

# 基于语音语义的电话预警通知技术研究

王彦霞<sup>1</sup> 王斌<sup>2</sup> 李翔<sup>2</sup>

1 河北省公安厅信息中心 2 清华海峡研究院(厦门)

DOI:10.32629/acair.v3i4.17942

**[摘要]** 随着互联网技术发展,电信诈骗手段不断升级,加之我国虚拟运营商数量较多,运行经验和水平参差不齐,导致容易发生电信诈骗,给社会诚信带来严峻挑战,严重威胁公众安全感。据统计,中老年人已成为受电信诈骗侵害的重点人群,以虚假信息诈骗为代表的电信网络新型违法犯罪严重扰乱正常生活秩序,危害老年人财产安全,成为影响老年人安全感和社会稳定的突出问题。本文聚焦基于语音语义的电信网络电话预警通知技术,通过融合智能语音、智能语义及大数据分析技术,构建电信网络诈骗预警体系。综合运用先进的人工智能算法模型,快速实现通话语音的反电诈分析与预警,并通过AI对话机器人实现快速触达提醒,有效防范电信诈骗的发生。

**[关键词]** 人工智能; 电信诈骗; 反电诈; 预警通知

**中图分类号:** TV **文献标识码:** A

## Research on Telephone Early Warning Notification Technology Based on Voice Semantics

Yanxia Wang<sup>1</sup> Bin Wang<sup>2</sup> Xiang Li<sup>3</sup>

1 Information Center of Hebei Provincial Public Security Department (PSD)

2 Tsinghua Straits Research Institute (Xiamen)

**[Abstract]** With the development of Internet technology, the means of telecommunications fraud have been constantly upgraded. Moreover, there are a large number of virtual operators in China, and their operation experience and service levels vary greatly, which leads to the frequent occurrence of telecommunications fraud, posing a severe challenge to social integrity and seriously threatening public security. According to statistics, the elderly have become the key group affected by telecommunications fraud. Telecommunications and network-based new types of criminal activities represented by false information fraud have seriously disrupted normal life order and endangered the property safety of the elderly, becoming a prominent issue affecting the elderly's sense of security and social stability. This article focuses on the voice and semantic-based telecommunications network phone warning notification technology. By integrating intelligent voice, intelligent semantics, and big data analysis technologies, a telecommunications network fraud warning system is constructed. Advanced artificial intelligence algorithm models are comprehensively applied to quickly achieve anti-fraud analysis and warning of phone call voices, and AI dialogue robots are used to achieve rapid contact and reminder, effectively preventing the occurrence of telecommunications fraud.

**[Key words]** Artificial intelligence; Telecommunication fraud; Anti-telecommunication fraud; Warning notice

随着互联网技术的发展,电信诈骗手段不断升级,加之我国虚拟运营商数量较多,运行经验和水平参差不齐,虚拟运营商诚信经营意识不强,监管不到位,导致容易发生电信诈骗,成为人们享受信息发达便利的绊脚石,严重威胁公众安全感。据统计,中老年人已成为受电信诈骗侵害的重点人群,以虚假信息诈骗为代表的电信网络新型违法犯罪,严重危害老年人财产安全,成为影响老年人安全感和社会稳定的突出问题。传统的电信网

络诈骗防范手段,如基于规则的关键词匹配等,在面对新型、复杂的诈骗手段时,显得力不从心,难以做到及时、精准的预警。在此背景下,基于语音语义的电信网络电话预警通知技术应运而生,旨在利用先进的人工智能技术,对电信网络通话进行监测与分析,实现对电信网络诈骗的高效预警,从而保护公众的财产安全。基于语音语义的电信网络电话预警通知技术研究,将结合国内运营商资源优势 and 人工智能企业的技术优势,综合运用先进

的人工智能算法模型、大数据分析技术,快速实现通话语音的反电诈分析与预警,并通过AI对话机器人实现快速触达提醒,有效防范电信诈骗的发生。

## 1 国内外相关研究

随着互联网技术、数字经济等加速发展,许多国家的电信网络诈骗案件持续增多,且呈现组织化、链条化、技术手段多变、跨境实施等特征,严重危害民众的财产和信息安全。不少国家持续出台措施,通过完善反诈骗立法、设立专门机构、升级技术手段等,强化打击电信网络诈骗犯罪,积极引导民众加强防范,积累了一些反诈防诈的有益经验。

针对层出不穷的电信网络诈骗案,多国政府都采取多种措施加强治理。法国政府推出专门服务网站,帮助用户有效屏蔽垃圾电话,用户也可在该网站举报诈骗电话。为应对花样繁多的电信网络诈骗,巴西政府要求巴西各大银行尽快完成对银行应用程序的技术升级,对每笔交易都进行人脸识别、指纹识别等生物识别技术认证,以提升账户安全性。阿联酋为遏制电信网络诈骗犯罪,阿联酋检方此前启动了一款名为“我的安全社会”应用程序,公众可通过上传图片、视频或语音片段,将诈骗行为发生的位置、时间和内容等信息发送给公诉机关。

我国电信网络诈骗犯罪形势严峻,已成为发案最多、上升最快、涉及面最广、人民群众反映最强烈的犯罪类型,其中刷单返利、虚假投资理财、虚假网络贷款、冒充客服、冒充公检法5种诈骗类型发案占比近80%,成为最为突出的5大高发类案。仅2023年,国家反诈中心累计下发资金预警指令940.6万条,会同相关部门拦截诈骗电话27.5亿次、短信22.8亿条,处置涉诈域名网址836.4万个<sup>[1,2]</sup>。中宣部、公安部、教育部、财政部等坚持广泛宣传和精准宣传相结合,组织开展“五进”宣传活动,不断提升群众识骗防骗能力。除了监管、执法部门外,科技公司也走在反金融诈骗第一线,作为市场的直接参与者,它们也逐渐摸索出了自己的一套反诈模型与方法论,将智能语音、智能语义技术应用到反电诈领域。智能语音识别技术将通话中的语音实时转换为文字,智能语义分析技术则能够理解文字内容的含义,通过对语义结构、情感倾向、话题主题等多维度的分析,精准识别出诈骗意图,并通过智能语音外呼机器人进行劝阻,满足电信网络高并发、实时性的要求,为大规模的诈骗预警提供了可能。

## 2 项目技术原理与架构

### 2.1 智能语音识别技术

智能语音识别技术的核心算法是整个技术体系的基石,20世纪80年代,语音识别开始从孤立词识别系统向大词汇量连续语音识别系统发展,基于GMM-HMM的框架成为语音识别系统的主导框架;但在进入21世纪后,语音识别系统的错误率依然很高,再次陷入漫长的瓶颈期。直到2006年提出用深度置信网络初始化神经网络,使得训练深层的神经网络变得容易,从而掀起了深度学习的浪潮。DNN代替传统高斯混合模型进行声学似然评估,保留了声学模型、语言模型和词典模型等所有组件,构成混合

ASR系统。最近几年兴起的是端到端(E2E)方法,E2E建模使用单个网络,将输入语音序列直接转换为输出令牌训练(output token sequence),推翻了传统ASR系统中的所有建模组件。神经网络部分一般采用Conformer神经网络。Conformer(Convolution-augmented Transformer)是一种结合了卷积神经网络(CNN)和Transformer的深度学习方法,旨在提升语音识别等任务的性能。它的设计目标是结合卷积层的局部特征提取能力和Transformer的长程依赖建模能力,从而有效地处理时序数据,尤其适用于语音和音频信号处理。本项目利用预训练模型已经学习到的语音通用特征和模式,减少为新应用场景开发语音识别系统所需的时间和资源,提高模型的训练效率和泛化能力,使模型能够快速适应新的领域或语言的语音识别任务。

为了让语音识别系统在实际应用中表现更出色,我们根据反电诈场景需要,对模型针对性优化,在核心的声学模型、语言模型、解码器上都有创新和突破,使其能更好识别专业术语和口音适配,更贴合实际需求。

### 2.2 智能语义分析技术

智能语义分析技术作为诈骗意图识别的核心支撑,需兼顾传统规则质检与深度神经网络语义模型的协同应用。该语义模型具备强大的逻辑表达能力,已实现十余种基础逻辑关系的覆盖,并支持从词级别到句子级别的多层级处理。通过逻辑关系的灵活组合与嵌套,能够有效应对各类反电诈场景中的复杂任务需求,为精准识别诈骗行为奠定技术基础。

技术架构上,该方案以自然语言处理(NLP)技术为核心,对语音识别转换后的文本内容开展深度语义解析。语义分析体系包含四大关键层面:

词法分析:通过精准分词与词性标注,构建文本分析的基础语义单元,为后续处理提供结构化输入;

句法分析:聚焦句子语法结构的构建,深入解析主谓宾、定状补等成分间的逻辑关系,厘清文本的语法脉络;

情感分析:专门针对文本情感倾向进行判断(积极/消极/中性),尤其对识别诈骗分子的诱导性话术、情绪操控手段具有重要价值;

话题识别:通过动态追踪通话内容的主题演变,捕捉话题切换中的异常模式,从而挖掘潜在的诈骗线索。

在实际应用中,该技术主要服务于对话内容的意图分类与风险研判。算法选型上,采用NLP领域领先的BERT算法作为核心框架,并创新性地引入迁移学习、知识蒸馏等优化手段,有效解决了原生BERT模型计算量大、响应速度慢的痛点,目前在通话内容语义提取精度上已达到业界领先水平。

针对实时性要求较高的场景,方案同步提供轻量级CNN算法作为补充。该算法借鉴图像领域的视觉感知思路,能够灵活捕捉不同语义视野与距离的特征信息;同时融合attention机制,实现跨距离语义关联的快速定位,在保障分析效率的同时进一步提升了识别准确性,可自适应多样化的语义分析场景需求。

### 2.3 智能语音合成技术

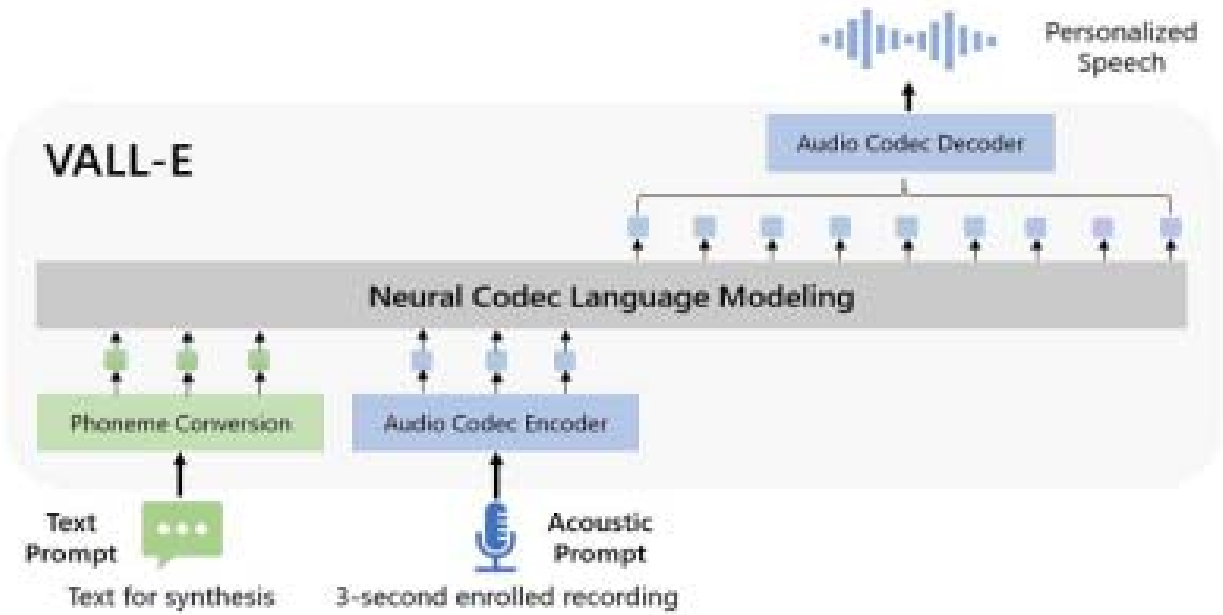


图1 语音合成工作原理图

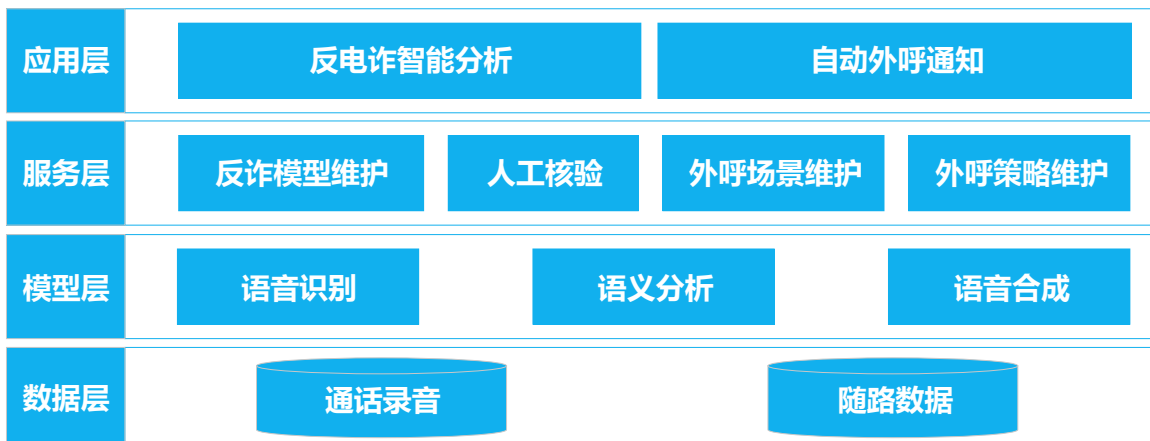


图2 电话预警通知系统架构图

智能语音合成技术通过算法、模型训练与调优,让机器能够生成自然流畅的语音,实现了从文本到语音的高效转换,具有合成音色优美、音质清晰、自然度高、韵律感强的特点,并支持多语种,满足不同地域应用需求。

本研究采用当下最流行的语音大模型 (Speech LLM),不但语音自然度达到真人水准,而且支持多语言混读;同时支持语音极速复刻 (Zero-Shot Clone),只需提供3-15秒示例语音,即可即时模拟该语音的音色和风格播报,从而具备高度的定制化能力,满足不同场景和用户的多样化需求。具体结构以经典VALL-E X为基础,适配本项目的具体需求改进。

在效率上,智能语音合成技术能快速将文本转化为语音,在自然度和表现力方面,越来越接近真人发音,增强内容的感染力。同时,它具备高度的定制化能力,满足不同场景和用户的多

样化需求。

#### 2.4系统架构设计

基于语音语义的电信网络电话预警通知系统采用分层架构设计,主要包括数据层、模型层、服务层和应用层。数据层负责从电信网络中获取通话数据,包括语音信号和相关的元数据,如通话时间、主叫号码、被叫号码等;模型层主要提供语音识别、语义分析、语音合成等人工智能模型能力,语音识别模型负责把电话录音转写成文字,为后续语义分析提供结构化文本数据,语义模型结合反电诈模型对通话录音进行反电诈分析,语音合成负责将文本转化为语音,用于外呼提醒的通话;服务层反电诈业务所需的反诈模型维护、人工核验,以及外呼场景维护和策略维护,支撑应用层反电诈智能分析子系统和自动外呼通知子系统。

### 3 关键技术实现与创新点

#### 3.1 语音语义融合的诈骗识别模型

传统的电信网络诈骗识别方法往往单独使用语音特征或文本特征,难以充分捕捉诈骗行为的复杂性。本项目创新性地提出了一种语音语义融合的诈骗识别模型<sup>[3]</sup>。该模型首先分别对语音信号和对应的文本内容进行特征提取。在语音特征提取方面,除了常用的梅尔频率倒谱系数(MFCC)、线性预测倒谱系数(LPCC)等特征外,还利用深度学习模型,如卷积神经网络(CNN)对语音的时频特征进行深层次的学习和提取。在文本特征提取方面,采用预训练语言模型,如BERT,获取文本的语义表示向量。然后,将提取到的语音特征和文本特征进行融合,通过拼接、加权求和等方式,得到融合特征向量。最后,将融合特征向量输入到分类器中,如支持向量机(SVM)或多层感知器(MLP),进行诈骗行为的分类判断。通过实验验证,这种语音语义融合的诈骗识别模型在准确率、召回率等指标上均优于单独使用语音特征或文本特征的模型,能够更精准地识别电信网络诈骗行为。

#### 3.2 实时动态更新的反电诈模型库

电信网络诈骗手段不断变化,新的诈骗话术和场景层出不穷。有专家指出,大数据侦查中数据失真导致的情报错误风险一定程度存在<sup>[4]</sup>。数据失真属于数据科学和情报学领域中“脏数据”范畴<sup>[2]</sup>,为了使预警系统能够及时适应这种变化,本项目建立了实时动态更新的反电诈模型库。模型库中存储了大量的诈骗案例、诈骗话术模板、语义特征等信息。系统通过实时监测电信网络通话数据,当发现新的疑似诈骗模式时,自动提取相关的特征信息,并与模型库中的已有信息进行比对。如果发现新的模式与知识库中的信息存在显著差异,则将其作为新的案例添加到模型库中。同时,利用智能语义技术,对模型库中的信息进行关联和整合,构建一个结构化的知识网络,预警系统能够不断学习和适应新的诈骗手段,经实测,预警命中精度保持在85%以上,具备较高推广价值。

#### 3.3 多模态数据融合与协同处理

除了语音和语义数据外,电信网络通话随路数据还包含丰富的其他模态数据,如通话时长、通话频率、主叫号码和被叫号码的归属地、通话时间分布等。本项目充分利用这些多模态数据,通过数据融合与协同处理技术,进一步提高预警的准确性和可靠性。在数据融合方面,采用特征级融合和决策级融合相结合的方式。特征级融合将不同模态数据的特征进行融合,形成一个综合特征向量,如将语音特征、文本特征与通话时长、频率等统计特征进行拼接。决策级融合则是分别利用不同模态数据训练独立的分类模型,然后根据各个模型的预测结果进行综合决策,如采用投票法、加权平均法等方式确定最终的预警结果。在协同处理方面,建立多模态数据之间的协同关系模型,通过分析不同模态数据之间的相互影响和关联,挖掘潜在的诈骗线索。例如,结合通话时长和语义内容分析,如果通话时长较长且语义内容中存在明显的诱导转账话术,则更有可能是诈骗通话。通过多模态数据融合与协同处理,预警系统能够从多个维度对通话数据

进行分析,有效提高了对电信网络诈骗的识别能力。

### 4 项目成效与社会影响

#### 4.1 预警拦截能力提升

通过实际应用案例可以明显看出,基于语音语义的电信网络电话预警通知系统在预警拦截能力方面取得了显著提升。与传统的防范技术相比,该系统能够更准确地识别电信网络诈骗行为,大幅降低了漏报率和误报率。在处理海量通话数据时,系统能够保持高效的实时处理能力,及时发现并拦截潜在的诈骗通话。

#### 4.2 保护公众财产安全

该项目技术的应用直接保护了公众的财产安全。通过及时准确的预警通知,民众能够在遭受电信网络诈骗时第一时间得知风险,避免进行转账汇款等操作,从而有效避免了财产损失。原本可能上当受骗的民众在接到预警后,成功避免了经济损失,保障了自身的财产权益。这不仅使民众个人受益,也在一定程度上减少了社会财富的流失,维护了社会的经济稳定。

#### 4.3 助力执法部门打击犯罪

对于执法部门而言,该系统提供了强有力的技术支持。系统能够快速准确地提供疑似诈骗线索,帮助执法部门缩小侦查范围,提高破案效率。通过对预警数据的分析,执法部门可以掌握电信网络诈骗的新趋势、新特点,为制定针对性的打击策略提供依据。

#### 4.4 推动行业技术发展

基于语音语义的电信网络电话预警通知项目技术的成功应用,为电信行业和人工智能领域的技术发展提供了有益的借鉴。该项目所涉及的人工智能语音识别、智能语义分析等技术的融合应用,为解决其他相关领域的问题提供了新的思路和方法。同时,项目的实施也促进了相关技术的进一步优化和创新,推动了整个行业技术水平的提升。

### 5 结论与展望

#### 5.1 研究结论总结

本文详细阐述了基于语音语义的电话预警通知技术,通过对电信网络诈骗现状的分析,明确了传统防范技术的局限性以及基于语音语义技术的优势。深入探讨了系统的技术原理与架构,包括智能语音识别技术、智能语义分析技术以及分层架构设计。重点介绍了项目的关键技术实现与创新点,如语音语义融合的诈骗识别模型、实时动态更新的知识库以及多模态数据融合与协同处理技术。通过实际项目实施过程和应用案例分析,验证了该技术在提升预警拦截能力、保护公众财产安全、助力执法部门打击犯罪等方面取得的显著成效。

#### 5.2 未来发展方向展望

随着人工智能技术的不断发展和电信网络环境的持续变化,基于语音语义的电话预警通知技术也面临着新的机遇和挑战。随着诈骗手段的日益复杂,在未来的研究和发展中,需要不断优化智能语音识别和智能语义分析模型,进一步提升模型的性能和泛化能力。

同时,衷心感谢研究团队中胡学营等成员在项目推进中的卓越贡献。他们在研究设计、数据采集与分析过程中展现了高度的专业素养和协作精神,为本研究的顺利完成提供了坚实基础。

最后,向所有为本研究提供间接支持的同仁、家人及朋友致以谢意。你们的理解与鼓励是我持续前进的动力。

#### [致谢]

本研究得到了河北省省级科技计划资助(项目名称:基于语音语义的电信网络诈骗电话预警处置技术及应用;项目编号:23375602D),在此致以诚挚的感谢。

#### [参考文献]

[1]徐婷.打击治理电信网络诈骗犯罪向纵深推进[N].人民公安报,2024-01-06(001).

[2]庄华,马忠红.电信网络诈骗犯罪预警的失灵与优化[J].情报杂志,2025,44(02):116-123.

[3]熊凯,杜理,丁效,等.面向文本推理的知识增强预训练语言模型[J].中文信息学报,2022,36(12):27-35.

[4]李小猛.大数据侦查情报的分析运用风险及其规制[J].情报杂志,2023,42(02):17-21.

#### 作者简介:

王彦霞(1983--),女,汉族,河北人,硕士,高级工程师,现任职于河北省公安厅信息中心,长期从事人工智能技术研究,致力于推动公安信息化建设的创新与应用,多次参与省级重大科技项目研发工作,具有丰富的理论基础与行业实践经验。

王斌(1984--),男,汉族,河北人,硕士,高级工程师,现任职于清华海峡研究院(厦门),拥有多项专利及学术论文,致力于人工智能语音领域研究与产业化解决方案落地。

李翔(1982--),男,汉族,湖南人,硕士,高级工程师,现任职于清华海峡研究院(厦门),致力于前沿AI技术的产业应用与创新,具备丰富的实践型经验。