

基于深度学习的方言语音识别技术探究

陶薇薇

四川文化产业职业学院，四川 成都 610000

DOI: 10.61369/TACS.2026040006

摘 要： 方言属于地域文化的重要载体，它承载着特定群体的历史记忆和文化基因，但是，其在传承时会遇到较为严峻挑战，这样会对方言的发展与传承产生极大阻碍作用。通过将深度学习技术引入语音识别领域，能够为方言语音的识别提供一个更为高效、精准的技术路径，它也可以成为方言数字化保护的核心。鉴于此，本文将针对深度学习下的方言语音识别技术展开分析，并结合发展中的问题提出一些策略，仅供各位同仁参考。

关 键 词： 深度学习；方言；语音识别；技术

Exploration of Dialect Speech Recognition Technology Based on Deep Learning

Tao Weiwei

Sichuan Vocational College of Culture and Industry, Chengdu, Sichuan 610000

Abstract： Dialects are important carriers of regional culture, bearing the historical memories and cultural genes of specific groups. However, they face severe challenges in inheritance, which greatly hinder the development and inheritance of dialects. The introduction of deep learning technology into the field of speech recognition can provide a more efficient and accurate technical path for dialect speech recognition, and it can also become the core of dialect digital protection. In view of this, this paper will analyze the dialect speech recognition technology under deep learning and put forward some strategies combined with the existing problems in development, for reference only.

Keywords： deep learning; dialects; speech recognition; technology

一、基于深度学习的方言语音识别技术的价值

（一）有利于完善语音识别研究体系

传统的研究多是集中在普通话以及主流的外语方面，人们对于方言语音识别的研究较为不足，通过引入深度学习技术，能够有效打破传统算法的瓶颈，我们可以更为精准地捕捉方言的韵律与声调，这样可以有效填补现有技术的空白，有利于推动语音识别向多语种、多方言的方向拓展^[1]。此外，开展基于深度学习的方言语音识别技术的研究还能进一步推动深度学习跨领域创新，方言语音识别本身有较强的特殊性，它也对深度学习模型的特征提取提出了更高的要求，这就需要对现有模型展开改革与优化，设计更多适配低资源场景的轻量化模型，这样能大幅提升模型在复杂语音场景的应用能力。

（二）有利于方言文化活态传承

方言是地域文化的重要载体，它也是中华民族文化多样性的直观体现。现阶段，受到城市化、普通话普及等方面的影响，很多方言面临着传承断层的问题，很多小众方言开始逐渐消亡。通过开展基于深度学习的方言语音识别技术改革与应用，能够更为有效地实现方言语音精准转写、存储与检索，这样可以逐渐打造一个方言语音语料库与文化数据库，能够更好地将口头文化遗产数字化保存^[2]。同时，我们还可结合深度学习开发一些方言学习APP等，这样可以更好地推动方言活态传承，有利于守护地域文

化根脉。

（三）有利于推动智能语音产业升级

现阶段，我国的智能语音产业快速发展，但是很多产品的方言语音识别存在一定的不足，这样会对产品的覆盖产生阻碍作用。通过开展基于深度学习的方言语音识别技术改革与创新，能够进一步推动智能语音产业的升级，更好地填补方言市场的空白，有利于拓展产业应用边界，各个相关企业可以开发更多方言智能终端以及方言语音输入等产品，这样可以极大扩大市场覆盖面，创造更多经济价值^[3]。

二、基于深度学习的方言语音识别技术面临的问题

（一）数据稀缺、质量参差不齐

数据是开展高质量模型训练的基础，充足高质量的方言语音数据是保证方言识别精度的重要因素，但是，当前很多方言语音的数据存在诸多问题。在实际工作中，方言数据本身较为稀缺且采集难度很大。方言通常使用的范围比较小，而且缺乏一个较为统一的标准，多数方言也没有一个统一的书面形式作为记录，这样对方言的数据采集产生很大阻碍作用。此外，方言本身的发音随意性较强，这也在一定程度上增加了采集与标注的难度，很多方言的使用者多是集中在一些特定的地域且多是中老年人，这也为方言的数据采集增加了难度^[4]。

在进行方言的数据采集时，会受到设备、环境等多方面因素的影响，这样会导致一些语音存在噪声、失真等问题，这样会对后续的资源应用产生不良影响。在采集过程中，也缺乏一个较为统一的标准，标注人员的专业水平也参差不齐，经常会出现标注错误、遗漏等问题，还有一些数据存在重复冗余的情况，这样也会大幅降低数据的利用效率。

（二）模型适配性不足、泛化能力弱

现阶段，主流的深度学习模型多是针对普通话或主流外语设计，这就导致其对方言的适配性较为不足，存在能力弱、训练成本高等问题，这样会对基于深度学习的方言语音识别技术的应用产生阻碍作用。一般来说，方言与普通话在声调、韵律等方面存在明显不同，如粤语九声六调等，但是，主流的模型并没有充分考虑这些特性，这就导致其很难精准捕捉方言的核心特征。此外，很多模型的泛化能力较弱，不同方言的语音差异较大，同一方言内部也会因为口音、人群等存在一定的差异，这就导致基于特定数据训练的模型在遇到没有见过的口音以及场景时，会出现识别精度大幅下降的情况^[9]。

（三）环境抗干扰能力弱，场景适配性不足

实际应用中，复杂的环境干扰与多样场景需求会进一步凸显该技术在环境适配方面的不足。当前的方言语音识别技术本身抗干扰能力较弱，在一些实际的场景中，环境噪声、设备噪声等都会对技术的应用产生干扰，这样会在很大程度上影响语音特征的提取，从而导致识别精度下降。方言数据稀缺也会让数据的增强效果有限，使其很难应对复杂的噪声环境，嘈杂场景以及低成本设备下，很容易出现识别精度下降的情况^[10]。此外，不同场景的方言语音特点差异大，当前的语言模型多是基于单一场景数据训练产生的，它们缺乏对多场景特征的学习，这样会导致日常对话模型难以识别方言戏曲等专业场景语音，整体的场景适配性较为不足。

三、基于深度学习的方言语音识别技术的优化路径

（一）数据层面：构建高质量、多元化方言语音数据集

数据作为深度学习模型训练的核心，其质量、规模以及分布会在很多程度上影响方言语音识别的精度，为此，我们要提升基于深度学习的方言语音识别技术改革，需要进一步解决当前的数据瓶颈。在实际工作中，我们可以尝试从采集、预处理、共享以及增强四个方面展开协同发力，这样可以逐渐打造一个更高质量的方言语音数据集

我们要不断加大对各个方言区的数据采集力度，重点关注一些小众方言、濒危方言等，在采集时，我们可以尝试与政府、企业、高校等展开合作，组建一个更为专业的采集团队，而后方可深入方言核心区域全面采集不同年龄段、性别、口音的方言语音数据。在采集数据时，我们应保证数据涵盖日常对话、专业领域以及民俗活动等诸多场景，这样可以有效扩展数据覆盖的全面性。同时，我们应进一步规范采集流程，尽可能选用一些高精度的采集设备，在采集时，还需严格控制采集环境的噪声干扰，这

样可以有效避免语音数据的失真^[7]。

为提升采集数据的效率，我们可以尝试引入众包的模式，鼓励方言使用者主动参与到数据的采集中，这样能够大幅降低采集数据的成本，也能在一定程度上扩大方言数据采集的规模，有利于弥补专业采集团队的覆盖短板。在数据预处理环节，我们可以尝试采用傅里叶变换、小波变换等先进的信号处理技术，这样可以有效去除语音数据中的环境噪声、设备噪声等，还能进一步增强语音信号的强度，有利于提升方言数据的清晰度。不仅如此，我们还需对所有语音数据进行标准化处理，尽可能统一采样率、时长以及格式等关键参数，这样可以确保数据的一致性。为提升基于深度学习的方言语音识别技术优化效果，我们可以尝试建立一个更为严格的标注标准，组建一个由语言学专家与技术人员组成的专业标注团队，这样可以对语音数据的音素、声调等进行更为精准的标注。

（二）模型层面：优化模型结构，提升适配性与泛化能力

深度学习模型是方言语音识别技术优化的核心，它的适配性、泛化能力以及训练效率会在很大程度上影响技术的应用效果，为此，我们可以结合方言语音的独特性对模型进行一个全方位的优化，以此方可实现更为精准地识别与高效训练。在模型结构优化方面，我们要主动打破单一的声学特征的局限，尽可能融入 MFCC、FBANK 等互补声学特征，还可以结合实际情况设计一个专门的特征融合模块，这样可以更好地整合不同特征在频谱细节、时序动态性的优势，构建一个更为全面的方言声学表征。

在提升模型泛化能力方面，我们应进一步扩大训练数据的覆盖面，将不同地域、口音、年龄段的方言语音数据纳入训练范围，这样可以让模型更为充分地学习多样化的方言语音特征，从而有效减少地域、人群差异带来的识别偏差^[8]。我们可以广泛采用迁移学习、半监督学习等先进训练方法开展基于深度学习的方言语音识别技术优化，我们可以基于迁移学习将普通话语音识别模型的参数迁移到方言模型训练中，利用普通话数据的学习成果辅助方言模型训练，这样可以有效缓解方言数据稀缺的问题。在降低训练成本以及推动模型轻量化方面，我们要进一步加强与云计算平台合作，利用云计算的强大计算资源降低中小科研机构、企业的计算成本，同时，这样也能进一步优化模型训练的算法，减少训练过程中的参数调整次数，从而大幅缩短训练时间，提升训练效率。通过模型压缩、量化等技术，可以进一步减少模型的参数数量与计算量，在保证识别精度的前提下降低模型的训练成本与部署成本。

（三）环境适配层面：增强模型抗干扰能力，提升场景适配性

在开展基于深度学习的方言语音识别技术优化时，我们要进一步提升方言语音识别技术的抗干扰能力与场景适配性，这就需要针对环境干扰和场景的差异采取一些更具针对性的优化措施，这样才能确保技术在不同场景下的精准识别。在增强模型抗干扰能力方面，我们可以重点优化噪声抑制以及语音增强技术，利用基于深度学习的噪声抑制模型精准识别并过滤环境噪声、设备噪声等干扰，而后方可结合空间滤波技术强化目标声源信号，这样

可以大幅提升语音数据的信噪比与清晰度^[9]。

在模型训练过程中，我们可以主动加入一些含有噪声的方言语音数据，这样可以让模型更为充分地学习噪声环境下的语音特征，有利于提升模型对噪声的适应能力。同时，我们可以采用数据增强技术模拟不同强度、类型的噪声环境，生成更为多样化的含噪声语音数据，这样可以扩大训练数据的多样性，有利于增强模型的抗干扰能力。此外，我们可以进一步优化语音端点的检测技术，通过精准识别方言语音的起始与结束位置避免噪声片段对

识别结果的影响，这样可以确保即使在复杂噪声环境下也能精准提取有效语音特征，有利于提升基于深度学习的方言语音识别技术的精度。在提升场景适配性方面，我们可以尝试构建一个多场景方言语音数据集，全面采集日常对话、方言戏曲、政务服务等各类场景的方言语音数据。针对不同场景的语音特点，我们可以对模型展开一些针对性的优化，以此促使基于深度学习的方言语音识别技术优化质量提升到一个新的高度^[10]。

参考文献

[1] 黄雪琴, 张良峰, 耿强. 基于 WeNet 模型的海南方言语音识别技术研究 [J]. 电脑知识与技术, 2025, 21(36): 26-28.

[2] 陈嘉, 陈章华, 何侠, 等. 基于多模态标注技术的潮汕方言语音识别系统在广电融媒体中的应用研究 [J]. 广播电视信息, 2025, 32(12): 62-66.

[3] 吕冰清, 李秀志, 何军成, 等. 方言语音识别技术在老年人智能设备中的应用研究 [C]// 郑州市社会学学会, 郑州大学社会工作系. 2025 年社会学研讨会新质生产力分论坛论文集 (下册). 桂林电子科技大学北海校区; , 2025: 91-92.

[4] 郝博阳. 基于深度学习的方言语音识别技术研究 [D]. 长春大学, 2024.

[5] 许鹏远. 基于深度学习的河南方言识别研究 [D]. 西安石油大学, 2024.

[6] 毛跃辉. 方言语音识别关键技术研究及其在空调上的应用 [J]. 家电科技, 2022, (S1): 167-171.

[7] 何铭. 面向藏语的语音识别技术研究 [D]. 天津大学, 2021.

[8] 张敬康. 基于深度学习的大同方言语音情感识别 [D]. 中北大学, 2020.

[9] 刘晓峰. 山西大同地方方言语音识别技术及应用研究 [D]. 中北大学, 2020.

[10] 许彬彬, 洪青阳, 李柏翰, 等. 基于闽南方言语音识别技术的数据库设计与标注 [C]// 中国中文信息学会语音信息专业委员会. 第十四届全国人机语音通讯学术会议 (NCMMSC'2017) 论文集. 厦门大学中文系; 厦门大学智能科学与技术系; 厦门第六中学; 厦门大学电子工程系; , 2017: 32-37.