

诈骗嫌疑人声纹数据库创建对防治电信诈骗电话的价值

康锦涛¹ 黎荣晋²

1. 公安部鉴定中心 广东深圳 518000

2. 深圳声扬科技有限公司 广东深圳 518000

摘要: 随着科技水平的逐渐发展, 市场上出现很多新的行业, 在行业的蓬勃发展时却出现了电信诈骗电话等的不法行业。以电信诈骗为代表的新型犯罪行为持续高发, 在很大程度上危害了人们的生命和财产安全, 对诈骗人员建立声纹数据库能把犯罪人员的生理和行为的声纹特征登记在库内, 在拨入来电时实时地对说话者的语音提取出声纹特征, 并进行库内匹配识别, 快速定位出疑似犯罪者的身份并调取出相应的个人信息。通过建立声纹数据库来核验犯罪人员的身份, 可以加强对犯罪人员的甄别, 加强对市民的保护, 谨防诈骗犯罪的发生。

关键词: 声纹数据库; 电信诈骗; 防止诈骗

在科技水平的发展中我国社会信息化水平进一步提升, 电信诈骗电话的行为越来越高发, 诈骗手段和骗术层出不穷, 并且呈现出从广撒网到精准撒网的趋势, 这对人们的生命和财产安全都造成了巨大的影响, 扰乱了社会秩序, 破坏了社会的安定, 对社会诚信造成了极大的危害, 严重侵害我国人民的切身利益。虽然公安机关一直在加大对电信诈骗的宣传力度和打击力度, 但是由于电信诈骗的犯罪行为的打击难度较大, 电信诈骗还处于高发阶段, 并且诈骗的金额和数量一直在持续上升, 电信诈骗电话打破了传统的犯罪中的时间和空间限制, 使受害人遍布全球。

1. 电信反诈系统的声纹识别整体构架

声纹识别是电信反诈系统中的重要应用, 它主要是通过声纹识别系统对说话人的语音信息进行截取和采集, 然后进行编码压缩, 生成对应的声纹特征后, 再与提前准备好的海量声纹数据库中的诈骗人员声纹进行全量比对, 最终从数据库中确定出诈骗人员的身份信息^[1]。声纹识别通过生物认证技术, 提取出语音中的语音信息, 例如音色、基频、共振峰等发音器官所产生的生理特征, 与个人后天的言语习惯、发音方式、方言语种等的行为特征, 生成出独一无二的生物信息后达到识别身份的目的, 其具备远程非接触式的认证特点与难伪装的真实性。由于以上种种原因, 不仅不同人之间的声纹信息差异大, 而且同一人的声纹信息是稳定持久, 这一切使得通过分析语音信号来识别百万级别的说话人身份成为了可能。

建立声纹识别系统要分为注册阶段和识别阶段。

注册阶段, 即建立声纹数据库的过程, 它能为后续的识别阶段提供涉案嫌疑人的注册声纹特征, 并同时在数据库中录入涉案嫌疑人对应的姓名、性别、身份证号码等的个人信息, 与注册语音的语种方言、采集设备、文本内容、采集时间、采集地点等语音信息。对声纹数据库的分析和评价一般受录音者的数量, 录音的次数和时间间隔, 录音对应的文本类型, 语音采集的设备, 语音采集的环境, 语音采集的传输通道等众多因素的影响。为了建立能统一、规范、实用的声纹识别数据库, 必须在入库之前清洗出对语音造成影响的各种因素^[2], 包括错误标记, 如个人信息与语音信息; 包括噪声干扰, 如无效的静音片段等等。

识别阶段, 即对接入的诈骗电话进行实时查询, 提取出电话音频中的声纹特征并与声纹数据库进行比对, 命中疑似嫌疑人后返回相应的个人信息与对应注册音频的语音信息, 以便操作人员查阅分析。声纹识别电信反诈系统(如图1)主要分为电话接入检测层、传输层和声纹识别服务层。建立电话接入检测层主要是感知和采集说话人的语音信息上传到公安系统, 实现对语音数据信息的分析和处理; 传输层是通过通讯链路实现对电话接入检测层的语音输入和上传, 网闸会对要输入到公安专网系统的音频进行专用链路转换并检测合规性; 声纹识别服务层是调取已在注册阶段建立好的声纹数据库、诈骗人员数据库以及管理平台服务器和客户端, 来进行综合声纹识别和声纹聚类技术。当出现

新的通话接入时,在声纹提取和声纹比对的技术下,实现对诈骗人员语音的精准识别。

2. 声纹识别技术分析

声纹识别,又叫做说话人识别,是一种可以根据语音波形来对说话人的生理特征以语音参数进行建模表征的算法,与实现说话人的声纹特征进行身份识别比对的技术。

声纹识别技术有 GMM-UBM、i-VECTOR 和基于深度学习的 x-VECTOR 声纹识别算法^[3]。本系统提出的算法架构同样是基于深度学习的 CNN-LSTM-DNN (CLDNN) 神经网络模型,该架构可分为 3 个部分,分别是前端、编码层、后端。前端是常见的语音特征,对原生音频数据进行傅里叶变换压缩得到声学特征,如 MFCC、Fbank 等,并进行端点检测、频谱均值方差规整等操作。编码层是各类 CNN、DNN、RNN/LSTM 网络,该部分以有监督学习的方式对输入的声学特征和标签信息进行迭代收敛损失,其中包括隐藏层、损失函数、池化层等,主要用于提高不同人的声纹特征类间协方差和减少相同人声纹特征类内协方差。后端是对注册和验证的声纹特征进行规整、补偿和降低维度等操作,使得两部分的声纹特征处于同一超空间,实现比对计算出它们的相似度。相比于基于深度学习的 x-VECTOR 算法,本系统则采用了 CNN-LSTM-DNN 的联合结构:先用数个一维卷积 CNN 把输入特征进行帧级别的编码降维,后采用 LSTM 长短期记忆循环层对 CNN 的特征进行时序编码,增强上下文的信息相关性;接着是池化层,把片段中的数帧汇集计算出统计量,实现不定长到定长的转换;之后采用 DNN 全连接层做段级别的编码处理。CLDNN,会输出训练集每个说话人的预测概率,在损失函数中与真实标签计算误差,并以有监督学习的方法逆向传播误差的梯度,以实现收敛。最终,在模型定型后的 DNN 全连接层中,提取出声纹特征,以作声纹数据库入库或者比对用途。

本论文采用的是学术界常用的 VoxCeleb1 作为测试集,对声纹数据库建库的过程进行分析与探讨。VoxCeleb1 是一个从 YouTube 的采访视频中提取出的人类语音短片所组成的视听数据集,它包含了超过 10 万名名人的话语。

以下是基于该测试数据集,对现有系统和本论文提出的 CLDNN 系统做出的可靠性和效率的比对结果(如表 1)。设识别语音的长度为 2 秒,性能指标的单位是识别次数/秒,实验的 GPU 型号为 GTX2080TI,服务器的 CPU 型号是

Intel i7-11700^[4]。

表 1: 主流的声纹识别系统对比

系统	性能指标 (等错误率%)	识别效率
基于深度学习的声纹识别方法 (时延神经网络)	2.56 ± 0.10	约 200 次/秒 (GPU)
i-VECTOR	9.10 ± 0.05	约 300 次/秒 (CPU)
GMM-UBM	23.00 ± 0.05	约 500 次/秒 (CPU)
本系统提到的方法 (CNN-LSTM-DNN 的联合结构)	2.21 ± 0.05	约 250 次/秒 (GPU)

3. 建设声纹数据库的步骤

声纹数据库的建设需要进行合理的规划,为数据库制定统一的建设标准,这是为了避免重复建设和系统不兼容等的相关问题。在声纹库的建设前应制定相关的标准,对声纹数据库的采集、传输、储存、检索、比对等工作进行专业的指导和规范,推动标准统一的声纹数据库方案在全国实施,保证在全国高效推进声纹数据库的建设。由于声纹数据库建设是一个长期且持久的项目,所以声纹数据库的建立需要在公安部门和相关业务部门的指导下进行,由具有领先的核心技术的专业公司来承担任务的推进工作,实现建立全国联网的重点人员声纹数据库的目标,使得声纹数据库能在打击电信诈骗电话的犯罪活动中发挥积极的作用,并满足实现大规模应用的要求^[5]。

3.1 构建阶段

声纹数据库建立的构建阶段的工作内容是建立声纹采集系统,搭建出声纹采集的环境,并根据地区(如华南地区、华中地区、华东地区、华北地区、东北地区、西南地区、西北地区)进行分布式数据库建立。可在每个地区采集约 1 万人的声纹样本数据,针对这些数据对以上提到的多种声纹模型算法做出展开分析,实现对声纹识别技术的验证,基于神经网络算法的声纹识别技术评价准确性,并对声纹识别技术的效果进行调优。

3.2 深化阶段

针对这个阶段的深化研究,我们会在每个地区相应地采集约 20 万人的声纹样本数据来进行研究开发,根据实际运行的环境对系统运行的效果加以观察和验证,对采集到的语音建立全局汇总的数据库,推进不同地区联网可入库可查询的工作,实现对语音的声纹识别技术的验证。同时优化全国范围内的声纹比对的实现细节,在这些研究的基础上探索出具体可操作性的声纹数据库建设方案与跨地区的声纹数据库查询比对方案。

3.3 实现阶段

该阶段的主要工作内容是把上述的研究结果进行产品化,使声纹识别技术能实现声纹对比服务,在实际的运行过程中根据实际需求进行面向公安系统的专门的功能定制,并实现对声纹识别技术的持续升级和优化。

4. 建设声纹数据库的作用

对诈骗人员建立声纹数据库,能在防治电信诈骗电话中发挥出积极的作用,给人民群众的生活和生活带来了便利,维护了社会秩序的稳定,即使诈骗人员总是不断翻新作案的手法,利用手机、电话等通信平台进行诈骗的违法犯罪活动,对人民群众的经济造成损失,影响社会安定与和谐。针对这类非接触性、隐蔽性、跨地域的诈骗违法犯罪活动的侦破,需要投入大量的警力和物力,但是其实际效果并不佳。由于电信诈骗电话会留下犯罪嫌疑人的声音信息,根据声音信息会对个人独特的生物特征的声纹进行识别,可在我国的刑事诉讼法、行政诉讼法和民事犯罪中作为有效的视听资料,以证据形式呈堂。声纹识别技术对电信反诈骗电话案件的侦破具有重要作用,把诈骗人员的语音样本使用深度学习的技术来进行训练,再对检测的结果进行二次验证和识别,准确识别出诈骗人员的身份信息,其中由于声纹比对的实现是通过计算机和服务端完成的,所以其检测结果一般比较精准。

4.1 预警服务

建设声纹数据库系统可以对可疑的和已经备案的手机、固定电话、网络电话等通讯工具的实时语音通话实现重点关注,通过抓取语音通话中的声纹进行识别,在数据库中进行匹配,将比对结果显现,破获和识别诈骗人员的身份,实现实时对诈骗人员的行为路径的掌握,辅助公安机关进行案件的侦破。

4.2 取证服务

声纹数据库系统可以对实时固定电话、手机、网络电话的等语音通讯工具的实时通话语音进行保存。公安机关如果通过系统的客户端和管理平台发现诈骗人员的行为,可以通过对声纹的身份标注并将录音储存的方式来实现对诈骗人员的诈骗犯罪行为的自动取证,并在第一时间将数据库中的案件语音文件和声纹识别结果回传,在整理和分析后又将结果推送到公安系统的办案客户端。

4.3 反制服务

建设声纹数据库系统的反制服务针对的主要是具有反

侦察意识诈骗人员和犯罪分子实施的,诈骗人员有的时候会伪装成政府机构,以获得受害人的信任而实施诈骗行为;有的诈骗人员甚至会通过虚拟信号码来实施诈骗行为,这无疑给公安机关对诈骗人员的身份识别造成极大的困扰。而本系统可以通过建立联动的预案来主动推送预警信息,通过在移动、电信、联通等通讯运营商的联合积极主动推送预警信息,以通话或短信的方式提醒受害人及其家人谨防电信诈骗,拦截诈骗行为,使得诈骗人员在追呼的过程中无法实现通话。

4.4 溯源服务

建设声纹数据库系统,可对已经确定的诈骗电话和诈骗人员等参与者的身份进行溯源查询,系统可以通过对声纹数据库进行检索、查询和回传通话记录,对诈骗的通话路径和行为进行反向的溯源查询,实现辅助公安机关侦破诈骗案件的功能。

5. 结束语

通过提取与鉴别语音中的声纹,能在一定范围内排查和确认声纹数据库中人的身份。犯罪者在电话中留下的语音信息,在很大程度上能对案件的侦破起到关键的作用。对诈骗人员建立声纹数据库,通过语音来甄别诈骗人员的身份信息,不仅能提升公安机关侦破诈骗案件的效率,减少公安机关的警力、物力的浪费,使公安机关能迅速打击诈骗犯罪行为,还能有效防治电信诈骗电话,直接减少人民群众的经济损失。

参考文献

- [1] 赵婉序,王海涛,蒋煜,杨舒凡,呼上枫,邢启明.数字经济时代大学生防范新型诈骗的对策研究[J].电信快报,2022(9):42-46.
- [2] 景晓东,魏庆凯,马杰,张国强.声纹识别辅助电信反诈骗系统研究[J].警察技术,2021(5):15-18.
- [3] 付立挺,董德武,夏利松.自然语言理解技术在电话诈骗识别预警领域的应用[J].警察技术,2021(5):19-23.
- [4] 刘晓晨,潘孝勤,曹金璇,芦天亮.声纹识别和语音识别技术在公安领域的应用[J].网络安全技术与应用,2021(4):153-155.
- [5] 张振清,张旭毅.国家声纹库在建过程中发现的问题及建议——以郑州声纹库试点工作为例[J].警察技术,2021(2):60-62.