

基于深度学习的学习者情感识别技术算法优化研究

曾光辉¹
ZENG Guanghui

摘要

学习者情感的表现具有模糊性和时空分离特性，情感状态之间的界限不清晰，如反映在频谱图像上时，不同时间的语音对应的频谱特征可能差异较大。传统识别技术算法往往基于较为固定的特征提取模式，难以准确地从复杂多变的频谱图像中提取关键局部特征，如频谱峰值、谷值等，导致加权和未加权准确率较低。为此，文章提出了一种学习者情感识别技术算法的优化方法。利用 K-均值聚类方法，根据语音信号的内在结构，将相似的语音片段进行聚类，生成学习者的语音频谱图像。采用传递函数对生成的语音频谱图像进行加重处理，以改善频谱图的平滑性和均衡性，增强频谱图像中的关键信息。将经过加重处理后的学习者语音频谱图像作为输入向量，输入到深度学习网络的卷积层。在卷积层中，针对加重处理的频谱图像进行特征提取，捕捉频谱峰值、谷值等关键的局部特征。在全连接层，将这些特征映射到情感类别空间，以实现情感分类。根据深度学习网络的输出，将识别样本划分到概率最高的情感类别中，完成学习者情感的识别。经实验证明，经过优化后的学习者情感识别算法，在加权和未加权准确率方面均实现了显著提升，显示出在学习者情感识别领域良好的应用潜力和前景。

关键词

深度学习；K-均值聚类方法；学习者；情感识别；算法优化

doi: 10.3969/j.issn.1672-9528.2025.03.004

0 引言

在数字化教育蓬勃发展的今天，理解学习者的情感状态对于提升教育质量具有至关重要的意义。学习者在学习过程中所经历的情感，如专注、困惑、沮丧或兴奋等，不仅影响着他们的学习体验，还对学习效果、知识吸收以及学习动力有着深远的影响。随着人工智能技术的不断进步，学习者情感识别技术应运而生，可以通过各种传感器和数据挖掘手段，自动识别学习者的情感状态。然而，现有的学习者情感识别技术算法仍面临诸多挑战。一方面，学习者情感表达的复杂性和多样性使得准确识别变得困难。不同的学习者可能在相同的学习情境下表现出不同的情感反应，而且情感的外在表现可能受到多种因素的干扰，如个体性格差异、文化背景以及学习环境等。另一方面，目前的算法在准确性、实时性和普适性等方面存在一定的局限性。在实际的教育场景中，需要能够快速且精准地识别学习者的情感，以实现及时的教学

干预，但现有的算法往往难以完全满足这些要求。因此，对学习者的情感识别技术算法进行优化研究，具有重要的理论意义和实践价值。

目前情感识别技术算法研究已经取得了一定的成果。例如，文献 [1] 提出了基于改进 K-均值聚类的识别技术，优化聚类中心初始化和距离度量来分类语音特征以识别情感。学习者情感模糊时，情感界限不清使语音特征分布复杂、模式难定。该技术聚类中心初始化难以确定代表不同情感的初始中心，可能错分语音特征，影响准确率。文献 [2] 提出了多模态数据融合的识别技术，整合多种数据并提取、融合特征以识别情感状态。在学习者情感具有模糊性和时空分离特性时，不同模态数据在不同时间点情感表达可能不一致，特征融合时会干扰关键局部特征提取，使情感识别准确性下降。文献 [3] 从语音生成的角度重新探讨了声门波与声道特征对语音情感识别的可能性，提出一种基于源-滤波器模型的声门波和声道特征语音情感识别方法。首先，利用 Liljencrants-Fant 和 Auto-Regressive eXogenous (ARX-LF) 模型从语音信号中分离出情感语音的声门波和声道特征；然后，将分离出的声门波和声道特征送入双向门控循环单元 (BiGRU) 进行情感识别。该技术在情感识别时，会受到使用环境的影响，延长识别耗时。

1. 广州工程技术职业学院 广东广州 510900

[基金项目] 2024 年度广东省普通高校特色创新类项目“基于深度学习的在线学习智能评价方法研究”(2024KTSCX318)；2024 年度广州市高等教育教学质量与教学改革工程产教融合实训基地项目“新一代信息技术产教融合实训基地”(2024CJRHJD008)

针对上述方法存在的一些不足,本文提出了一种基于深度学习的学习者情感识别技术算法的优化研究方法。

1 学习者语音频谱生成

语音信号具有时空分离特性,是复杂的非平稳信号,其情感信息多隐藏于频谱特征中。因此,需将语音信号从时域转换到频域来有效提取情感特征,通过生成学习者语音频谱能直观展现语音的频谱特性,像不同频率成分的能量分布、频谱峰值和谷值等,这些特征对情感识别分析极为关键,可反映发音者的情感状态(如高兴、悲伤、愤怒等)。传统情感识别技术用短时傅里叶变换(STFT)处理语音信号得到时频谱图,但包含较多冗余信息,加大了后续计算复杂度。为此,本文提出用K-均值聚类算法生成学习者语音频谱^[4]。先对学习者的音频文件进行切割和标签分配,然后通过K-均值聚类进行分组,生成学习者语音频谱图像,以此提高数据的处理效率和准确性。

在处理学习者音频信号以生成用于情感识别的关键频谱图像时,首先面临的是如何有效地从原始音频中提取出包含情感信息的关键片段。为此,将原始的学习者音频文件切割成多个均匀且相对独立的片段。经过对不同帧持续时间的细致观察与分析,最终优化选择了500 ms作为分割窗口大小,以此将单个学习者的话语转换成多个片段^[5]。同时,为了确保学习者情感信息的一致性和连贯性,本研究为每个话语的所有片段分配了一个统一的标签。这一过程可以形式化地表示为:

$$K_n = \int_{n=1}^S x_n \cdot d \quad (1)$$

式中: K_n 表示被分割的 n 个学习者语音片段; S 表示学习者语音; x_n 表示分割窗口; d 表示分割窗口时间长度,即500 ms。

接下来,利用K-均值聚类算法对这些相似的学习者语音片段进行分组^[6]。具体而言,随机选取 k 个观测值(即语音片段)作为初始的类簇中心点,然后计算每个采样点到这些簇中心点的距离,并将采样点分配给距离最近的簇。在本研究中,采用欧几里得距离矩阵来度量样本间的距离。但考虑到人脑的视觉感知部分是基于非线性处理系统进行语音处理和识别的,引入径向基函数(RBF)来计算两帧之间的差值,以优化学习者情感识别技术算法的性能^[7]。径向基函数的基本思想是通过核函数将向量从线性不可分的低维度空间映射到线性可分的高维度空间^[8]。在本文中,径向基函数的输出用于表示学习者音频中两帧之间的差值,其形式化表示为:

$$y = \sum_{n=1}^{K_n} \omega_n R \quad (2)$$

式中: y 表示输出的差值结果; R 表示权重系数; ω_n 表示径向基函数的一维高斯激活函数。使用径向基函数作为映射函数,可以确定学习者音频片段之间的相似度。当学习者语音信号的特定片段相关性较大时, y 的值会相对较小;如果语音片段不相关,则 y 的值会很大^[9]。根据径向基函数计算出的差值 y ,对学习者的音频片段进行聚类。从每个聚类中选出一个靠近聚类中心的片段作为关键片段,然后采用短时傅里叶变换算法(STFT)将选定的关键片段转化为谱图进行二维表示。这一过程可以形式化地表示为:

$$x = \psi(t) \left[e(c_y + b) \right]^t \quad (3)$$

式中: x 表示关键音频的频谱图; $\psi(t)$ 表示窗函数; e 表示傅里叶变换的复指数; c_y 表示算法输入的学习者语音关键片段; b 表示信号在时间的复数幅度; t 表示信号在频率的复数幅度^[10]。通过以上步骤,成功生成了学习者语音的关键频谱图像,这些图像将用于后续的情感识别分析。

2 学习者语音频谱图像加重处理

在语音信号处理中,特别是针对情感识别任务,高频与低频区域的能量分布差异以及能量波动模式是重要的分析对象。然而,由于个体发声时的口腔辐射特性,高频区域能量衰减较大,低频区域衰减较温和,这种能量分布不均衡现象对语音分析和情感识别构成了挑战。为了缓解这一问题,本文引入了加重技术来调整频谱分布。这一技术的核心目的是通过增强高频区域的能量,以减小高频与低频区域之间的能量差异,改善频谱图的平滑性,使得频谱图像更加均衡,有利于后续的情感特征提取和识别分析。这一过程的实现依赖于传递函数,其公式表示为:

$$x^* = F - ux^{-1} \quad (4)$$

式中: x^* 表示加重处理后的频谱图像; F 表示传递函数; u 表示学习者语音频谱图像低频与高频差值。利用以上公式,可以对生成的频谱图像进行加重处理。通过设计的传递函数 F ,确保高频区域的能量得到显著增强,而低频区域的能量则保持相对稳定,从而改善频谱图的平滑性和均衡性,为后续的情感特征提取和识别分析打下坚实的基础。

3 基于深度学习的情感识别

频谱图中的频谱峰值、谷值等局部特征是反映学习者情感状态的重要线索。深度学习网络中的卷积层能够对加重处理后的频谱图像进行多层卷积操作,通过预设的卷积

核进行点积运算精确地捕捉这些局部特征。这种自动学习特征的方式相较于传统方法更加高效和准确，能够挖掘出频谱图中隐藏的与情感相关的复杂模式。接着，通过池化操作进一步筛选和压缩这些特征，保留最具代表性的信息，同时去除冗余信息和噪声。以提取到的重要情感特征为依据，实现对学习者情感的识别。在这一过程中，深度学习网络的参数通过训练过程来学习获取，目的是最小化损失函数以提高情感分类的准确性，进而完成学习者情感识别技术算法的优化。

深度学习网络由卷积层、池化层和输出层组成，在优化后的学习者情感识别技术算法中，将加重处理后的学习者语音频谱图像作为输入向量，输入到深度学习网络的卷积层中。在卷积层对输入的学习者语音频谱图像进行局部特征的提取^[11]。具体来说，卷积层通过预设的卷积核，在学习者语音频谱图像上滑动并进行点积运算，从而精确地捕捉到如频谱峰值、谷值等关键的局部特征。这些特征反映了学习者语音中的情感状态，是情感识别的重要线索。这一过程可以用公式表示：

$$v(t) = \sum_{k=1}^L f x^*(T-k) \quad (5)$$

式中： $v(t)$ 表示通过卷积层提取的语音频谱图像的局部特征； L 表示卷积层卷积核长度； k 小时输入学习者语音频谱图像维度； f 表示激活函数； T 表示卷积过程的序列索引。

通过卷积操作，深度学习卷积层中的每个卷积核都学习到不同的频谱图像局部特征，这些特征随后被作为池化层的输入。在池化层中，对提取到的学习者语音频谱图像局部特征进行降维处理，以减少计算量并提取最具代表性的特征。通过最大池化操作，保留每个池化区域内的关键信息，同时去除冗余和噪声。这一过程用公式表示：

$$p(t) = \max[v(t) + s] \quad (6)$$

式中： $p(t)$ 表示池化层降维处理后的频谱图像的关键特征； s 表示池化核的索引。经过卷积层和池化层的处理后，学习者语音频谱图像的特征被有效地提取和降维。接下来，全连接层将这些特征映射到情感类别空间，实现情感分类。全连接层通过计算每个情感类别的概率，并应用激活函数进行归一化处理，从而得到最终的情感分类结果。这一过程可以用公式表示：

$$O_j = f[Wp(t) + b] \quad (7)$$

式中： O_j 表示提取的学习者语言频谱图像特征属于第 j 类情感的概率； W 表示深度学习网络全连接层与池化层的连接权重系数； b 表示偏置项。

全连接层的参数（权重和偏置）是通过深度学习网络的训练过程学习得到的。在训练过程中，深度学习网络根据标注的情感标签，不断调整参数，以最小化损失函数（如交叉熵损失），从而提高学习者情感分类的准确性。最终，将识别样本划分到概率最高的情感类别中，以此实现学习者情感识别。

4 实验论证

4.1 实验数据及环境

为了验证优化后的基于深度学习的学习者情感识别技术算法的可靠性与可行性，开展了一项对比实验，将本文算法与现行的两种识别算法进行对比：文献[1]中提出的基于改进K-均值聚类的识别技术算法，以及文献[2]中提出的多模态数据融合的识别技术算法。实验数据来源于某学校学习者语音数据库，该数据集中包含了200名学习者的语音序列，总大小为2.62 GB。

实验环境为：操作系统为Windows KM1.6，处理器为i8-UHNG62，内存容量为32 GB。

4.2 实验流程

在实验过程中，首先利用本文提出的算法对数据集中的200名学习者进行情感识别。为了评估算法的性能，将数据集中的40%数据作为深度学习网络的训练样本，剩余60%作为测试样本。深度学习网络的参数设置为：学习率设置为0.01；输入图像的维度设置为2维，大小调整为512 px × 512 px；全连接层与池化层之间的连接权重初始化为2.62；偏置项设置为0.63。

在预处理阶段，利用式（1）（2）（3）将学习者语音信号转化为频谱图像，并利用式（4）对学习者的频谱图像进行加重处理以改善频谱图的平滑性和均衡性，增强特征。然后，利用式（5）（6）（7）对测试样本进行情感识别。为了全面评估算法的性能，采用了加权准确率和未加权准确率两个指标。加权准确率通过对每个类别的样本量赋予相应的权重来调整准确率，减少了样本数量不平衡对整体算法评估结果的干扰。而未加权准确率则能够平衡算法的整体性能，在计算准确率时，无论某个类别的样本数量多少，其在最终评估结果中的贡献都是相同的。加权准确率和未加权准确率越高，说明算法识别精度越高。

4.3 实验结果与讨论

根据实验中记录的3种算法对学习者的情感识别结果，绘制了图1和图2，分别展示了3种算法的加权准确率和未加权准确率结果。

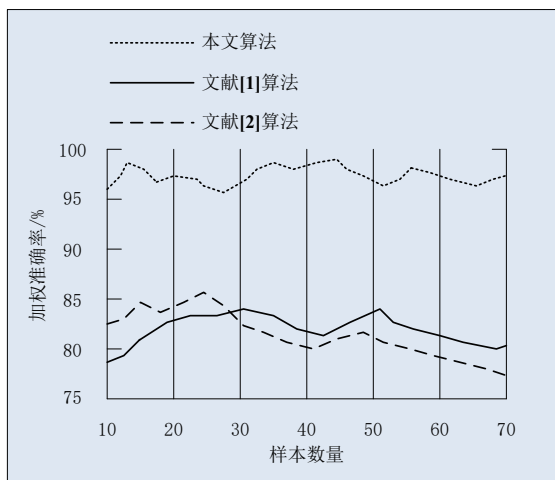


图1 学习者情感识别技术算法加权准确率

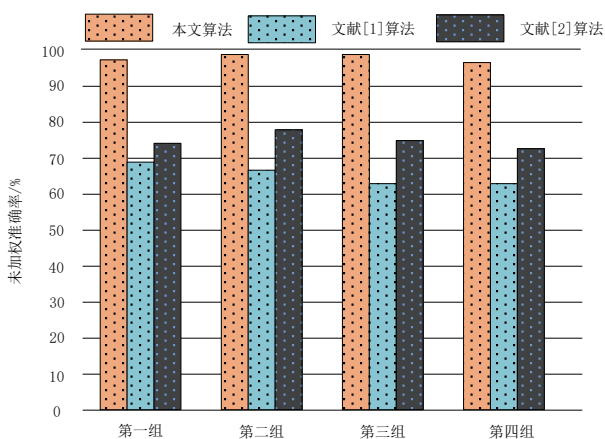


图2 学习者情感识别技术算法未加权准确率

从图1、图2可以看出，在学习者情感识别的场景中，本文提出的优化算法在加权和未加权准确率上均实现了显著的提升。两个准确率指标均稳定地保持在90%以上，这一表现远优于其他算法。因此，实验结果有力证明了本文提出的优化算法在学习者情感识别方面更具优势，所提出的优化方案不仅可行，而且可靠。进一步剖析本文方法的优势能够发现，运用深度学习网络进行多层卷积和池化操作，可高效捕捉频谱图像中的局部特征，例如频谱峰值、谷值等，这些特征对情感识别极为关键。此外，深度学习网络还能自动学习并优化参数，以实现损失函数最小化，进而提升情感分类的准确性，使得算法能够更准确地识别学习者的情感状态。

5 结语

本文提出一种基于深度学习的学习者情感识别技术算法的优化方法，通过集成K-均值聚类与深度学习模型，增强了从复杂频谱图像中提取关键情感特征的能力，显著提高了情感识别的精确度和稳定性。此优化不仅为在线教育领域带来了更加智能、个性化的学习辅助工具，还通过理论分析与

实验验证，深化了对情感识别技术挑战与未来发展趋势的理解。展望未来，随着人工智能技术的持续革新，学习者情感识别技术将迎来更为广阔的发展机遇。进一步深化算法研究，致力于开发更加高效、精确的情感识别解决方案，并强化跨学科合作，推动该技术在在线教育、智能辅导等多个领域的广泛应用，为构建智能化的教育生态系统贡献力量。

参考文献:

- [1] 李巧君, 郭彧. 基于改进K均值聚类的语音情感识别深度学习方法[J]. 计算机应用与软件, 2024, 41 (9): 224-229.
- [2] 司俊勇, 付永华. 多模态数据融合的在线学习情感计算研究[J]. 图书与情报, 2024(3): 69-80.
- [3] 李永伟, 陶建华, 李凯. 基于声门波和声道特征的语音情感识别[J]. 信号处理, 2023, 39(4):632-638.
- [4] 张石清, 陈晨, 赵小明. 采用双阶段多示例学习网络的语音情感识别[J]. 计算机科学与探索, 2024,18(12): 3300-3310.
- [5] 王思丽, 杨恒, 刘巍. 基于混合机器学习的网络舆论情感识别方法研究[J]. 情报工程, 2024, 10 (2): 11-26.
- [6] 黄佩瑶, 程慧慧, 唐小煜. 具有互补特征学习框架和注意力特征融合模块的语音情感识别模型[J]. 新疆大学学报(自然科学版)(中英文), 2024, 41 (1): 52-58.
- [7] 李磊, 朱永同, 杨琦, 等. 基于多任务学习与注意力机制的多层次音频特征情感识别研究[J]. 智能计算机与应用, 2024, 14 (1): 85-94.
- [8] 李丹丹, 陈俊. 基于集成算法的在线课程评论情感识别与主题挖掘研究[J]. 北京印刷学院学报, 2023, 31 (12): 59-68.
- [9] 苗博瑞, 许云峰, 赵少杰, 等. C-BGA: 结合对比学习的多模态语音情感识别网络[J]. 计算机工程与应用, 2024, 60 (16): 168-176.
- [10] 宫岐伟, 禹可, 吴晓非. 基于ERC Roberta的提示学习实现对话情感识别[J]. 北京邮电大学学报, 2023, 46 (5): 106-111.
- [11] 杨卓东, 金卫, 生慧, 等. 基于深度学习的脑电信号情感识别研究进展[J]. 北京生物医学工程, 2023, 42 (3): 315-321.

【作者简介】

曾光辉(1972—), 男, 湖南长沙人, 本科, 副教授, 研究方向: 智能信息处理。

(收稿日期: 2024-11-08)