

文章编号: 2095-2163(2021)11-0137-07

中图分类号: TM73

文献标志码: A

基于 SmoteEnn_XGBoost 模型的路况感知方法研究

杨黎娜, 姚凯学, 何 勇, 席雷鹏, 刘文才, 赵继露

(贵州大学 计算机科学与技术学院, 贵阳 550025)

摘要: 针对由于异常路况数据量较小, 导致数据集数据不均衡问题, 从而引起路况预测模型准确度不高以及运行效率低等问题, 提出一种基于 SmoteEnn_XGBoost 的路况感知方法。设计研发了路况数据采集系统, 捕获并处理实时的路况数据; 使用 SmoteEnn 算法对数据集进行数据合成处理; 采用 XGBoost 算法对提取的主要特征进行训练和测试, 从而识别出正常路况和异常路况。结果表明, 基于 SmoteEnn_XGBoost 的路况感知方法, 对比目前较常用的分类模型 SVM、逻辑回归、GBDT、随机森林等, 在提高路况分类效果的同时大幅缩短算法的运算时间。

关键词: 路况; 三轴加速度; SmoteEnn 算法; XGBoost 算法

Research on Road Condition Classification Method Based on SmoteEnn_XG Boost Model Consumption

YANG Lina, YAO Kaixue, HE Yong, XI Leipeng, LIU Wencai, ZHAO Jilu

(School of Computer Science and Technology, Guizhou University, Guiyang 550025, China)

[Abstract] Aiming at the problem of data imbalance in the data set due to the small amount of abnormal road conditions, resulting in low accuracy of the road condition prediction model and low operating efficiency, a road condition sensing method based on SmoothEnn_XGBoost is proposed. Designed and developed a road condition data collection system to capture and process real-time road condition data; use the SmoothEnn algorithm to perform data synthesis processing on the data set; use the XGBoost algorithm to train and test the extracted main features to identify normal road conditions and abnormal road conditions. The results show that the road condition perception method based on SmoothEnn_XGBoost, compared with the more commonly used classification models SVM, logistic regression, GBDT, random forest, etc., can greatly reduce the calculation time of the algorithm while improving the effect of road condition classification.

[Key words] road conditions; three-axis acceleration; smoothEnn algorithm; XGBoost algorithm

0 引言

2020年5月,由交通运输部公布的《2019年交通运输行业发展统计公报》显示:中国公路里程快速增长,公路的养护管理压力也随之快速上升,公路养护行业面临任务加重和资金紧张的双重难题^[1]。近年来,由于道路养护的不及时,导致交通事故频发,不仅危害人民群众的生命财产安全、给保险公司增大了赔付压力,也增大了城市交通的通行压力。目前,路况检测主要依靠路政部门周期性地使用三米直尺平整度仪、激光平整度仪和车载式颠簸累计仪等专用道路平整度检测仪器对道路平整度进行检测。这种方式检测精度高,使用门槛高,需要花费较高的人力和物力成本,不能及时的发现道路异常。

通常,路面不平是汽车振动的基本因素。汽车的振动是简谐振动的,而简谐振动的动力学特征是回复力与位移的平方成正比。由牛顿第二定律可知:加速度与位移的平方成正比,因此用加速度来表现汽车的振动大小是可行的^[2]。近年来,基于智能手机传感器的路况识别研究进展很大。文献^[3]基于自主开发的狼獾系统,采用机器学习算法中的K-均值聚类和支持向量机模型进行路况识别,可以识别出坑洞路面和平坦路面。国内学者进行路况识别问题的研究通常是基于智能手机终端的传感器系统进行数据采集并采用统计学方法与支持向量机、K-均值聚类模型相结合的方法^[4-7]。通过分析可知,汽车在不平道路上的振动数据跟语音的振动原理是类似的,故近年涌现出基于语音识别算法的路

作者简介: 杨黎娜(1988-),女,硕士研究生,主要研究方向:物联网与嵌入式;姚凯学(1961-),男,硕士,教授,硕士生导师,主要研究方向:计算机控制、嵌入式技术;何 勇(1974-),男,博士,副教授,硕士生导师,主要研究方向:物联网技术、嵌入式系统;席雷鹏(1996-),男,硕士研究生,主要研究方向:物联网与嵌入式;刘文才(1993-),男,硕士研究生,主要研究方向:物联网与嵌入式;赵继露(1997-),女,硕士研究生,主要研究方向:物联网与嵌入式。

通讯作者: 姚凯学 Email: 734077976@qq.com

收稿日期: 2021-08-04

况分类方法,比如印度旁遮普大学的学者提出了基于 DTW 算法的路况分类算法^[8],东帝汶国立大学的研究人员使用改进后的 MFCC 和改进后的 PLP 算法进行特征提取进而对铺装和未铺装道路进行分类^[9-12]。

在上述研究的基础上,本文提出一种基于 SmoteEnn_XGBoost 的路况感知算法。通过数据采集软件采集到道路的振动数据,将数据中的三轴加速度进行滤波处理,并采用 Smote_ENN 算法解决数据集中数据不均衡问题;提取三轴加速度、三轴加速度的均方根、车速等特征,使用 XGBoost 模型进行路况的训练和识别,从而对具体路况进行分类识别。

1 数据采集

1.1 数据采集设备

本文采用 OPPO Reno 智能手机进行数据采集;处理器为高通 SDM710 八核,运行内存为 6 G,机身内存为 256 G;手机搭载的三轴加速度传感器和陀螺仪为 lsm6ds3c。数据采集软件 RoadDataCollection 采用 Java 语言编写,调用 Android 系统提供的传感器数值变化回调函数,来获取传感器数据;三轴加速度采样频率为 400 Hz,陀螺仪采样频率为 50 Hz, GPS 数据采样频率为 1 Hz。手机搭载车辆为别克君威 2015 款,发动机为 1.6 T 184 马力 L4,前后轴距 2 737 mm。数据采集软件界面如图 1 所示。



图 1 RoadDataCollection 测量界面

Fig. 1 RoadDataCollection Measurement interface

1.2 测量设备摆放位置

考虑到井盖、坑洞的直径较小,故只会对一侧轮胎引起振动,而铁路道口和减速带会对两侧轮胎引起振动,故分别在两个位置采集数据。一是将手机固定在主驾方向盘左侧的空调出风口,二是将手机固定在副驾的座位下方,分别对两个位置的振动数据进行采集。经分析发现,由于车辆经过不平道路时引起的振动很大一部分是由于车轮的刚性运动,导致整个车架的振动,因此这两个位置对数据分类算法的精确度影响较小。本采集路段包含平滑路段、井盖、减速带以及铁路道口等引起车辆颠簸的主要因素。

1.3 采集频率设置

根据奈奎斯特采样定理可知:在进行模数转换的过程中,当采样频率大于信号中最高频率的 2 倍时,采样之后的数字信号可以完整地保留原始信号中的信息。因此,一般在实际应用中要保证采样频率为信号最高频率的 2.56~4 倍。由于汽车的振动来源是发动机、传动系统、车轮旋转引起的振动以及道路不平引起的汽车振动。旋转部件引起的振动频率一般为 20~30 Hz,而道路不平引起的振动形式一般为刚体运动,刚体运动的振动频率一般为 0~15 Hz。本文采用 3 种采集频率进行测试,分别是 10 Hz、50 Hz、100 Hz。经过测试发现,50 Hz 的采样频率相对于另外两种采样频率效果更好。

1.4 数据标记

本文通过设计并开发路况数据采集程序,采集汽车运行过程中的三轴加速度以及车速和 GPS 等信息。由于汽车在经过道路坑洼、井盖、减速带、铁路道口等不平路况时,前后轮压到不平路段时都会有一个加速度变化的最大值,而一般家用汽车前后轮间距为 2.2~3 m,当车速为 10 m/s (36 km/h) 通过不平路段时,两个波峰间的时间为 0.2~0.3 s,故很难利用 APP 按键进行准确标记。因此,本文采用录音的方式,对收集加速度数据的全过程进行录音,然后通过人工标记的方式对采集到的数据集进行标记,以保证训练数据标记的准确性。图 2 是在郑州市高新区枫香街采集到的一段经过 31 个井盖的加速度数据。0 值附近的两条曲线分别是 acc_x 和 acc_y ,代表 x 、 y 轴的加速度值;10 附近的曲线是 acc_z ,代表的是 z 轴的加速度,也是重力加速度。本数据集中,将经过井盖、坑洼、减速带、铁路道路时的加速度数据通过 'type' = 1 进行标记,平顺路段的加速度数据通过 'type' = 0 进行标记。图 2 中明显的

突起,代表经过井盖时的三轴加速度值;中间接近直线的部分,代表的是停车时的三轴加速度值;在两个井盖之间的波动是汽车行驶在正常路段时的三轴加速度值。

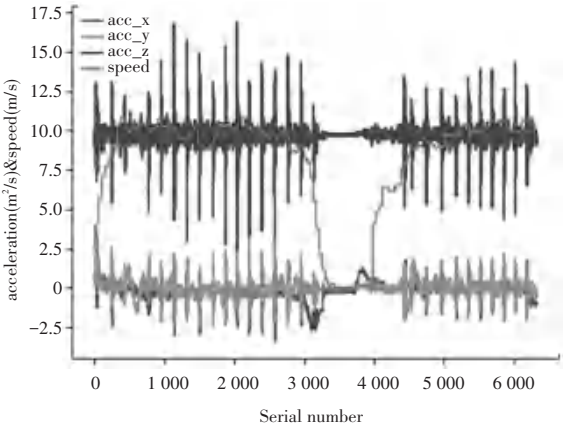


图 2 三轴加速度以及速度曲线

Fig. 2 The collected triaxle acceleration and velocity curves

2 数据处理

2.1 加速度的处理

由于三轴加速度数据有 3 个坐标轴分量,分别是 x 轴、 y 轴、 z 轴。在手机屏幕朝上静止放置时,三个轴的初值分别是 x 轴为 $0 \pm 0.1 \text{ m/s}^2$, y 轴为 $0 \pm 0.1 \text{ m/s}^2$, z 轴为 $9.8 \pm 0.1 \text{ m/s}^2$ 。由于司机在开车期间手机可能以任何角度放置在车上,而这 3 个轴在静止状态下只有重力加速度分量。无论手机如何放置,这 3 个轴的取值虽然会发生变化,但都是重力加速度在各个轴的分量,故本文采用 3 个轴取值的均方根作为加速度的取值。

2.2 滤波

将加速度数据进行低通滤波,去除旋转设备以及共振引起的高频分量。在低通滤波阈值的选取上,本文采用汽车的发动机振动频率作为低通滤波阈值。经采集的汽车打火状态,在未进行任何振动的情况下加速度数据进行快速傅里叶变换发现:车本身的振动频率大部分是由发送机产生的,约为 23 Hz,与理论上四缸发动机的振动频率 20~30 Hz 相符。道路不平引起的轿车振动一般属于刚性运动,产生的振动频率一般在 20 Hz 以下,故通过低通滤波器将采集到的加速度数据进行低通滤波,去除 20 Hz 以上的车自身产生的高频振动干扰。如图 3 所示。

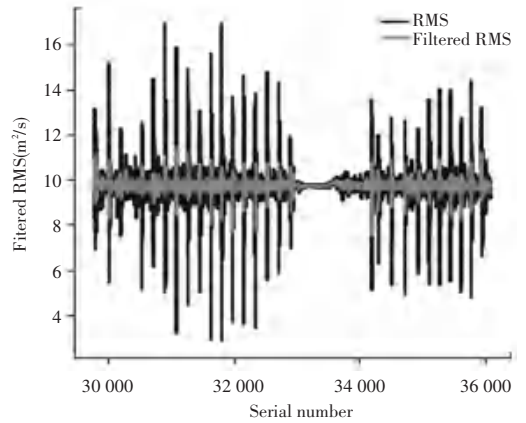


图 3 低通滤波后的 RMS 曲线

Fig. 3 RMS curve after low pass filtering

2.3 数据不均衡问题处理

汽车通过坑洞、减速带等不平路段的时间较短,通过分析采集来的数据发现:平顺路况数据占据了整个数据集 80% 以上的比例,故异常路况和平顺路况数据存在不均衡的问题。数据不均衡会导致数据量大的那部分数据主导整个模型预测的准确度,从而减弱另一部分数据的准确度对整个模型的准确度。解决数据不均衡问题通常从数据和算法两个层面去解决。从数据层面一般采用欠抽样、过抽样和数据合成等方法对数据进行处理从而平衡两类数据数量,而从算法层面一般采用加权、一分类、K-fold 交叉验证和采用适应不均衡样本的模型等方法来解决数据不均衡问题。

SmoteEnn 算法是在 Smote 算法基础上进行过采样,再通过 Enn 算法进行清洗,最终得到数据均衡的样本集。通过这种方式解决数据不均衡问题一方面增加了少数类样本数据量,一方面改善了 Smote 算法增加的少数类样本和多数类样本重复的问题。SmoteEnn 算法的实现步骤如下:

输入 路况数据集 $X = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ 、数据集中少数类异常路况数据样本数量 T 、少数类异常样本需要增加的倍数 $N\%$ 以及少数类异常路况数据的最近邻数据的点的个数 K 。

输出 含有 $(N/100) * T$ 个新合成的数据均衡的路况数据集。

具体步骤:

(1) 判断 N 是否小于 100, 如果小于 100 说明合成样本的数量少于少数类样本的原始数量。这时要在原有的少数类异常路况数据样本中随机选择 $T * N\%$ 个样本作为合成样本的依据。

(2) 对 X 中的少数类异常路况数据样本中的每个数据样本 (x_i, y_i) 求其在 X 中的 K 个最近邻的点

$X_{inn}, nn \in \{1, 2, \dots, K\}$ 。

(3) 从 X_{inn} 中选取任意一个样本 X_{iab} , 计算其余当前遍历样本 X_{icd} 的向量差, 将此差值与 $0 \sim 1$ 之间的随机数相乘, 再加上当前遍历样本, 得到新的少数类异常数据类样本 $x_s = x_{icd} + (x_{iab} - x_{icd}) * rand(0, 1)$ 。

(4) 根据 KNN 算法, 对新生成的数据样本进行预测分类。如果与其 K 近邻样本的大部分类型相同, 则保存该数据, 否则删除。

(5) 重复(3)、(4)步骤, 直到得到 $(N/100) * T$ 个少数类异常路况数据样本, 并将此样本添加入原样本中。

表 1 不同算法处理后的数据集效果

Tab. 1 Comparison diagram of the effect of data sets processed by different algorithms

数据集名称	时间/s	召回率	精确率	F1	AUC	误判率
原始数据	0.238 4	0.965 9	0.965 1	0.964 1	0.867 5	0.034 1
过采样数据	0.250 3	0.958 7	0.958 8	0.958 7	0.958 7	0.041 3
SmoteEnn	0.228 4	0.967 8	0.968	0.967 8	0.968 1	0.032 2
欠采样数据	0.048 9	0.938 7	0.939	0.938 7	0.938 9	0.061

3 路况感知方法

3.1 XGBoost 模型介绍

XGBoost 是极端梯度提升的简称, 2014 年 3 月由陈天奇博士提出的, 在梯度提升框架下实现的机器学习算法。提升方法是比较常用的统计学习方法, 应用广泛并且有效率高, 其将加法模型与前项分步算法相结合。在分类问题中, 通过改变训练样本的权重, 学习多个分类器, 并将这些分类器进行线性组合, 提高分类的性能。XGBoost 算法是在 GBDT (Gradient Boosting Decision Tree) 等提升树的模型上进行了优化, 将 GBDT 用一阶负梯度代替残差的基础上, 加上二阶梯度, 同时在采样时还借鉴了随机森林的 Sub-Sampling 方法, 不仅可以处理更大规模的数据, 而且运行效率进一步提升^[13-15]。

3.2 XGBoost 目标函数优化

XGBoost 算法采用 CART 树作为基分类器, 模型由多棵 CART 树组成, 每棵 CART 树对样本进行预测, 预测结果进行累加求和, 最后通过 Sigmoid 函数将其映射到 $0 \sim 1$ 的区间, 来表示分类结果的概率。其目标函数由损失函数和正则项两部分组成, 具体定义如下:

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

其中, 路况数据集 (X_i, Y_i) , $X_i \in R_m, Y_i \in R$, X_i

通过采用 Smote 过采样算法、SmoteEnn 算法、ClusterCentroids 欠采样算法生成 3 个数据集, 分别将生成的数据集使用 XGBoost 模型进行测试, 结果见表 1。SmoteEnn 算法处理后的数据集在后续的模式训练和预测中能够大幅提升 AUC 的值。AUC 是 ROC 曲线下的面积, 其计算方法同时考虑了分类器对正负例的分类能力, 可以对分类器做出合理的评价。AUC 的值越大, 证明这个模式的分类效果越好。因此, 本算法采用 SmoteEnn 算法进行数据不均衡处理后的数据集, 进一步提高了模式的召回率、精确度、F1 等评价参数, 将 XGBoost 模式的分类效果 AUC 指标提升了 10% 以上。

表示路况数据集的 m 维特征向量; Y_i 表示路况的分类标签; \hat{y}_i 是样本 X_i 预测出的路况类型; f_k 是 CART 树模型。

在路况分类中采用 LogLoss 函数作为损失函数: $L(y, \hat{y}_i) = \sum_i [y_i \ln(1 + e^{-\hat{y}_i}) + (1 - y_i) \ln(1 + e^{\hat{y}_i})]$, 正则项为 L2 正则, 定义为: $\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2$ 。 T 为树的叶子节点个数; ω 为叶子节点的权值向量; γ 和 λ 为要通过目标函数求解的系数。

XGBoost 模型是通过分布前向算法, 对加法模型进行训练, 逐步进行优化, 得到最优的模型。XGBoost 模型的定义为:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), \quad f_k \in F \quad (2)$$

其中, f_k 为 CART 树, K 表示树的棵数, F 为所有可能的 CART 树。

利用加法模型, 分步骤对目标函数公式(1)进行优化。在第 r 轮优化时 XGBoost 模型为:

$$\hat{y}_i^{(r)} = \sum_{k=1}^r f_k(x_i) = \hat{y}_i^{(r-1)} + f_r(x_i) \quad (3)$$

其中, $\hat{y}_i^{(r-1)}$ 为第 $r-1$ 轮优化后的模型, $f_r(x_i)$ 为第 r 轮新加入的 CART 树。由此可知, 第 r 轮的目标函数可表示为:

$$Obj^{(r)} = \sum_{i=1}^n L(y_i, \hat{y}_i^{(r-1)} + f_r(x_i)) + \Omega(f_r) \quad (4)$$

采用 *sigmoid* 函数对预测结果进行映射, 可知损失函数为:

$$L(y, \hat{y}^{(r-1)}) = (1 - y) \log(1 - \text{sigmoid}(\hat{y}^{(r-1)})) + y \log(\text{sigmoid}(\hat{y}^{(r-1)}))$$

利用泰勒公式进行近似求解, 可将目标函数公式(4)变形为:

$$Obj^{(r)} = \sum_{i=1}^n [L(y_i, \hat{y}_i^{(r-1)}) + g_i f_r(x_i) + \frac{1}{2} h_i f_r^2(x_i)] + \Omega(f_r) + \text{constant} \quad (5)$$

其中, g_i 为损失函数的一阶梯度统计; h_i 为损失函数的二阶梯度统计; $g_i = \partial_{\hat{y}_i^{(r-1)}} L(y_i, \hat{y}_i^{(r-1)})$, $h_i = \partial_{\hat{y}_i^{(r-1)}}^2 L(y_i, \hat{y}_i^{(r-1)})$ 。

将 *sigmoid* 函数带入, 则一阶梯度统计和二阶梯度统计分别变形为: $g_i = \text{sigmoid}(\hat{y}_i^{(r-1)}) - y_i$, $h_i = \text{sigmoid}(\hat{y}_i^{(r-1)}) (1 - \text{sigmoid}(\hat{y}_i^{(r-1)}))$ 。因常数项对目标函数的优化没有影响, 故可将常量去除。目标函数公式(5)变形为:

$$Obj^{(r)} = \sum_{i=1}^n [g_i f_r(x_i) + \frac{1}{2} h_i f_r^2(x_i)] + \Omega(f_r) \quad (6)$$

其中, $L2$ 正则项通过对叶子节点个数以及节点的权重进行约束, 从而达到对模型复杂度进行控制的目的。

假设 *CART* 树定义为:

$$\begin{cases} f_r(x_i) = \omega_{q(x_i)} \\ \omega \in R^T \\ q: R^d \rightarrow \{1, 2, \dots, T\} \end{cases} \quad (7)$$

T 表示叶子节点的个数, $q(x_i)$ 代表 *CART* 树的结构。由上式(7) 带入目标函数式(6) 可得式(8):

$$Obj^{(r)} \approx \sum_{i=1}^n [g_i \omega_{q(x_i)} + \frac{1}{2} h_i \omega_{q(x_i)}^2] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2 = \sum_{j=1}^T [(\sum_{i \in I_j} g_i) \omega_j + \frac{1}{2} (\sum_{i \in I_j} h_i + \lambda) \omega_j^2] + \gamma T \quad (8)$$

其中, I_j 为叶子节点的样本集。

$f_s(x_i)$ 将样本划分到叶子结点, 计算得到该叶子结点的分数 ω 。因此, 当 $i \in I_j$ 时, 可以用 ω_j 来代替 $f_s(x_i)$ 。将 $G_j = \sum_{i \in I_j} g_i$, $H_j = \sum_{i \in I_j} h_i$ 带入上式进行简化可得:

$$Obj^{(r)} = \sum_{j=1}^T [G_j \omega_j + \frac{1}{2} (H_j + \lambda) \omega_j^2] + \gamma T \quad (9)$$

从而求出各个叶子节点的最佳值以及此时目标函数的值:

$$\omega_j^* = -\frac{G_j}{H_j + \lambda} \quad (10)$$

$$Obj_j^{(r)} = -\frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \gamma T \quad (11)$$

3.3 基于 XGBoost 模型的路况感知算法

鉴于 XGBoost 模型在分类问题上高效、准确的特性表现, 本文提出的路况感知算法主要基于此模型^[16-18]。其算法主要包括: 数据采集、特征提取、低通滤波、使用训练好的 XGBoost 模型进行路况感知和标记路况等步骤。具体步骤如图 4 所示。

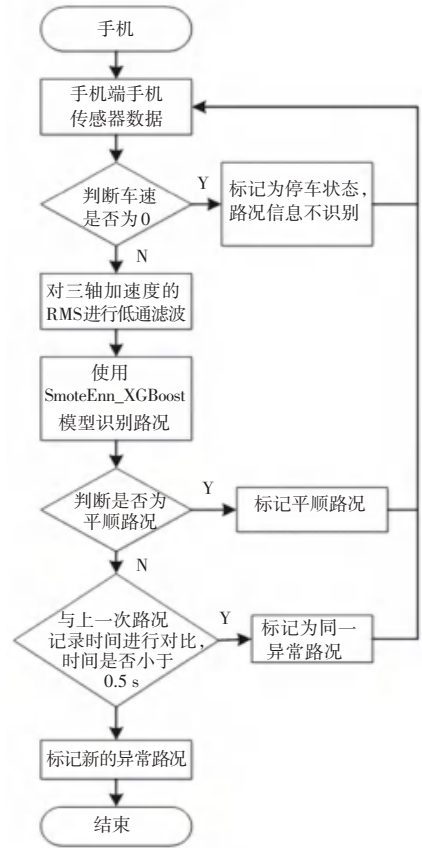


图 4 基于 SmoteEnn_XGBoost 模型的路况感知算法流程图

Fig. 4 Flow chart of road condition awareness algorithm based on SmoteEnn_XGBoost model

(1) 采集数据: 通过手机, 应用 Road Data Collection 采集设备的 IP、时间、三轴加速度、GPS 信息、车速等信息;

(2) 判断车速: 用于排除由于开关门等非道路因素引起的加速度波动, 从而降低运算能耗;

(3) 将三轴加速度的均方根 *RMS* (公式 12) 进行低通滤波处理, 去除发动机等转动设备的高频干扰;

(4) 提取特征: 将三轴加速度 (分别用 x 、 y 、 z 表

示)、三轴加速度的 RMS 、车速、相邻两条加速度之差的绝对值 Abs_Acc (公式 13)、相邻两条三轴加速度的 RMS 之差的绝对值 δ (公式 14) 等, 作为 XGBoost 模型分类的特征;

(5) 将特征值传入 XGBoost 模型得到路况的预测值;

(6) 判断是否为平顺路段: 若为平顺路段标记 'type' = 0, 否则与上一次标记的异常路段时间进行对比; 如果小于阈值时间则判断为是同一个异常路段, 否则标记为新的异常路段, 并记录当前时间、GPS 信息。

$$RMS = \sqrt{x^2 + y^2 + z^2} \quad (12)$$

$$Abs_Acc = |A_i - A_{i-1}| \quad (13)$$

$$\delta = |RMS_i - RMS_{i-1}| \quad (14)$$

3.4 不同模型结果对比

本文采用 XGBoost 模型作为路况感知模型, 并与当前使用较多的 SVM 模型、GBDT、随机森林等方法进行对比实验。本文采用 10 191 条数据进行实验, 其中 7 643 条数据用来训练模型, 2 548 条数据用来测试模型。对比结果见表 2。

表 2 不同模型关于路况分类的评价值对比

Tab. 2 Comparison of evaluation value of different models on road condition classification

模型名称	时间/s	召回率	精确率	F1	ROC	误判率
XGBoost 模型	0.228 4	0.967 8	0.968	0.967 8	0.968 1	0.032 2
SVM 模型	2.7	0.931 7	0.932 5	0.931 7	0.932 4	0.068 3
GBDT	2.317 8	0.952 5	0.952 6	0.952 5	0.952 6	0.047 5
随机森林	7.761 2	0.931 7	0.931 9	0.931 7	0.932	0.068 3

由表 2 可知, 4 种模型的评价指标值均在 90% 以上, 都能较好的分类出平顺路况和不平路况。但从运行时间上看, XGBoost 模型的运行时间远远小于其它 3 种模型, 这让基于 XGBoost 模型的路况感知算法在实际中推广的可能性大大增加。此外, XGBoost 模型的 AUC 值相对其图 3 种模型有较大的提高。证明 XGBoost 模型应用于路况分类问题中有较好的分类效果。

本算法也存在进一步提升的可能。从图 5 可以看出通过对输入特征进行量化可知相邻两条 Z 轴加速度的绝对值在本算法中起到决定性作用。因此, 在后续的算法优化的过程中, 可以通过修改每个特征值的权重来进一步优化该算法, 提高模型的适

应性。

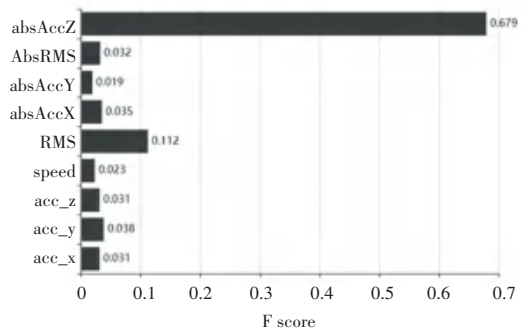


图 5 XGBoost 模型中特征值权重图

Fig. 5 A graph of eigenvalue weights in the XGBoost

4 结束语

本文提出的基于 SmoteEnn_XGBoost 模型的路况感知算法, 利用智能手机终端采集汽车运行期间的振动数据, 以此提取主要特征进行处理, 识别出具体的路况信息, 并将具体的路况感知情况和 GPS 信息进行记录。利用 SmoteEnn 算法将路况数据集进行不均衡数据处理, 得到数据均衡的路况数据集, 可以减少多数类样本在分类评价结果中的影响, 使应用于本系统的 XGBoost 模型准确度更高。XGBoost 模型相对于其他分类算法, 将正则项加入目标函数, 并对目标函数进行泰勒展开, 同时应用了一阶、二阶偏导数, 在提高算法效率的同时, 能够控制模型的复杂度。

参考文献

- [1] 余志生. 汽车理论[M]. 机械工业出版社, 2019: 230-232.
- [2] 周晓青, 孙立军, 颜利. 各国路面平整度验收规范[J]. 中外公路, 2006, 26(1): 52-56.
- [3] BHORASKAR R, VANKADHARA N, RAMAN B, et al. Wolverine: Traffic and road condition estimation using smartphone sensors[C]// 2012 fourth international conference on communication systems and networks (COMSNETS 2012). IEEE, 2012: 1-6.
- [4] 仵晓煜. 基于 PDA 平台的便携式路况采集系统研究与实现[D]. 西安: 长安大学. 2012.
- [5] 李若童, 申威. 基于手机端的道路病害管理系统设计与实现[J]. 科技视界, 2015(24): 129-130.
- [6] 邵志超. 基于多传感器数据融合的路面平整度检测系统[D]. 西安: 长安大学. 2015.
- [7] 王端宜, 陈颖. 智能手机在路况调查中的应用研究综述[J]. 公路工程, 2017(5): 100-104.
- [8] A G S, A D B, A S S, et al. Smart patrolling: An efficient road surface monitoring using smartphone sensors and crowdsourcing[J]. Pervasive and Mobile Computing, 2017, 40: 71-88.
- [9] CABRAL F S, FUKAI H, TAMURA S. Feature Extraction Methods Proposed for Speech Recognition Are Effective on Road Condition Monitoring Using Smartphone Inertial Sensors[J]. Sensors, 2019, 19(16): 3481.

(下转第 147 页)