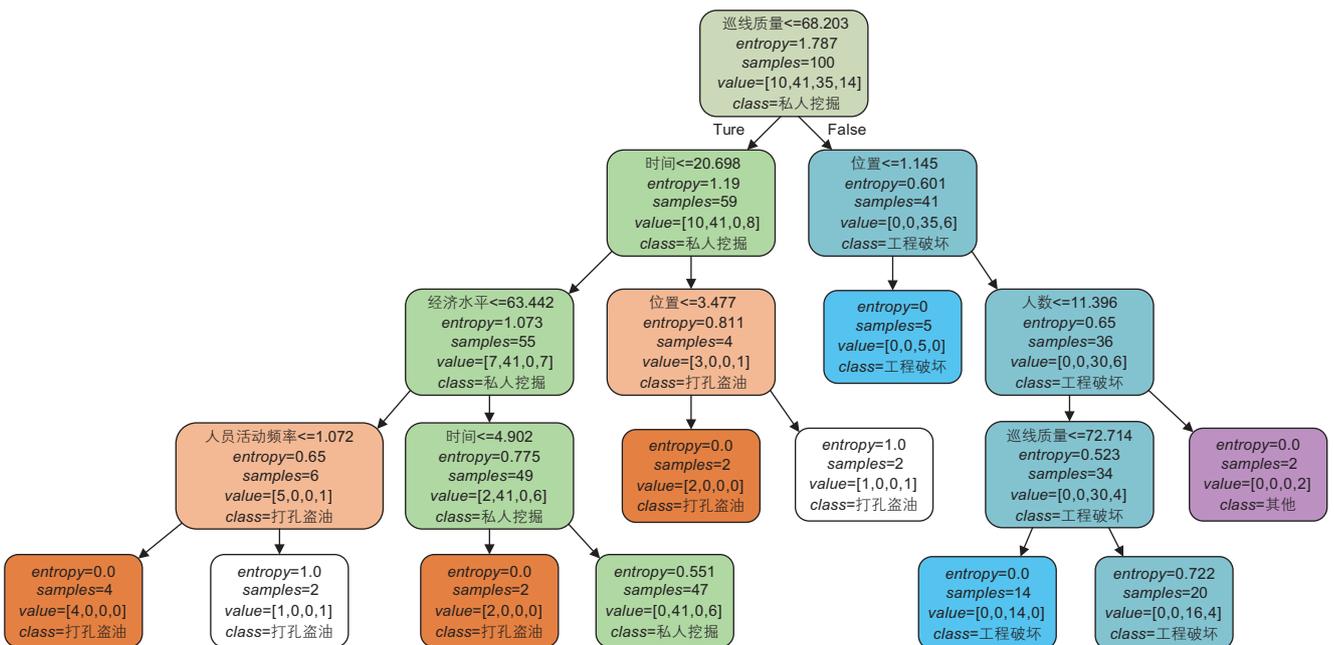


图4 异常活动类型判断决策图

Fig. 4 Abnormal activity type decision diagram

示,该决策树的结构表明了根据各类特征对第三方异常活动类型进行判断的过程,用测试集对模型准确率进行测试,该识别模型准确率为90.9%。

异常活动类型判断决策图如图4所示,为5层决策树,第一层首先对巡线质量特征进行判断,比较其对应特征值与68.203的大小关系并进入决策树第二层,对时间和位置特征进行判断,以此类推,直到判断出最终的活动类型。在该决策树中, *entropy* 为不纯度指标, *samples* 值表示样本个数, *value* 值表示属于不同类型破坏样本个数,如 *value*=(10, 41, 35, 14) 表示属于打孔盗油类别样本数为10个,属于私人挖掘类别样本数为41个,属于工程破坏类别样本数为35个,属于其他类别样本数为14个; *class* 代表最终分

类结果,不同种颜色代表所属不同的破坏类型,其中两个分类结果为打孔盗油的白色方框不纯度指标为1,难以进行判断,其分类结果不准确。

对不同类型的第三方破坏活动判别特征分别如下,打孔盗油破坏的判别含巡线质量、时间、经济水平、位置、人员活动频率共5个特征,私人挖掘破坏含巡线质量、时间、经济水平共3个特征,工程破坏含巡线质量、位置、人数共3个特征。各类特征的重要程度如表6所示,权重越大,对应特征对模型贡献度越大,所提取的8个特征因素中,公众宣传效果和安全标志在该模型中对第三方破坏类型的判断无影响,其余6个影响因素对破坏类型判断的影响权重依次为:巡线质量、时间、经济水平、人数、位置、人员活动频率。

表6 各特征权重

Table 6 Feature weight

特征	时间	人数	位置	经济水平
权重	0.15	0.04	0.02	0.13
特征	公众宣传效果	巡线质量	安全标志	人员活动频率
权重	0	0.64	0	0.02

在该模型中,决策树各分枝判断依据及模型的准确率将根据数据量的变化有所更新,当有更多的历史数据作为训练集输入到模型中时,需要重新调整各特征参数,并对模型进行优化。

4 结论

(1)通过对某长输管段附近第三方活动进行监测,基于时空聚类法提取异常停留点,得到了停留点识别决策图与聚类图,直观地反映了管道附近用户移动规律。

(2)结合轨迹位置特征与速度、加速度、转角多个移动特征的轨迹行为差异度计算,综合体现了轨迹分段在其邻域内移动的异常程度,为管道监测范围内第三方异常轨迹的识别提供重要依据。

(3)针对差异度值较大的用户轨迹,基于管道第三方破坏行为识别决策树模型挖掘第三方风险特征与第三方破坏行为的潜在关系,有助于及时发现私人挖掘、工程破坏和打孔盗油等第三方管道破坏活动,实现第三方破坏智能防范。

参考文献

- [1] 帅健,单克.基于失效数据的油气管道定量风险评估方法[J].天然气工业,2018,38(9):129-138.618-638.[SHUAI J, SHAN K. A quantitative risk assessment method for oil and gas pipelines based on failure data [J]. Natural Gas Industry, 2018, 38(9): 129-138.618-638.]
- [2] 张强,杨玉锋,郑洪龙,等.第三方挖掘作用下管道可靠性评估研究[J].中国安全生产科学技术,2017,13(2):143-147.[ZHANG Q, YANG Y F, ZHENG H L, et al. Study on reliability evaluation of pipeline under the effect of third-party excavation [J]. Journal of Safety Science and Technology, 2017, 13(2): 143-147.]
- [3] CUI Y, NOOR Q, CHAD V M. Bayesian network and game theory risk assessment model for third-party damage to oil and gas pipelines[J]. Process Safety and Environmental Protection, 2020, 134: 178-188.
- [4] 李新宏,韩子月,卢才武,等.老龄城镇油气管道失效风险评估方法[J].中国安全科学学报,2020,30(02):93-98.[LI X H, HAN Z Y. Research on failure risk evaluation methodology of aging urban oil and gas pipeline [J]. China Safety Science Journal, 2020, 30(02): 93-98.]
- [5] GUO X Y, ZHANG L B, LIANG W, et al. Risk identification of third-party damage on oil and gas pipelines through the Bayesian

- network[J]. *Journal of Loss Prevention in the Process Industries*, 2018, 54: 163–178.
- [6] 高强, 张凤荔, 王瑞锦, 等. 轨迹大数据: 数据处理关键技术研究综述[J]. *软件学报*, 2017, 28(4): 959–992. [GAO Q, ZHANG F L, WANG R J, et al. Trajectory big data: A review of key technologies in data processing [J]. *Journal of Software*, 2017, 28(4): 959–992.]
- [7] XU G X, GAO S Y, MAHMOUD D, et al. A survey for mobility big data analytics for geolocation prediction[J]. *Journal of IEEE Wireless Communications Magazine*, 2017, 111–119.
- [8] PO R L. A framework for anomaly detection in maritime trajectory behavior[J]. *Knowledge and Information Systems*, 2016, 47(1): 189–214.
- [9] 马文耀, 吴兆麟, 李伟峰. 船舶异常行为的一致性检测算法[J]. *交通运输工程学报*, 2017, 17(5): 149–158. [MA W Y, WU Z L, LI W F. Conformal detection algorithm of anomalous behaviors of vessel [J]. *Journal of Traffic and Transportation Engineering*, 2017, 17(5): 149–158.]
- [10] 王璞, 鲁恒宇, 谭倩, 等. 手机信令与出租车GPS数据融合车源定位方法[J]. *哈尔滨工业大学学报*, 2018, 50(9): 96–100, 107. [WANG P, LU H Y, TAN Q, et al. A data fusion approach for locating driver sources using mobile phone signaling data and taxi GPS data [J]. *Journal of Harbin Institute of Technology*, 2018, 50(9): 96–100, 107.]
- [11] 李智. 手机基站数据采集与定位系统的设计与实现[D]. 上海交通大学, 2016. [LI Z. Design and implementation of mobile phone base station data acquisition and positioning system [D]. Shanghai: Shanghai Jiaotong University, 2016.]
- [12] 浦佳祺, 陈德旺. 基于最小二乘法和BP神经网络的TOA定位算法[J]. *计算机技术与发展*, 2018, 28(5): 5–8. [PU J Q, CHEN D W. A TOA positioning algorithm based on least square method and BP neural [J]. *Computer Technology and Development*, 2018, 28(5): 5–8.]
- [13] 崔邓, 沈敬伟, 周廷刚. 一种面向移动智能手机用户轨迹停留点提取的时空聚类算法研究[J]. *西南师范大学学报(自然科学版)*, 2016, 41(10): 81–87. [DENG C, SHEN J W, ZHOU T G. On spatial-temporal clustering algorithm based on extraction of user trajectory of smartphone [J]. *Journal of Southwest China Normal University*, 2016, 41(10): 81–87.]
- [14] 李帆, 夏吉喆, 黄赵, 等. 顾及停留位置特征提取的个人位置预测方法[J]. *武汉大学学报(信息科学版)*, 2020, 45(12): 1970–1980. [LI F, XIA J Z, HUANG Z, et al. Predicting personal next location based on stay point feature extraction [J]. *Geomatics and Information Science of Wuhan University*, 2020, 45(12): 1970–1980.]
- [15] MAO J L, WANG T, JIN C Q, et al. Feature grouping-based outlier detection upon streaming trajectories[J]. *Transactions on Knowledge and Data Engineering*, 2017, 29(12): 2696–2709.
- [16] 胡开喜. 移动轨迹的异常检测及其聚类研究[D]. 重庆: 重庆大学, 2018. [HU K X. Study on outlier detection and clustering of moving trajectories [D]. Chongqing: Chongqing University, 2018.]
- [17] 毛嘉莉, 金澈清, 章志刚, 等. 轨迹大数据异常检测: 研究进展及系统框架[J]. *软件学报*, 2017, 28(1): 17–34. [MAO J L, JIN C Q, ZHANG Z G, et al. Anomaly detection for trajectory big data: Advancements and framework [J]. *Journal of Software*, 2017, 28(1): 17–34.]
- [18] 仇功达, 何明, 杨杰, 等. 异常轨迹数据预警与预测关键技术综述[J]. *系统仿真学报*, 2017, 29(11): 2608–2617. [QIU G D, HE M, YANG J, et al. Key technologies of precaution and prediction of abnormal spatial-temporal trajectory: A review of recent advances [J]. *Journal of System Simulation*, 2017, 29(11): 2608–2617.]
- [19] 何明, 仇功达, 周波, 等. 基于改进密度聚类与模式信息挖掘的异常轨迹识别方法[J]. *通信学报*, 2017, 38(12): 21–33. [HE M, QIU G D, ZHOU B, et al. Abnormal trajectory detection method based on enhanced density clustering and abnormal information mining [J]. *Journal on Communications*, 2017, 38(12): 21–33.]
- [20] 刘振宇, 宋晓莹. 一种可用于分类型属性数据的多变量决策树算法[J]. *东北大学学报(自然科学版)*, 2020, 41(11): 1521–1527. [LIU Z Y, SONG X Y. An applicable multivariate decision tree algorithm for categorical attribute data [J]. *Journal of Northeastern University*, 2020, 41(11): 1521–1527.]
- [21] LING J T, ZHANG H, DONG S H, et al. Intelligent prevention method for third-party damage of long-distance pipeline based on mobile devices location information[C]. 13th International Pipeline Conference, Canada, 2020, V2.