

基于深度学习的智能语音识别技术探索

倪序南

天津国土资源和房屋职业学院 天津 300270

摘要: 智能语音识别技术作为人工智能领域重要分支,在近年来取得了突破性进展,该技术通过把人类语音信号转化成可识别数字信息,实现了人机交互方式的革命性变革,研究显示基于深度学习的语音识别系统准确率已突破95%,在多个应用场景展现出了显著优势,具体来说在智能家居领域实现了对家电设备的声控操作,在医疗健康领域辅助完成了电子病历的自动化录入,在工业制造领域提升了生产流程的智能化水平,本文运用文献分析法系统梳理语音识别技术发展历程,还对其未来在5G通信、物联网等新兴技术领域应用前景进行前瞻性探讨。

关键词: 语音识别; 应用; 发展

1 前言

跟着科技以飞速的状态发展,特别是信息技术不断实现突破,语音交互技术也就此应运而生,它的应用范围从传统人际交流领域,跨越到了人机交互的全新领域。现代技术给语言赋予了全新的生命力,人们能够通过语音指令直接操控各类机器设备,让机器设备可以依据接收到的语音信息,自动去执行相应的操作内容,这极大提高了人机交互的便捷性与效率。这一技术革新不光改变了人们的日常生活,还预示着未来人机交互模式会出现重大转变。语音识别技术作为智能家居的核心组成部分,正逐渐发挥着十分关键的作用,凭借语音识别技术,智能家居系统能够精准理解并响应用户的语音指令,进而实现对家电设备的智能控制操作。就像搭载语音识别功能的智能电视,用户不用进行繁琐的遥控器操作,只要简单说出相关指令,电视就能快速切换频道、调节音量、搜索节目等,这极大提升了用户的使用体验感受。人工智能技术的广泛运用,正在深刻改变着人类的生活方式,语音识别技术的进步是人工智能技术在日常生活中应用的一个体现。它不但让智能家居设备变得更加人性化,还使科技对生活智能化的推动作用更加显著。在未来的发展进程中,随着语音交互技术不断成熟且普及开来,我们能够预见到人工智能会更深入地融入人们的日常生活,为人们打造更加便捷、高效、舒适的生活环境。

2 智能语音交互

随着人工智能技术持续不断地进步,智能语音交互系统的性能与适用性显著提升,在语音识别领域,深度学习算法的应用让系统对语音信号解析能力大幅增强,能有效识别多种口音和不同说话人,为提高语音识别准确性和鲁棒性,研究者致力于优化算法模型采用端到端

架构等策略,在语义理解模块,自然语言处理技术进步为系统提供更精准文本解析能力,通过对上下文信息进行综合分析,系统能更好理解用户意图实现复杂智能交互体验,为适应不同场景和各类用户需求,语义理解模块采用多模态信息融合技术结合多种信息来源提高理解准确性,语音合成技术也处于不断进化的状态,从早期规则基础合成到如今基于深度学习方法,音质和流畅度显著提升,为满足用户个性化方面的需求,语音合成模块支持自定义音色和语调让交互体验贴近自然,在系统部署方面,随着云计算技术不断普及,越来越多智能语音交互系统采用混合云架构实现资源高效利用。同时,为了保障用户隐私和数据安全,系统设计者还需关注数据加密、访问控制和隐私保护等方面的技术。未来,随着人工智能、大数据、物联网等技术的深度融合,智能语音交互系统将在更多领域得到应用,如智能家居、智能医疗、智能交通等。

3 理论基础

智能语音识别系统的核心技术架构主要涵盖三个关键环节:首先是对语音信号进行特征参数提取,其次需要建立合适的声学模型以确定识别单元,最后通过算法训练优化模型性能。由于人类语音具有显著的个体差异和环境变异性,这给智能语音识别技术的研发带来了诸多技术难题。

3.1 特征提取技术

在语音信号处理这个领域当中,特征参数提取技术起着关键作用,它的主要功能是从人类发声信号里获取有表征意义的参数集,以此来解析说话者的意图表达,语音信号属于复杂的多维信息载体,其处理过程一开始就要进行特征提取,这一步骤能够筛选出对智能语音识别系统有实用价值的键信息,通过这一技术手段,可

以实现语音数据的深度解析与处理,能有效剔除冗余信息并保留反映语音本质特征的核心参数,其中线性预测倒谱系数(LPCC)作为基于声道建模理论的代表性方法,已经成为语音信号分析领域广泛应用的技术方案,线性预测分析(LPC)技术通过建立语音样本间的线性关系模型,能够获取一组精确描述语音频谱特性的模型参数。

3.2 语音信号建模

在语音识别进行建模的过程当中,模式匹配法是主要采用的技术手段,从建模单元的角度来看,可分为单词、音节和音素这三个层级,其中音节单元在中小词汇量的汉语语音识别系统里具有较好的适用性,这主要是源于汉语作为单音节语言这样的特点,然而对于英语等多音节语言的大词汇量识别系统而言,音节单元的适用性则相对比较有限,研究表明在中大规模汉语语音识别系统当中,采用音节作为识别单元具有较高的可行性,从语音学方面的角度来看,音素作为最小的语音单位是区分不同词汇的基础,在实际应用过程当中通过将声母和韵母进行差异化组合形成细化声母,虽然会导致模型数量有所增加但能显著提升对易混淆音节的区分能力。

3.3 模型训练技术

模型训练过程按照特定优化准则来进行,通过分析海量已知模式样本提取本质特征参数,构建可实现未知输入模式与模型库最优匹配的数学模型,以此显著提升声学模型的识别精度。当前智能语音识别领域主要采用两种模型训练的方法,一种是基于高斯混合模型-隐马尔可夫模型(GMM-HMM)的声学建模技术,另一种是人工神经网络(ANN)建模方法。其中GMM-HMM方法着重优化高斯混合模型里均值向量和协方差矩阵等关键参数,这些参数的精确训练直接决定了声学模型的性能表现。而人工神经网络方法是通过模拟生物神经元工作机制,采用类似原理来计算状态发射概率。

4 基于深度学习的智能语音识别技术应用

4.1 智能家居方面

基于人机交互技术的智能家居系统大幅提升居民生活便捷程度,对老年群体和儿童等特殊用户特别友好,当前主流智能家居解决方案普遍整合语音识别模块,该技术广泛应用于照明调节、温控系统等多个功能模块,除上述应用场景外,语音交互技术还可拓展到厨房电器和清洁设备等生活领域,在视听设备控制方面,传统交互模式难适应用户不断增长功能需求,引入智能语音识别系统后视听设备能即时语音响应,显著提升操作效率,具体实现机制是智能终端设备预装专用语音识别应

用程序,采集语音数据经网络传至云端服务器解析,再将处理结果反馈至终端执行指令,这种交互方式支持丰富功能操作,比如通过语音指令“切换频道”可实现节目快速检索功能。

4.2 车联网方面

目前的研究显示传统车载人机交互系统要求驾驶者在操控车辆期间还要进行手动操作,这样的操作模式明显增加了驾驶员的认知负荷与注意力分散风险,相关统计数据表明驾驶过程中操作电子设备让视线偏离前方道路是引发交通事故的重要因素之一,近些年来随着智能语音识别技术在车联网领域深入应用,该系统有效减轻了驾驶员在复杂路况下的操作负担,借助语音交互方式大幅提升了行车安全性,我国在车载语音交互系统研发领域已经取得突破性进展,尤其在有限词汇量语音识别技术方面已达到国际先进水平,这项技术能够实现包含车辆控制指令、语音通讯、导航设置以及娱乐系统操作等多项功能,就拿车载免提电话系统来说驾驶员只需通过语音输入电话号码系统就能自动完成拨号操作,需要注意的是在车辆密闭空间内语音识别准确率保持在较高水平,这种创新的人机交互模式不但解放了驾驶者的双手和视线还显著改善了驾驶体验。

4.3 医疗方面

医疗业务量一直持续攀升,使得医护人员在书写检查报告、病程记录还有门诊病历等医疗文书方面耗费大量时间资源,智能语音识别技术在医疗领域的应用有创新性改变,通过语音输入替代传统手工或键盘录入模式,显著提升了医疗文书处理的效率,实证研究表明,该技术能让门诊医生通过语音描述患者症状,经系统自动转化为高准确率的电子文档,再经快速校对便可完成病历录入,大幅降低了医务人员的工作负荷,从系统运行层面来说,这一技术革新不但优化了医疗资源配置的效率,还显著降低了医疗机构的运营成本,临床数据显示,语音识别技术的应用让单例患者就诊时间平均缩短23.5%,有效提升了医疗机构的服务吞吐量,为改善医疗服务供给能力提供了技术支撑。

4.4 军事方面

智能语音识别技术在国防军事领域体现出明显应用价值,当下针对军事应用场景研发的专用语音识别系统,在识别准确率、实时响应和环境适应性等关键性能指标上有显著提升,在指挥自动化系统中这项技术已经被深度整合应用,军事模拟训练领域引入智能语音识别技术后,其人机交互界面得到明显优化,借助自然语言这一最契合实战需求的信息载体,有效提高了部队协同

训练的真实性,在飞行控制系统方面语音识别技术的应用,让飞行员能够用语音指令替代传统手动操作,既提升了航空作战效能又降低了飞行员工作负荷。

5 智能语音识别技术面临的问题

5.1 环境中的噪声干扰

语音信号具有显著的多样性和复杂性特征,在存在环境噪声干扰的情况下,其识别准确率会明显下降。现有研究表明,在实验室理想条件下(室内安静环境)语音识别准确率可达97%,然而现实应用场景中普遍存在各种噪声干扰,目前尚未形成能够有效消除环境噪声影响的系统性解决方案。

5.2 不标准的语音识别率比较低

在机器学习技术快速发展的背景下,智能语音识别系统取得了显著进展,然而其应用仍存在若干局限性。现有语音识别系统主要针对标准普通话使用者设计,对带有方言特征的语音样本识别准确率较低。值得注意的是,当前我国相当比例人群的普通话发音存在不同程度的方言特征,这种语音变异现象导致识别系统输出结果与用户预期产生偏差。

5.3 容错率的处理

在语音识别技术的实际应用过程中,由于发音者可能存在方言口音、个人发音习惯差异、临时性口误或专业术语发音不标准等多种复杂情况,导致语音信号在声学特征提取环节就出现偏差,进而在语义解析和自然语言处理阶段产生连锁反应式的识别错误,最终显著降低了系统的整体识别准确度。针对这一行业普遍存在的技术瓶颈问题,当前主流的解决方案主要包括两种途径:其一是依赖人工干预的后处理机制,即通过专业标注人员对识别结果进行逐条修正;其二是要求用户重新录入语音数据,但这会显著降低用户体验和工作效率。值得注意的是,部分领先的科技企业如谷歌、科大讯飞等已开始投入大量研发资源,致力于开发基于深度学习和上下文理解的新型语音识别系统,这类系统通过构建更强大的语言模型和语义理解网络,能够自动识别并修正语序混乱、语法错误或语义不连贯的语句,同时结合用户

个性化发音特征进行自适应调整,从而在保持实时响应速度的前提下,显著提升复杂场景下的语音识别准确率,这一技术突破有望在未来3-5年内实现规模化商业应用。

6 未来展望

智能语音识别作为跨学科前沿技术需建统一标准化体系,以此优化各专业领域协同合作机制推动社会经济高质量发展,在互联网技术迭代升级、企业研发投入增加及国家政策扶持等多重因素驱动下,这一关键性人机交互技术正吸引越来越多科研力量投入当前我国智能语音识别领域面临噪声抑制、方言识别准确率提升和系统容错能力优化等技术瓶颈,这些问题严重制约该技术产业化进程,值得注意的是随着智能终端设备快速普及,语音识别技术蕴含巨大商业价值和发展潜力必将吸引更多行业领军企业布局该领域,目前国内互联网巨头都在进军语音市场,像百度小度语音助手、小米小爱音箱等都已陆续问世,截至目前我国在语音识别技术研究上取得一定进步,智能手机是应用语音识别功能最多的设备可实现语音打字、语音聊天等功能,未来我国智能语音识别技术会随科技发展逐渐进步并最终应用到生活方方面面。

参考文献

- [1]任丽娜.一种抗噪声的语音识别方法研究[J].电子测试,2021(13):70-71.
- [2]李素梅,侯秀丽.基于语音识别的智能晾衣系统设计[J].科学技术创新,2021(12):163-164.
- [3]侯强,侯瑞丽.机器翻译方法研究与发展综述[J].计算机工程与应用,2019(10):30-35.
- [4]惠子,郑昱,等.一种基于离线语音输入技术的检测数据记录系统[J].电子测试,2018(11):68-69.
- [5]古典.语音识别中神经网络声学模型的说话人自适应研究[D].中国科学技术大学,2018:77.
- [6]李晓朋.语言识别技术的应用[J].电子技术与软件工程,2016(23):156.