计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2021,30(8):311−316 [doi: 10.15888/j.cnki.csa.007922]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于用户行为分析的诈骗电话识别①

杨建昆[1], 夏文财[2]

[1](中国电信股份有限公司 云南分公司, 昆明 650299)
[2](云南电信公众信息产业有限公司, 昆明 650118)
通讯作者: 杨建昆, E-mail: 13398716608@189.cn

摘　要: 本文的研究目的是提高诈骗电话的识别率和识别准确性. 基于大数据平台采集用户通话行为、上网行为等通信过程数据, 结合用户基本属性、手机终端信息等进行综合分析, 并采用合适的识别算法进行机器学习建立识别模型, 能更好的发现诈骗电话与普通电话的内在差异, 相比传统基于呼叫行为的分析, 能有效提高骚扰诈骗电话识别的准确度和覆盖率, 降低漏判、减少误判. 经实际数据验证, 对诈骗电话识别效果有明显提升, 可作为诈骗电话识别的一种新的技术选择.

关键词: 诈骗电话; 机器学习; 用户行为分析

引用格式: 杨建昆,夏文财.基于用户行为分析的诈骗电话识别.计算机系统应用,2021,30(8):311–316. http://www.c-s-a.org.cn/1003-3254/7922.html

Fraud Call Identification Based on User Behavior Analysis

YANG Jian-Kun[1], XIA Wen-Cai[2]

[1](Yunnan Branch of China Telecom Co. Ltd., Kunming 650299, China)
[2](Yunnan Telecom Public Information Industry Co. Ltd., Kunming 650118, China)

Abstract: The purpose of this study is to improve the recognition rate and accuracy of fraud calls. We collect the communication process data such as users' behavior of having telephone communications and surfing the Internet by a big-data platform and conduct a comprehensive analysis combined with users' basic attributes and mobile terminal information; also, an identification model is built by the appropriate recognition algorithm for machine learning. The proposed method can better find the internal differences between fraud calls and ordinary ones. Compared with traditional analysis based on call behavior, it can effectively improve the identification accuracy and coverage of prank and fraud calls and reduce false negatives and false positives. The proposed method performs prominently better in fraud call identification, which can be used as a new technology choice, according to actual data verification.

Key words: fraud call; machine learning; user behavior analysis

1　概述

诈骗电话对用户的正常通信造成严重干扰, 并给部分用户造成经济损失. 国家从政策法规、执法力度方面加强打击和遏制. 运营商在电话入网开通时进行严格实名认证并进行个人开通号码数限制. 这些措施的执行取得了一定效果, 但诈骗电话依然较多, 个别地区反诈形式严峻, 社会反响较大[1,2].

近年, 公安部门、通信运营商也积极从技术方面采取防范措施, 从电话使用行为的监管和分析入手, 发现有诈骗行为特征的电话及时进行关停, 尽量做到早发现、早制止, 从而减少危害. 笔者所在省电信公司从2017年开始向公安反诈平台推送实时话单做反诈分析, 同时公司也建立了自己的异常呼叫识别平台, 对异常呼叫号码进行识别, 对达到一定阈值条件的号码进

① 收稿时间: 2020-07-27; 修改时间: 2020-08-29; 采用时间: 2020-10-30; csa 在线出版时间: 2021-07-31

行关停处理. 判别依据主要是根据诈骗电话的高频主叫特征. 存在的突出问题就是误判较多, 因为很多客服类、快递、送餐类电话也具有高频主叫特征[3,4].

关于诈骗电话识别方面的研究近年也有很多, 但主要也是基于语音话单和呼叫信令分析方面. 文献 [5] 提出通过通信记录数据去发现电信诈骗通话行为与普通用户的通话行为特征差异,并结合用户属性, 使用支持向量机 (SVM) 进行电信诈骗行为的学习, 完成电信诈骗行为的识别. 文献 [6] 介绍了通过对海量呼叫信令的大数据分析进行诈骗电话预警.

本文在基于通话行为特征分析的基础上, 采集用户上网行为数据, 再结合用户年龄、入网时长、使用手机终端类型等基础信息进行综合分析, 建立机器学习模型, 对历史数据进行学习, 能更好的发现骚扰、诈骗电话特征, 更好的区分诈骗电话与快递类、网约车司机、客服类电话的不同. 可有效减少误判, 降低漏判, 提升诈骗识别的准确率和覆盖面.

近年, 机器学习相关技术和工具也有了较快发展, 出现了一些新的效率更高、分类效果更好的机器学习算法. 针对我们采集的样本数据, 神经网络算法具有明显的优势, 本文介绍的诈骗电话识别方法最终采用神经网络算法建立机器学习模型实现. 为方便识别方法描述, 文中针对主要技术环节给出用 Python 语言编写的具体实现代码.

2 样本号码及用户行为特征数据采集

具有代表性的样本号码和区分度高的特征维度数据是建立一个好的诈骗电话识别模型的基础.

用户资料、话单等数据是受国家法律保护的个人隐私数据. 通信运营商需严格管控, 不能外泄, 仅可在本网内分析使用. 因此, 从全网收集样本数据较为困难, 本项目实验采集样本数据仅限于笔者所在电信公司本网数据. 实验证明有效的方法也可供其他运营商在自己网内参考使用.

2.1 样本号码的收集

为了能准确的区分不同类别电话特征差异, 减少误判, 同时结合收集数据的难易程度, 我们将号码分为诈骗电话、快递类电话、常规电话、骚扰电话、网约车司机电话 5 类进行分析. 在后面算法模型中分别用类别 0、1、2、3、4 表示.

项目分析用样本数据为 341 549 个电信号码, 其中诈骗电话 11 229 个; 骚扰电话 18 594 个; 网约车司机电话 2388 个; 快递类电话 9338 个; 普通电话 300 000 个.

骚扰和诈骗电话从 12321 举报中心和公安部门通报下发的数据得到. 其中诈骗电话含公安通报的实际涉案电话和用户举报且举报类型为诈骗的电话. 我们建立模型的目的就是监测和识别骚扰诈骗电话, 尽量做到早发现, 早关停, 减少危害, 降低投诉.

网约车司机电话和快递类电话从网约车平台和物流公司得到. 这类电话用户基本稳定, 数量也较少,基本采用全量收集. 收集这 2 类电话的目的是为了更好的学习他们与诈骗电话的不同特点, 使模型能更准确的对他们进行区分, 避免误判停机, 影响快递员、网约车司机的正常通信. 各种客服电话在实践中也是容易被误判的一类电话, 但客服类电话难于准确标识, 不便单独采集进行学习, 我们把它归到普通电话类别中.

普通电话为不属于骚扰诈骗、快递送餐类、网约车类的其他电话, 相当于训练集的负样本. 这类电话可从运营商数据库中直接抽取, 最容易得到, 但不是越多越好, 可根据模型分析需要适量抽取. 由于其他类别电话采集量受客观因素影响, 相对固定, 单方面增加普通电话样本量会增加正负样本的不平衡性和模型训练的复杂度, 对训练结果提升不一定有利. 实际训练过程中我们尝试取不同数量负样本进行训练, 初始阶段训练效果随样本数增加而提升, 当负样本数量达到正样本数 10 倍以后, 训练效果成下降趋势.

2.2 特征选择及特征数据采集

每类电话都有自己相对独立的特征, 要提高模型的分类能力, 就要让模型尽可能学习到对各类电话有区分度的各种维度特征. 一般先根据业务经验确定可能对区分不同类别电话有用的数据特征, 再结合数据采集难易程度及用于生产环境的可行性来确定需要采用的数据维度特征. 采集到数据后, 再利用特征工程分析各种特征对区分不同类别电话的实际作用, 对作用不明显或无实际意义的特征进行剔除.

本文采用了电话用户的基础属性、用户使用手机终端信息、用户通话行为信息、短信收发信息、用户上网行为信息等.

2.2.1 用户基础属性及资费数据采集

用户基础属性包括年龄、性别, 入网时间、积分、信用度等级、星级等. 用户资费数据包括套餐基本费、月均话费、预存余额、欠费情况等. 作为通信

运营商, 客户基础信息和资费信息就保存在运营商自己的客户关系管理系统 (CRM) 和账务系统中, 可以直接从数据库中获取, 是最容易采集到的数据.

由于诈骗电话号码容易被封停, 因此, 用于诈骗的电话通常是新开卡、临时卡. 从号卡入网时间等基础属性数据能做一定程度的区分.

2.2.2 用户通话行为数据采集

用户通话行为数据可从语音话单或呼叫信令中得到. 具体通过统计一段时期内的通话记录来构建, 包括主叫次数、被叫次数、通话时长、漫游通话天数、漫游地个数、主叫通话人数、每天不同时段通话情况等. 与商务人士、固定上下班人员不同, 诈骗活动通常是在一个相对固定的场所进行. 因此, 通话行为也是区分普通电话与骚扰诈骗电话的重要特征.

2.2.3 用户使用的手机终端信息采集

诈骗电话使用的手机一般为专用手机, 不作为个人正常通信使用, 所用手机通常性价比较高. 因此, 手机终端型号, 价格也是诈骗电话的一个较为显著的特征. 从自注册系统、大数据平台可以获取到终端型号数据. 从终端管理系统可以获取到终端价格信息.

绝大部分手机在首次使用时会通过短信将手机品牌、型号等终端信息发送给通信运营商的自注册系统, 因此, 从自注册系统可以采集到手机终端信息.

少部分品牌手机不会发送自注册信息给自注册平台, 这类手机终端信息可从用户上网日志的 HTTP 请求包头中解析得到. 前提是运营商建有可存储上网日志的大数据平台.

2.2.4 用户上网行为特征数据采集

用户上网行为包括用户一段时间内使用的数据流量、上网时长、上网内容等. 数据使用量和上网时长可从数据话单中得到. 上网内容即用户上网做什么, 经常使用什么手机 APP 等. 可从经常访问的网址和经常使用的 APP 来评判. 不同类别电话用户经常使用的 APP 有较大差异, 网约车司机使用导航类 APP 相对较多, 同时有自己专用的接单 APP; 快递送餐类同样有自己专用的 APP. 因此, APP 使用行为是一个重要的分类特征标识. 我们根据 APP 在各类电话用户中使用的频次情况, 收集了 8 个典型 APP 作为模型特征参与电话分类. 将一段时间内用户使用各个 APP 的次数作为特征值. 这是本方法与传统诈骗电话识别方法最大的不同点, 下面进一步说明 APP 特征数据的采集方法.

移动网络在为手机提供上网服务时, 会产生与上网行为相关的日志. 日志内容包含访问 IP、端口、HTTP 请求报头、通信协议等信息. HTTP 报头中包含手机终端使用的操作系统版本, 终端型号等信息, 通过对包头解析可提取出上网手机的终端信息. 另外, 每个手机 APP 都会与相对固定的远端服务器连接, 并有相对固定的访问 IP、端口和通信协议. 根据用户手机请求的 IP 地址、端口、协议信息我们就能知道用户使用的 APP 情况.

上网日志数据量巨大, 一个中等规模省份一天的日志量级达 TB 级, 记录数超百亿, 要长期存储和使用这些日志数据必须具有分布式的海量存储和强大的并行计算能力. 得益于近年大数据技术的发展, 三大运营商都建设了自己的大数据平台. 用户较长一段时间的上网日志都会留存在大数据平台中, 为上网行为分析提供了基础条件. 正是有了这样的环境, 我们可以使用上网行为特征参与电话分类, 进行更准确的诈骗电话识别.

在实际的生产环境中, 为提高响应速度, 合理分担负载, 运营商一般会建立省级、集团级 2 级大数据平台. 省级平台实时接收网络设备产生的上网日志, 并根据日志中的 IP、协议、端口等信息结合打标规则进行打标 (在日志记录中增加标签信息, 如标注是哪个已知性质的 APP 产生的日志), 然后再上传集团大数据平台. 集团再做关联汇总形成满足各种维度需求的宽表数据, 从而满足其他系统及各省对数据请求响应的及时性要求.

根据以上业务分析和数据采集方法, 本次研究共收集了 63 个维度特征数据参与电话分类识别.

2.3 维度特征相关性分析

为了进一步了解所选特征对电话分类的支持程度, 探索特征选择的合理性, 我们使用开源工具 feachselect 对特征与类别标签做相关性分析, 主要实现代码如下:

```
fs=FeatureSelector(data=X,labels=Y)
fs.identify_zero_importance(task='classification',eval_metric ='None',n_iterations =10)
fs.identify_collinear(correlation_threshold=0.9)
print(fs.record_collinear) #强的特征可进行相关合并
print(fs.feature_importances) #重要性低的可移除
fs.feature_importances.to_csv('/python/zpsb/data/imp.csv')
```

表 1 列出了部分特征的重要性分析结果.

表 1 分类特征重要性系数表

排序	特征	重要性系数	特征说明
0	open_date	0.043 121 693	开通时间
1	birth_date	0.043 012 346	客户生日
2	dt_1032	0.029 171 076	本月10 s内主叫通话数占比
5	dt_m_1052	0.025 921 223	近3个月主叫通话人数
6	app1_visits	0.025 441 505	应用1当月启动次数
7	price	0.025 266 314	手机上市价格
9	dt_1051	0.023 807 172	本月主叫通话人数
10	net_dt	0.023 561 434	客户入网时间
11	dt_1068	0.022 723 104	主叫对端号码区域分布数
16	dt_1630	0.018 892 416	本月发送短信次数
21	dt_1620	0.018 054 086	本月手机上网流量
23	dt_1617	0.017 486 185	本月手机上网次数
24	app4_visits	0.017 215 755	应用4当月启动次数
34	app5_visits	0.013 514 403	应用5当月启动次数
52	star_level	0.009 082 892	星级级别
61	gender	0.004 245 738	性别
63	Mem_level	0.000 424 456	会员级别

从特征重要性分析结果可看出, 选用的与上网行为相关的特征, 即手机 APP 使用次数对分类重要性排序都排在前 50, 尤其 app1 使用次数对电话分类的区分度更是好于很多通话行为特征.

会员级别对分类的重要性系数接近 0, 说明该特征对电话分类几乎没有意义. 主要原因是会员级别取值分布较为集中, 其中取值为 99 999 的记录占 99%, 即不管什么类别电话, 其对应的会员级别取值基本是一样的, 这种特征列对电话分类没有区分度, 可以移除, 留下反而会增加模型复杂度. 通过后面测试验证也证明相关性极低的特征从模型中移除更有利于模型效果的提升.

2.4 样本数据的拆分

对数据进行分析前, 需先对数据进行预处理, 包括将样例数据中的空值、异常值用合适的值进行替换; 将日期、字符型特征进行数值化转换; 用标准化变换消除量纲对特征取值的影响, 并做归一化转换. 数据预处理对保证模型效果具有重要意义, 处理过程基本都是经典方法, 本文不做详述.

将经过预处理的数据先按列拆分为特征列 (用 X 表示) 和标签列 (Y 表示). 再将样本特征数据按行划分为训练集 x_train 和测试集 x_test, 对应的标签数据为 y_train 和 y_test. 训练集数据用来训练模型. 未参与训练的测试集数据用来验证模型效果. 方法如下:

x_train, x_test,y_train, y_test= train_test_split(X,Y, test_size=0.25,stratify=y,random_state=0)

指定 stratify=y 是为了保持拆分后各类别电话在拆分集的比例与原样本集保持一致, 确保少数类电话不因拆分随机性在某个拆分集占比太少, 影响学习或效果评估.

3 算法选择及机器学习识别模型的建立

电话分类预测是一个多分类问题, 有多种分类算法支持建立多分类识别模型. 在开始学习前难于确定采用哪种算法较好. 一般先采用多种主流算法进行尝试性训练和评估, 如果试用模型表现都较好, 则分别调参优化, 再用集合算法进一步提升预测效果.

针对本次电话样例数据, 分别采用决策树、文献 [3] 介绍的改进 GA-SVM、微软的梯度提升框架 LightGBM (Light Gradient Boosting Machine)、神经网络进行机器学习. 测试发现 LightGBM 和神经网络效果明显好于其他方法, 训练宏平均得分在 85% 以上, 其他方法都在 80% 以下. 其中 LightGBM 在无 GPU 环境下训练速度最快. 神经网络验证得分最高, 在 GPU 环境下, 训练速度比 LightGBM 更快. 下面重点介绍用神经网络建立诈骗电话识别模型.

3.1 用 Keras 框架定义神经网络

Keras 是一个用 Python 编写的深度学习框架 API, 支持多种深度学习语言 (TensorFlow、CNTK 等) 并内置 GPU 支持. 利用 Kerase 能比较高效的完成模型建立.

通过 Keras 建立一个 MLP 神经网络诈骗电话识别模型的实现方法如下:

```
model = Sequential()
model.add(Dense(512, activation='sigmoid', input_dim=x.shape[1])) #定义输入层
model.add(Dense(256, activation='relu')) #添加隐藏层
model.add(Dense(5, activation='sigmoid')) #定义输出层
adam=keras.optimizers.Adam(lr=0.000 5) #定义优化器
model.compile(loss='categorical_crossentropy',optimizer='adam',metrics=['accuracy']) #编译模型
his=model.fit(x_train,to_categorical(y_train,num_classes=5),epochs=120,batch_size=2048,validation_data=(x
```

_test,to_categorical(y_test,num_classes=5)))

y_pred = model.predict_classes(x_test) #对测试号码分类

#输出模型评估报告 (实际值与预测值对比分析)

print(metrics.classification_report(y_test,y_pred))

model.save('/model/zpsb.h5') #保存训练好的模型

3.2 模型参数选择及调优

模型参数的取值与具体样本特征关系较大, 下面对几个重要参数的确定进行说明.

激活函数用于映射特征值与标签值的内在关系. 对于多分类, 输出层的激活函数通常使用双曲正切函数 Sigmod 或 Softmax. 本类中用 Sigmod 效果较好. 输入层和隐藏层一般选择线性整流函数 ReLU 效率更高.

损失函数用于计算真实值与预测值的差. 模型训练的目的就是使损失值最小. 对于多分类, 损失函数为 categorical_crossentropy, 同时标签值需用函数 to_categorical 进行向量化.

优化器是更新激活函数权重的特定算法, 用于确定优化方向, 降低损失值. Adam(自适应学习率) 在确保收敛的情况下, 能大幅提升学习效率, 学习效果也通常更好. 在本例中使用默认参数情况下, 采用 Adam 比采用 sgd (随机梯度) 综合得分提升约 2%.

迭代次数指定模型重复训练的轮数. 在每次迭代中, 优化器都重新调整权重, 提高训练准确率. 将保存在 his 中的训练过程数据显示出来可方便的看出最佳迭代次数, 如图 1 所示.

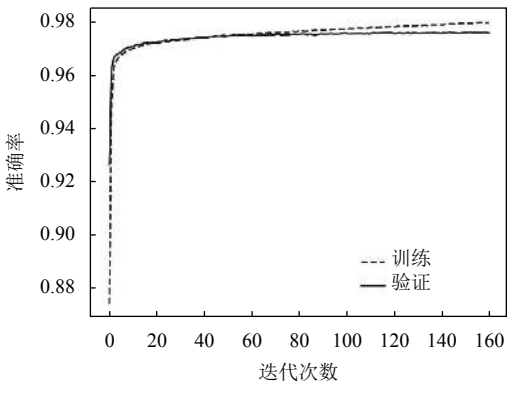

图 1 迭代次数与准确率关系

从图 1 中可以看出迭代次数越大训练拟合越好 (虚线), 但当迭代次数超过 120 次时, 验证效果不再提升反而有下降. 建立模型的目的是用于对未知标签的数据进行预测, 因此最佳迭代次数由验证曲线决定.

Batch_size 参数设定每次迭代时一次输入模型的样本数. 为使每次处理的数据都尽可能包含占比较小类别的电话, 将该参数值设置为远大于默认值的值. 本例设置为 2048 取得较好效果.

3.3 识别方法及模型效果评估

诈骗识别模型效果的评估一是要考虑分类准确性, 即查准率, 验证有多少识别为诈骗的电话是真的诈骗电话, 有多少被误判. 二是要考虑所有诈骗电话中有多少能被识别出来, 多少被漏判, 即召回率. 只有召回率和查准率都高, 才能达到最终的应用目的. 因此, 用综合了查准率和召回率的综合性指标 $F1$ 得分对模型效果进行评估.

使用 sklearn 工具包能比较完整的得到模型对验证数据分类结果的效果评估报告, 代码如下:

from sklearn import metrics

print(metrics.classification_report(y_test,y_pred))

为了分析上网行为特征对电话分类的实际作用, 分别使用不带上网行为特征数据和带上网行为特征数据分别训练模型, 再对测试数据进行分类并输出评估报告, 如表 2 和表 3 所示.

表 2 分类结果评估报告 (不使用上网行为特征)

行为	Precision	Recall	$F1$-score	Support
0	0.93	0.72	0.81	2807
1	0.93	0.86	0.90	2335
2	0.98	0.99	0.99	75000
3	0.85	0.89	0.87	4649
4	0.75	0.46	0.58	597
Accuracy	—	—	0.97	80690
Macro avg	0.89	0.79	0.83	80690
Weighted avg	0.97	0.97	0.97	80690

表 3 分类结果评估报告 (使用上网行为特征)

行为	Precision	Recall	$F1$-score	Support
0	0.94	0.75	0.83	2807
1	0.95	0.89	0.92	2335
2	0.99	1.00	0.99	75000
3	0.87	0.90	0.89	4649
4	0.85	0.81	0.82	597
Accuracy	—	—	0.98	85388
Macro avg	0.91	0.86	0.88	85388
Weighted avg	0.98	0.98	0.98	85388

从表 2、表 3 对比可以看出, 在采用同样算法模型和参数的情况下, 使用上网行为特征参与诈骗电话分类识别, 模型验证集 $F1$ 得分宏平均值为 88%, 比不使

用上网行为特征 $F1$ 得分提升了 5%. 诈骗电话 (类别 0) 本身的 $F1$ 得分提升了 2%, 召回率得分提升了 3%, 也即有更多的诈骗电话能被识别出来, 同时准确率也同步得到提升. 最主要的原因是通过使用上网行为特征, 网约车 (类别 4)、快递类电话得到准确区分, 网约车识别 $F1$ 得分从 58% 大幅提升到 82%, 误判率得到大幅降低.

由于训练集只是样本的一部分, 在所有参数都调整完成后, 可将测试数据并入训练集, 用全部样本数据再对模型做最终训练. 由于训练数的增加, 训练后的模型泛化能力会进一步提升, 实际分类效果会比表 3 所示的分类效果更好.

4 模型应用

训练好的模型可用于对给定号码集进行静态预测分类, 也可用于生产过程进行在线识别. 在线识别用于监测分析实时话单, 及时识别并拦截当天具有诈骗行为特征的电话. 静态分类多用于对当天之前的历史话单进行分析, 批量输出骚扰诈骗电话进行关停处理. 静态识别一般是对在线识别漏掉的诈骗电话进行重新分析和挖掘.

部署一台运行识别模型的服务器, 每天夜间从大数据平台抽取话单、上网行为特征等业务数据, 经清洗变换形成模型所需要的特征数据, 为电话分类识别做好准备. 再部署一台话单监听服务器, 实时接收交换机送来的话单文件, 解析话单入库, 并对每个主叫号码进行通话频次统计, 当在设定时间窗口内, 主叫超过一定次数, 则将号码送入运行诈骗电话识别模型的服务器做进一步识别, 如果被识别为疑似诈骗电话, 则通知设备进行关停处理, 从而实现诈骗电话的在线识别和及时拦截.

本文介绍的模型方法使用了用户入网时间、性别、上网行为日志等数据, 适用于拥有这些数据的本网运营商做诈骗识别分析, 并对识别出的骚扰诈骗电话及时进行关停操作.

对异网号码或国际呼入电话, 由于难于采集用户基本属性数据和上网行为数据, 只能建立只针对通话行为进行分析的模型, 判定为疑似诈骗的异网电话可在交换机进行拦截, 中止接续服务. 由于缺少上网行为等特征分析, 为减少误判, 只能放宽拦截条件.

5 结束语

本文介绍了基于用户上网行为和通话行为分析的诈骗电话识别方法, 并通过神经网络算法建立了可用于生产环境进行诈骗电话识别的机器识别模型. 所述方法容易在生产环境部署实施, 通过测试数据进行验证, 诈骗电话识别的准确率和召回率相比传统只基于通话行为的识别效果有明显提升.

电话诈骗形式多样且多变, 识别模型也需要不断改进提升. 一方面是不断的用新的能代表近期诈骗电话特点的样本数据重新学习. 在用于生产过程中, 及时将误判的号码加上标签作为样本数据参与新的轮次学习, 使模型不断自我学习, 自我提升, 越用越准确. 另一方面, 探索引入新的有利于提高模型区分度的新特征, 如用户开户证件类型、通话基站及位置等特征参与机器学习, 也可考虑引入第三方平台 (360 手机卫士、阿里钱盾等) 收集到的大众对电话的标识信息作为特征参与分类等, 这些可作为进一步提升诈骗电话识别效果的改进方向.

参考文献

1 程锦红, 萧瑶, 方雅丽, 等. 基于大数据的防范电话诈骗体系架构研究. 通讯世界, 2020, 27(4): 13–15. [doi: 10.3969/j.issn.1006-4222.2020.04.008]

2 张杰俊, 李爽. 电话诈骗识别系统的设计与实现. 软件, 2020, 41(4): 190–194. [doi: 10.3969/j.issn.1003-6970.2020.04.040]

3 王世豪, 蔡延光. 基于改进 GA-SVM 的电信客户欺诈识别方法. 东莞理工学院学报, 2019, 26(5): 14–20.

4 赵越, 王瑜, 葛阳, 等. 基于机器学习的大数据防诈骗能力研究与应用. 江苏通信, 2019, 35(4): 64–66, 74. [doi: 10.3969/j.issn.1007-9513.2019.04.017]

5 吉涵之, 马宇宸, 李爽, 等. 基于 SVM 的电信诈骗行为特征识别方法. 软件, 2017, 38(12): 104–109. [doi: 10.3969/j.issn.1003-6970.2017.12.020]

6 王志刚, 曲劲光. 基于大数据的电信诈骗治理技术研究. 电信工程技术与标准化, 2017, 30(4): 86–89. [doi: 10.3969/j.issn.1008-5599.2017.04.025]