

基于改进的 MFCC 与 CNN 心音信号识别方法的研究

王佳佳, 熊飞龙

(江西理工大学 能源与机械工程学院, 南昌 330000)

摘要: 心音分类在心血管疾病的早期检测中起着至关重要的作用, 特别是对小型初级卫生保健诊所、缺少专业人员陪护的家庭等检测; 为提高心音信号数据类别间的可辨别性, 提出了一种改进 MFCC 方法提取数据特征, 并与 PCA 算法组合, 作为样本输入 CNN 模型进行分类; 对心音信号数据集进行降噪与下采样, 减少数据量及噪声影响, 利用改进的 MFCC 对其进行特征提取, 并利用 PCA 算法进而抽取相关特征; 为验证不同特征数据集以及不同滤波算法在提取心音数据特征数据集方面对分类过程及分类结果所产生的影响, 将其分别输入 CNN 模型进行训练; 经实验验证, 改进的 MFCC 特征+PCA 算法与传统的 MFCC 相比较, 可提高训练模型的训练速度, 同时也可提高识别率。

关键词: 心音分类; MFCC; 滤波; PCA; CNN

Research on Recognition Method of Heart Sound Signals Based on Improved MFCC and CNN

WANG Jiajia, XIONG Feilong

(School of Energy and Mechanical Engineering, Jiangxi University of Science and Technology, Nanchang 330000, China)

Abstract: Heart sound classification plays a crucial role in the early detection of cardiovascular disease, especially in small primary health care clinics and families lacking professional care. In order to improve the distinguishability between the data categories of heart sound signals, an improved Mel-Frequency Ciptal Coefficients (MFCC) method is proposed to extract data features and combine with Principal Component Analysis (PCA) algorithm, with a sample input Convolution Neural Network (CNN) model for classification. The heart sound signal data set is denoised and downsampled to reduce the amount of data and the influence of noise, their features are extracted by using the improved MFCC, and then, the relevant features by using the PCA algorithm. In order to verify the effects of different feature data sets and filtering algorithms on the classification process and results in the heart sound data feature dataset extraction, they are input into the CNN model for training respectively. Experimental results show that compared with the traditional MFCC, the improved MFCC feature + PCA algorithm can improve the training speed and recognition rate of the training model.

Keywords: heart sound classification; MFCC; filter; PCA; CNN

0 引言

论文旨在利用深度学习算法通过对心音信号数据进行处理来判断病人心脏健康与否。目前心音听诊方面的电子设备逐渐智能化, 不再局限于简单听诊器, 很多电子听诊设备流入市场, 对人体心脏检测起到了很大的便利性。收集听诊信号并结合信号分类算法对信号计算、分析与处理输出分类结果是医学诊断和治疗中非常重要的一步。这种方法可以帮助医生更准确地诊断患者的病情, 并制定相应的治疗方案, 其中信号的降噪效果、分类的准确率以及分类算法的复杂度是信号诊断的关键, 也是人工智能领域研

究学者一直以来比较关注的问题^[1-6]。

人类心脏的两侧各有由心脏瓣膜连接的两个腔室, 分为左心室和右心房。心脏跳动形成心脏周期。心脏周期指从一次搏动开始到下一次搏动开始的心脏所有活动时间, 其分为两部分: 一部分因收缩产生的时间为收缩期, 因心脏舒张产生的时间则为舒张期。正常心音信号又分为第一心音信号 S1、第二种心音 S2, 而第三心音信号 S3 与第四心音 S4 成人一般甚少听到, 部分信号会因为心脏异常而产生^[7], 因此部分研究学者会根据心音产生的 S3、S4 信号来区分信号的分类。

收稿日期: 2023-10-26; 修回日期: 2023-12-07。

基金项目: 江西省教育厅项目(GJJ21084)。

作者简介: 王佳佳(1994-), 女, 硕士, 讲师。

通讯作者: 熊飞龙(1982-), 男, 大学本科, 工程师。

引用格式: 王佳佳, 熊飞龙. 基于改进的 MFCC 与 CNN 心音信号识别方法的研究[J]. 计算机测量与控制, 2024, 32(12): 201-207, 215.

尽管近年来在心音信号处理分类方面取得了很大进展,但大多数心音分类方法都是基于传统的分段特征和基于浅结构的分类器^[8-9]。这些传统的声学表征和分类方法在表征心音方面存在不足,且由于心声环境复杂多变,性能普遍下降。Morteza 等人^[10]从未分割的心音信号中提取 40 个时频特征,进行心音异常检测。Zeinali 等人^[7]将数据集的统计特征、信号特征、小波变换、信息理论(香农熵)作为心音信号的特征,输入 GBC、RFC、SVC 模型进行训练结果对比,实验结果表明在二分类时,GBC 模型分类准确率可达到 95%;Deng 等人^[8]采用改进的 MFCC 算法对无需分段的心音信号进行处理,将结果作为输入,传入 CRNN 模型,并对其他模型参数进行对比研究,取得了良好的成效;Schmidt 等人^[11]提取了不同类型的光谱特征,包括光谱参数模型、瞬时频率和幅值(IFA)和倍频功率来表征时频属性;Jamal 等人^[12]采用突变参考信息的包络方法和峰值检测算法分割心音信号,对其参数识别,达到心音信号分类的目的。

随着人工智能算法的快速发展,深度学习神经网络(DNN)近年来也被用于人类心音分类的探索。Chen 等人^[13]采用短时傅里叶变换(STFT)幅度谱作为 CNN 的输入样本。通过改变卷积层数达到优化模型的复杂度的目的。结果表明,具有两层卷积层的 CNN 取得了最好的效果。Tschannen 等人^[14]提出了一种结合了基于 CNN 的特征提取器和支持向量机(SVM)心音分类方法,并对训练结果进行了评估。Latif 等人^[15]提出了一种基于 RNN 的异常心跳检测算法。

心音分类问题主要取决于信号处理算法的选择、调整。通常情况下,通过信号处理软件例如:Matlab、Python 等编程软件,从音频中获取信号,一般信号采集频率多为 44 100 Hz,数据量比较大,可达几十万个样本点。本研究对获取的原始信号数据进行预处理,在此过程中,声音中的噪音首先通过基于频率和振幅的特殊滤波器,从而消除部分频率噪声的影响,可提高音质。第二方面也是重要的目标,即提取特征。针对此,本文提出了一种将基于改进的 MFCC 特征提取方法,并与 PCA 算法结合,作为样本输入卷积神经网络 CNN 模型进行分类。并对比了 3 个经典模型:VGG、FC、LeNet-5 对特征数据的训练结果,实验数据,得出不同层数对特征数据训练的计算复杂度是不相同的。论文在不分割心音信号的情况下计算传统的 MFCC 倒谱系数,为考虑信号的动态特征以及更好地模拟人耳感知频率的特性,对传统的 MFCC 倒谱系数特征进行改进,提取信号动态特征以及频率的特性,并将上述特征进行拼接,通过 PCA(主成分分析法)算法,形成特征数据集,输入分类模型。论文所设计的深度学习框架可用于解决处理维度减少,单通道数据集较小的问题。实验表明,在 2016 PhysioNet/CinC Challenge 数据库上,对于两类分类问题(病理或非病理),有效提高了分类准确率,并提高了速率。

这项研究与以往文献不同之处一方面在于可训练包含周期较少的心音信号,以往对于包含周期较少的心音信号

数据需要进行扩充或者摒弃,而在本研究中,仍然可保存此数据,减少了数据集的丢失,并可分析包含周期较少的心音信号,另外对于采集频率较高的数据,例如原始数据,即使数据样本点比较多,通过特征提取方法所得结果的数据维度仍然可保持一致,避免数据集数据维度差距较大,而需填充放大等操作,便于后期输入神经网络模型;第二方面降低了特征数据维度后,通过改变深度学习网络结构,减少了深度学习网络参数,加快了训练时间,并提升了算法训练的准确度。

1 改进的 MFCC 特征提取算法

本节从心音信号预处理、基于改进 MFCC 的信号特征提取+PCA、深度学习分类模型 3 个主要部分详细描述算法。预处理阶段:读取心音信号数据,对信号进行预处理,去除低频伪影、基线漂移和干扰。特征提取阶段:运用特征提取算法提取病理性和正常心音信号改进的 MFCC 特征,并将改进 MFCC 特征通过 PCA 算法,设置相关参数,提取出相关性较强的特征形成新的数据集。将数据集按 4:1 的比例分为:训练集与测试集,输入到 CNN 模型中,作为样本进行分类器学习和训练。

1.1 数据集

本研究中所运用的数据集为:“PASCAL 挑战数据”部分数据。PhysicsNet/Computing in Cardiology Challenge 2016^[6]提供了迄今为止最大的开源心音数据集,其中包含 3 000 多条心音数据,持续时间从 2~120 多秒不等,所包含的完整心动周期在 10~40 个之间。并标注了每个样本的“label”,正常心音信号的心动周期 0.8~1.2 s 之间。本文使用的数据集总共 1 613 条,所占内存为 50.4 kB 有余,“label”包含 murmur、normal、artifact、extrahls、extrastole 共 5 种类型。论文从 wav 数据集中读取数据,将 1 613 条“label”进行整理,分为正常与异常二类,用 1 和 -1 标记,其中正常心音信号数量为 769 条,异常心音信号数量为 844 条^[17]。

1.2 数据预处理

对心音信号预处理主要包括两个部分:下采样与滤波。原始心音信号的采样率 44 100 Hz,相较于信号本身的频率而言,采样点较为庞大,因为心音信号的频率范围在 10~1 000 Hz,两者相比相差一个数量级。原始数据由于采样频率较大,出现大量冗余的现象,并且这些大部分数据是其他无关的信号,甚至为干扰信号,因此增加了噪声的干扰性^[18],造成数据特征偏差较大,而且巨大的数据量也增加了对信号的处理的时间复杂度和空间复杂度,影响处理数据的速度。因此,有必要对原始心音信号进行下采样。

由于数据集中的心音通常是在有大量患者和医务人员的嘈杂环境中记录的,心音专家对这些心音进行了预处理,以便在实验中使用它们之前获得更高质量的数据。需要注意的是,心脏在人类肺附近工作,自然伴随着肺或人类呼吸的声音,这导致了高频声音的出现。因此下采样处理后信号需要初步降噪,本文采用参考文献中提到的五阶巴

特沃斯带通滤波器 (通频带: 25~400 Hz)^[19]。以便从原始读取的心音信号中去除低频伪影、基线漂移和高频干扰。图 1 为心音信号的原始信号和预处理信号的时域图。

下采样部分设置的频率为 1 000 Hz^[18], 采样效果如图 1 所示。图 1 (a) (b) 图从上往下依次为异常信号、正常信号的原始信号时域图、下采样频率为 1 000 Hz 的时域图与滤波之后的时域图。将下采样信号与原始数据信号相比较, 可以看见在缩小了数量级之后的采样点, 其波形本身不会产生较大的差距, 对波形整体结构影响比较小, 周期保持不变, 优势在于在数据量方面, 却近乎减少至原先的一半。对下采样的数据进行滤波处理后, 可得降噪之后的心音信号其心音周期分布更加明显, 可减少高频段信号噪声的影响, 有助于增强频段内的信号。

1.3 数据特征提取

1.3.1 MFCCs

梅尔频率倒谱系数 (MFCCs) 近年来主要应用于语音识别领域。它是由 Davis、Mermelstein 在 20 世纪 80 年代引入, 是迄今为止应用仍比较广泛的算法^[19-21]。语音信号通过预加重、分帧、加窗、快速傅里叶变换与功率谱、Mel 滤波与取对数、离散余弦变换 (DCT)、均值归一化 (Mean Normalization) 得到 Mel 频率倒谱系数 (MFCCs)^[8]。其中 Mel 滤波原理为:

$$Mel(f) = 2595 \lg(1 + f/700) \quad (1)$$

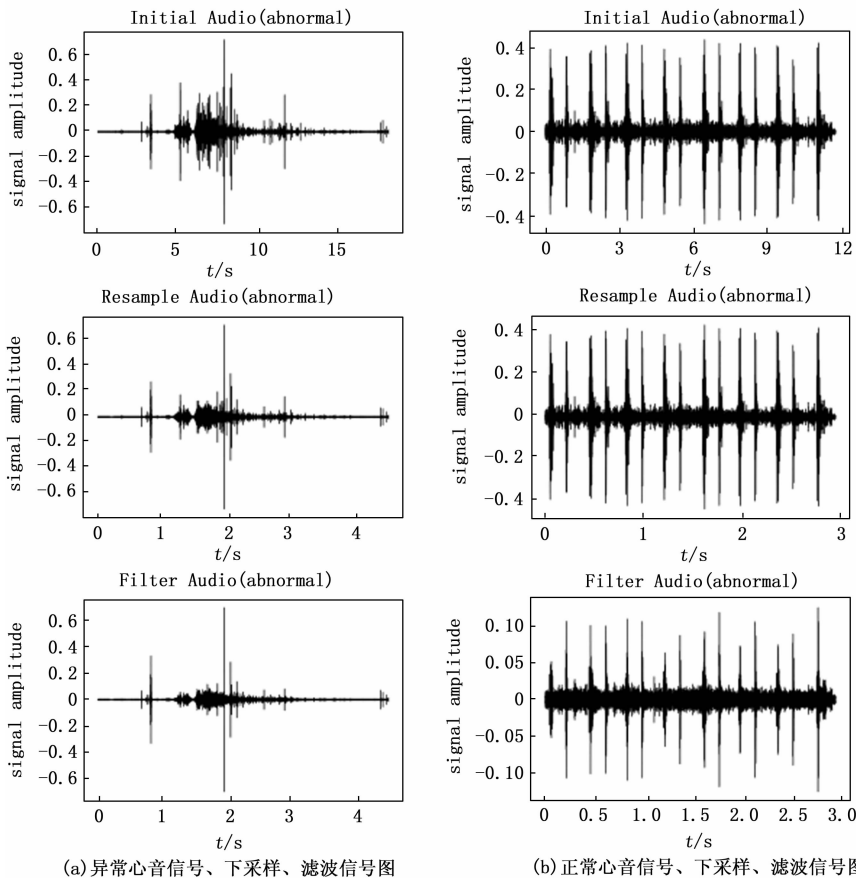


图 1 异常、正常心音信号预处理信号过程图

$$f_m = 700(10^{m/2595} - 1) \quad (2)$$

根据式 (1) (2), 可实现实际频率尺度 f 、Mel 频率尺度之间的转换。在 Mel 频率轴上配置 L 个通道的三角形滤波器组, 构建 Mel 刻度上的过滤器组为 $H_m(k)$, 如式 (3):

$$H_m(k) = \begin{cases} 0, k < f(m-1) \\ \frac{2(k-f(m-1))}{(f(m+1)-f(m-1))(f(m)-f(m-1))}, & f(m-1) \leq k \leq f(m) \\ \frac{2(f(m+1)-k)}{(f(m+1)-f(m-1))(f(m)-f(m-1))}, & f(m) \leq k \leq f(m+1) \\ 0, k \geq f(m+1) \end{cases} \quad (3)$$

式中, $k = \frac{(1+N)f_m}{f_s}$, N 为 FFT 点数, f_s 为采样频率, f_m

为梅尔刻度转化为频率后的值, m 代表第 m 个滤波器, k 为横坐标, $f(m)$ 为中心频率, $m = 1, 2, \dots, L$ 。论文设置 $m = 26$, 窗长尺寸比例为 0.025, 窗移尺寸比例为 0.01, 以单个信号的长度作为度量, 乘以相应的比例, 得出实际窗长与窗移动。由于传统的 MFCC 方法在计算倒谱系数时, 只考虑了只使用了每一帧的静态特征, 因此为考虑信号的动态特征, 对 MFCC 提取其一阶差分与二阶差分的动态特征; 采用频率加权 MFCC, 使用了梅尔频率刻度来替代线性频率刻度, 以更好地模拟人耳感知频率的特性, 提取出 MFCC 特征系数。因此通过此方法, 可得出 52 个信号特征数据。为更好地考虑了音频信号的能量分布, 可以提高特征的鲁棒性和区分能力, 并对 MFCC 采用功率加权的方法。它对于处理音频信号中的能量特性更为敏感, 能够更好地捕捉语音的关键信息。并对上述特征进行融合。原始信号的 MFCC 频谱与滤波之后的 MFCC 频谱分别如图 2 和图 3 所示, 正常与异常信号的改进 MFCC 特征频谱如图 4~6 所示。

将图 2、图 3 对比可得, 未进行预处理的信号所提取的 MFCC 特征在频率上强度较小, 频率特征不够明显, 而进行预处理之后的信号频率分布强度相对较大, 数据轻度增大。滤波后的频谱图比滤波前的频谱图更加清晰, 高频噪声被有效地抑制。因此未预处理的数据由于包含噪音且数据量较大, 在算法特征提取方面, 所提取的数据对整体影响比较大。

图 4~6 分别为原始数据 (44 100 Hz)、下采样与 SG 滤波处理的数据、下采样与带通滤波处理的混合 MFCC 特征频谱图。对比可得, 3 种不同处理方式的频谱图所计算出的特征分布在频段上强度存在明显

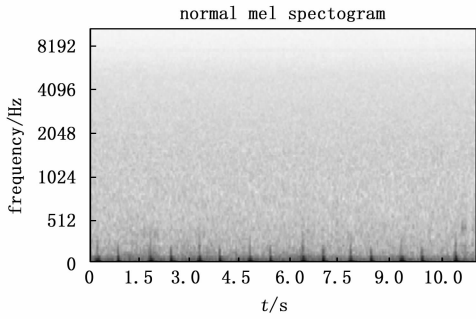


图 2 原始信号的 MFCC 频谱图

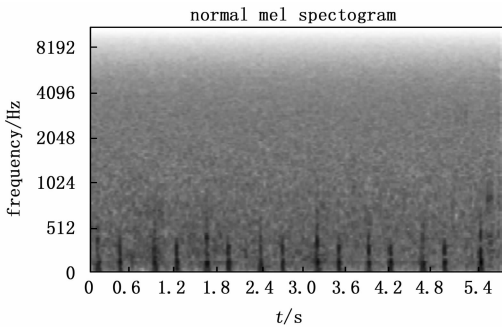
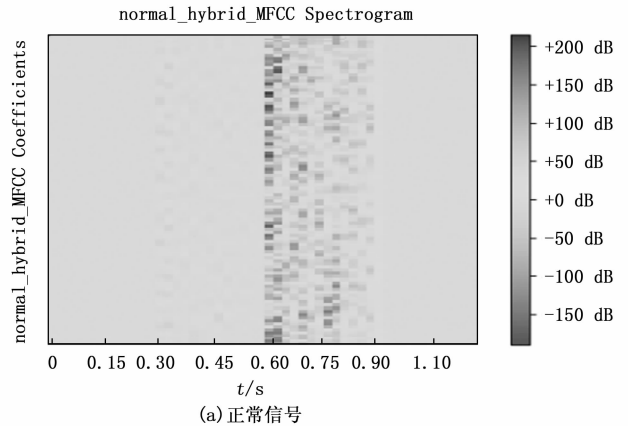
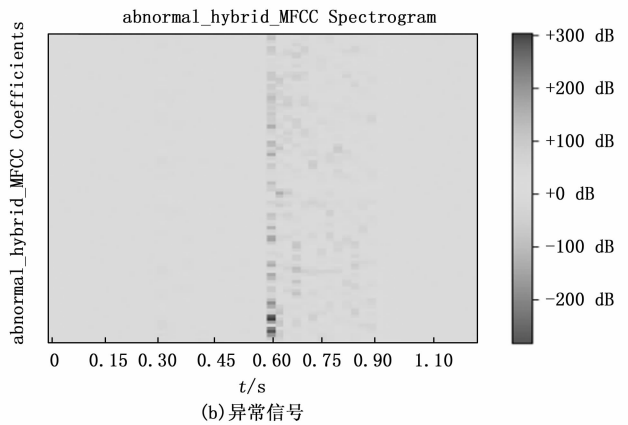


图 3 带通滤波之后的 MFCC 频谱图

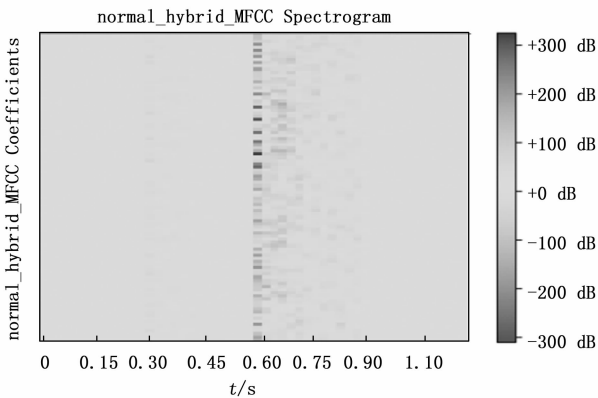


(a) 正常信号

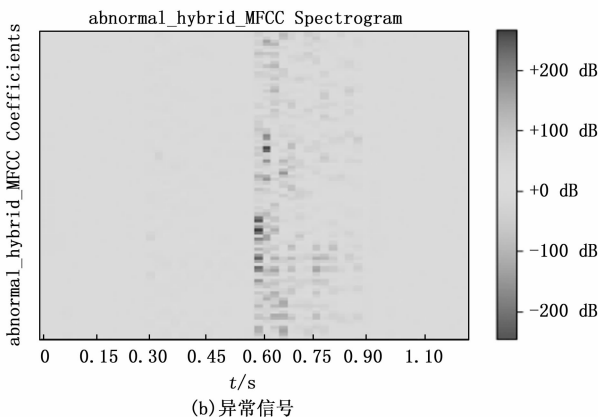


(b) 异常信号

图 5 下采样与 SG 滤波处理的信号改进 MFCCs 特征频谱图



(a) 正常信号



(b) 异常信号

图 4 未预处理（下采样与滤波）的信号改进 MFCCs 特征频谱图

的差异，更能反映帧与帧之间的强度关系。通过改进的 MFCC 算法，不仅可以提取信号特征，另外可有效解决：即使输入算法的数据维度不一致，周期不一致，输入改进的 MFCC 算法之后，可使得输出数据维度保持一致，减少数据的维度调整的复杂度。同样也可得出，不同的预处理方式下采用同一算法，正常信号与异常信号的特征在频域分布方面强度也存在很大的差异。图 6 与图 3 进行比较，可得出传统的 MFCC 所提取的频域特征信息少于混合特征所表现的信息，并且图 3 所包含的特征仅仅是每帧之间的特征，提取不到帧与帧之间的数据关系，而图 6 不仅可以体现每帧之间的特征，还可以体现对帧与帧之间的信息分布以及关联强度。为验证原始数据特征与预处理数据特征之间的相关性，判断预处理之后的数据特征是否保留了原始数据特征，本文采用皮尔逊相关系数来衡量两者之间的关系。

皮尔逊相关系数（Pearson Correlation Coefficient）是一种衡量两个变量之间线性相关程度的统计量，通常表示为 r 。它的取值范围在 $-1 \sim 1$ 之间，其中 1 表示完全正向线性相关， -1 表示完全负向线性相关，0 表示无线性相关。它是一种重要的数学统计量，常被用于衡量两个变量之间的线性关系强度和方向。如果两个变量同时向相同方向偏离均值，相关系数将为正值；如果向相反方向偏离，则相关系数为负值。它是由卡尔·皮尔逊从弗朗西斯·高尔顿

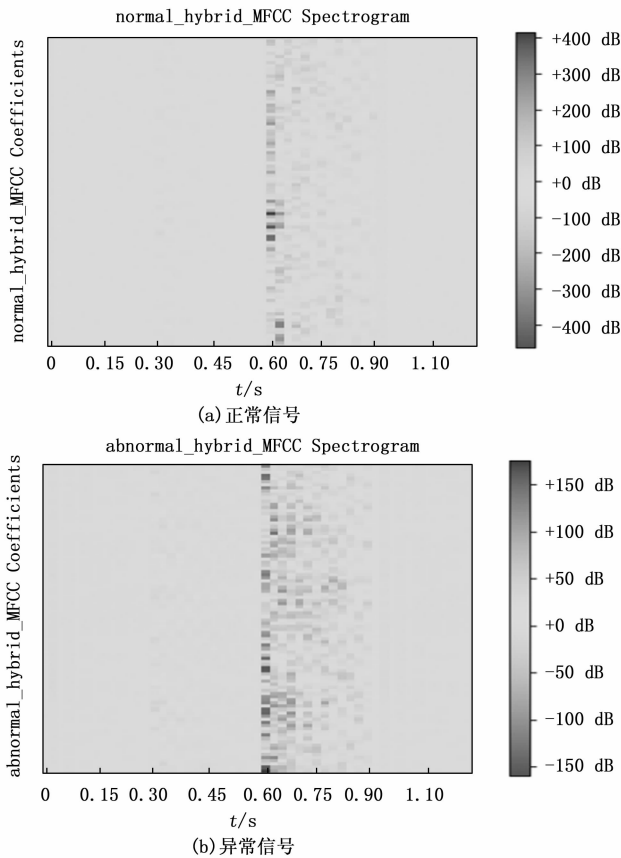


图 6 下采样与带通滤波处理的信号改进 MFCCs 特征频谱图

在 19 世纪 80 年代提出的一个相似却又稍有不同的想法演变而来的^[22]。表 1 为皮尔逊相关系数判断变量的相关强度的属性关系。

表 1 相关强度属性

范围	等级
0.8~1.0	极强相关
0.6~0.8	强相关
0.4~0.6	中等程度相关
0.2~0.4	弱相关
0.0~0.2	极弱相关或无相关

将预处理(带通降噪、下采样)所提取的改进的 MFCC 特征与原始数据的改进 MFCC 特征运用皮尔逊相关系数计算得出, 其相关系数的数值在 0.8~1 之间, 属于强相关, 因此预处理方法之后的数据仍然保留原始信号数据的特征, 减少了数据量, 并降低了数据的维度, 算法具有一定的有效性。

1.3.2 PCA 降维

当输入变量过多时, 深度学习算法复杂度增加, 性能下降。在特征空间中拥有大量的维度意味着一个巨大的空间体积。反过来, 空间中的点(数据行)通常代表一个小的和不具有代表性的样本, 这个会极大影响机器学习算法性能的问题, 被称为“维度诅咒”^[7]。另外通过 1.3.1 改进的 MFCC 特征之后, 数据维度有所增加, 可通过降维算法

增强数据的特征性。通常可采用主成分分析(PCA)和线性判别分析(LDA), 降低数据维数。但是 LDA 作为一种有监督的学习算法, 也属于一种分类学习算法。其降维过程需要输入标签数据, 受标签数据影响较大, 降维的维度及主成分的数量为标签类别减去 1, 不适用于后期的神经网络模型, 因此本文采用 PCA 算法实现降维。

PCA 是一种简单有效的线性变换方法。PCA 分析的主要目的是通过确定变量的相关性来识别数据中的模式。如果存在很强的相关性, 那么减少维度将是重要的。一般来说, PCA 在高维数据中找到方差的最大方向, 并在维数较少的子维中绘制, 以保留大部分信息。非常适用于数据量较大的降维算法。具体降维步骤如下:

1) 计算样本均值:

$$\mathbf{X}(m, n) = \mathbf{X}_0(m, n) - \frac{1}{n} \sum \mathbf{X}_0(m, n) \quad (4)$$

$\mathbf{X}_0(m, n)$ 代表 m 行 n 列的原始特征数据矩阵, $\mathbf{X}(m, n)$ 为原始特征数据去中心化处理后的矩阵。

2) 求矩阵 $\mathbf{X}(m, n)$ 的协方差矩阵 $\mathbf{X}_c(n, n)$, 即:

$$\mathbf{X}_c(n, n) = \frac{1}{n} \mathbf{X}^T \mathbf{X} \quad (5)$$

3) 用特征值分解方法求协方差矩阵 $\mathbf{X}_c(n, n)$ 的特征值与特征向量;

4) 对特征值从大到小排序, 选择其中最大的 k 个。然后将其对应的 k 个特征向量分别作为行向量组成特征向量矩阵 $\mathbf{P}(n, k)$, 本文将 k 设置为 30。

5) 将数据转换到 k 个特征向量构建的新空间中, 即:

$$A(m, k) = \mathbf{X}(m, n) \mathbf{P}(n, k) \quad (6)$$

为验证特征数据对信号分类的影响, 降低数据维度, 采用 PCA 提取改进 MFCC 的特征值, 选取有效成分为 30, 形成降维的二维数据特征集。

2 心音信号识别方法

研究的最后部分是对数据集进行验证分类。将数据特征作为样本, 使用 CNN 进行训练和测试有两个显著的优点: 1) CNN 可以使大多数数据既提取全局的、特定的特征, 又剔除无用的信息; 2) CNN 作为深度学习领域中比较简单的网络, 使用它作为模型进行训练和测试, 这样在保持准确率的同时, 可以提高诊断速度; 3) 本文提出的方法在准确性和诊断速度方面都有较好的效果。因此, CNN 这一经典且相对简单的深度学习网络可以实现高精度的分类任务。

在 CNN 开始训练一个模型和测试之前, 预处理的数据被转换成二维数据。连续三层通过叠加二维卷积层和最大池化层, 提取 CNN 的全局特征, 减少参数, 保留主要特征。然后, 在网络的末端采用 3 个完全连通层, 最终对信号特征数据进行分类。

论文设计的 CNN 块使用 3 个卷积层, 第一层 32 核, 第二层 32 核, 第三层 16 核。每一层的可学习核大小设置为 2×2 , 在每一卷积层中使用流行的 ReLU 激活函数。在第一、二层卷积层之后, 使用一个最大池(Max-pooling), 其

中使用 1×1 窗口, 步幅为 1×1 , 第三层卷积层之后使用最大池化层, 使用 2×2 窗口, 步长为 1×1 。由于数据为单通道数据, 因此不需要进行归一化 (BN) 层和 dropout 层。得到的特征数据通过 256 个神经元的全连接 (FC, full connected) 层学习全局特征, 再次通过 128 个神经元全连接。最后, 采用 softmax 层推导出正常心音和异常心音对应两类的概率分布。

3 实验结果与分析

3.1 测试评价

对于所提出的方法的评价, 每个实验使用 4 个测量值: 精度、召回率、精度和 F_1 。这些测量的定义如下:

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (7)$$

$$Recall = TP / (TP + FN) \quad (8)$$

$$Specificity = TN / (TN + FP) \quad (9)$$

$$Precision = TP / (TP + FP) \quad (10)$$

$$F = \frac{2 * (Precision * Sensitivity)}{Precision + Sensitivity} \quad (11)$$

式中, TP 表示模型预测为 “normal” 类的 “normal” 标签个数, TN 表示被模型预测为 “abnormal” 类的 “abnormal” 标签个数, FP 表示被模型预测为 “normal” 类的 “abnormal” 标签个数, FN 表示被模型预测为 “abnormal” 类的 “normal” 标签个数。Accuracy (准确率) 表征的是预测正确的样本比例, Precision (查准率) 表征的是预测正确的正样本的准确度。Precision 越大说明误检地越少, Precision 越小说明误检地越多。Recall (查全率) 表征的是预测正确的正样本的覆盖率, Recall 越大说明漏检地越少, Recall 越小说明漏检地越多。specificity (特异度) 表征的是预测正确的负样本的覆盖率。

3.2 测试结果

大多数心音信号分类方法通过 3 种途径实现较高的准确率。1) 使用原始数据提取特征, 并通过有监督于无监督的机器学习方法进行分类; 2) 将原始数据转换为图像, 并应用不同的卷积网络; 3) 在卷积神经网络基础上, 应用改进的模型, 将原始数据转换为二维图像。然而, 上述方法是针对神经网络的卷积层提取信号特征, 进行模型训练。为了考虑噪声数据的负面影响, 本文利用滤波器对原始数据校正, 并归一化, 提出了一种改进 MFCC 特征提取方法, 此方法可有效挖掘信号特征, 输入简单的 CNN 模型, 可在短时间内有效地实现准确分类心音的目的。

数据集测试硬件信息: 本文使用 Intel (R) Core (TM) i3-10100CPU@3.60 GHz 处理器、32.0 GB 内存和机带 RAM 8 GB 的 Windows 系统上进行测试, Python 被用作编译语言。本文数据集总样本数为 1 613 条, 其中正常心音信号为 769 条, 异常心音信号为 844 条。将总样本分为训练集与测试集, 其中训练集 1 290 条, 测试集 323 条。

为避免过拟合, 将改进的 MFCC 特征+PCA 特征数据进行随机打乱, 并按照 4:1 的比例将其划分为训练集与测试集, 输入 CNN 训练深度学习模型进行测试。在训练和测

试环节中, 批次大小为 100, 学习率为 0.000 1, 一次迭代训练为 13 次, 重复实验以确保结果的稳定性。结果表明, 该方法的精度和损耗在很小的范围内波动。训练集与测试集损失值与准确率过程如图 7 所示。

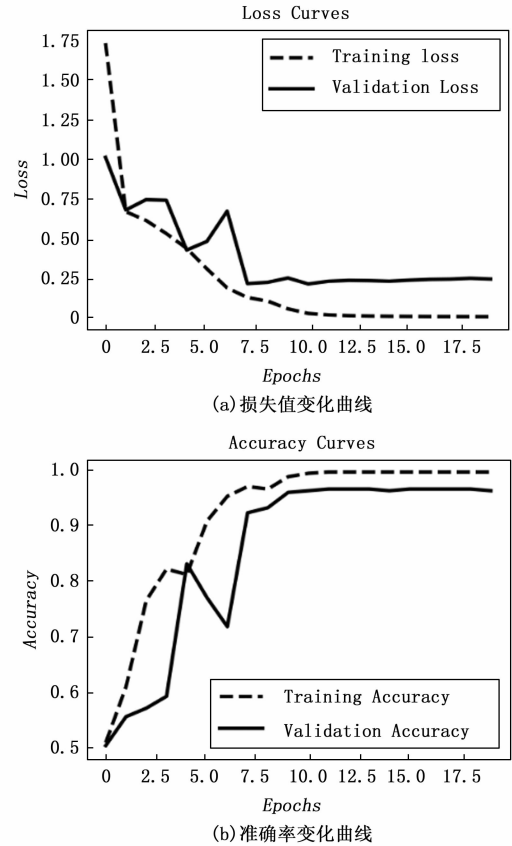


图 7 训练过程曲线

图 7 (a) 中, 训练集与验证集数据在迭代次数为 10 时, 训练集数据 Loss 值逐渐接近于 0, 测试集数据 Loss 值逐渐趋于平稳。(b) 图训练集准确率达到 98.59%, 测试集准确率为 96.59%。模型混淆矩阵结果为: TP 、 TN 、 FP 、 FN 分别为: 154、158、4、7, TN 数量 (154) 占总数 (323) 的百分率为: 47.68%, FN 数量 (7) 占总数 (323) 的百分率为: 2.17%, FP 数量 (4) 占总数 (323) 的百分率为: 1.24%, TP 数量 (158) 占总数 (323) 的百分率为: 48.92%, TN 数量 (158) 占 “abnormal” 数据总数 (165) 的百分率为: 95.76%, FN 数量 (7) 占 “abnormal” 数据总数 (165) 的百分率为: 4.24%, TP 数量 (154) 占 “normal” 数据总数 (158) 的百分率为: 97.47%, FP 数量 (4) 占 “normal” 数据总数 (158) 的百分率为: 2.53%, 由此可看出, 此模型对正常数据的预测准确率高于异常数据的预测准确率。

3.3 测试对比

为验证不同模型对数据特征分类的影响, 将改进的 MFCC+PCA 数据特征作为输入, 众所周知几个经典的模型结构: VGG、FC、LeNet-5 与所设计的 CNN 模型结构进行对比。表 2 为几个模型参数之间的数据对比。

表 2 不同模型的训练速率与准确率对比

算法	数值
VGG/s	93.58%/5 280
FC/s	91.56%/161
LeNet-5/s	91.74%/189
CNN/s	96.59%/74

表 2 中所涉及的时间仅指模型训练时间。由表 2 可看出, 当训练模型卷积层数与训练参数较多时, 由于实验模型所涉及的特征数据维度较小并且为单通道的数据, 对于输入数据组并不具优势, 反而会因为时间与空间计算的复杂度, 延长训练时间。当训练模型越复杂, 另外还可能会造成数据拟合的现象, 进而使训练集准确率较高, 测试集准确率较低的现象。层数较少的训练模型会降低算法训练时间及复杂度, 减少训练参数, 减少对 CPU 及 GPU 性能的依赖。因此, 通过对比使用各种网络的结果, 验证了所设计的简单卷积神经网络能够取得很好的性能。为探究预处理之后的数据特征与未处理数据特征输入 CNN 的模型的训练速率, 对带通滤波 + MFCC + PCA、SG 滤波 + 改进 MFCC、带通滤波 + 改进 MFCC、带通滤波 + 改进 MFCC + PCA、与未预处理原始数据特征进行了对比, 如表 3 所示。

表 3 不同特征对比

算法	CNN/s	准确率/%
MFCC+PCA	162	94.21
未预处理 + 改进 MFCC	128	93.50
改进 MFCC(带通)	126	94.74
改进 MFCC(SG)	125	95.36
改进 MFCC+PCA(带通)	74	96.59

上述所涉及时间也主要指模型的训练时间, 模型的批次大小为 100, 学习率为 0.000 1, 选用 Adam 优化器, 迭代次数保持一致, 为保证数据收敛, 将训练模型的迭代次数均设置为 20。由表 3 可得, 改进的 MFCC 算法所提取的特征数据集在训练时间与训练速度上高于 MFCC 特征数据。SG 滤波处理之后的改进 MFCC 特征数据的训练速率与准确率略高于带通滤波处理之后的改进 MFCC 特征数据的训练速率与准确率。当采用 PCA 算法进行进一步筛选之后, 发现改进 MFCC 与 PCA 的混合算法在训练速率与准确率方面更具有优势。

4 结束语

本文提出了一种数据驱动的心音分类方法, 在数据预处理中加入特征提取法(改进的 MFCC 特征提取算法, 并与 PCA 算法进行组合), 并使用 CNN 作为网络模型, 降低了计算复杂度。与未预处理数据相比较表明, 使用改进 MFCC+PCA 进行数据特征提取不仅可以达到预期的效果。而且还可以提取输入数据的全局特征, 同时通过降维有效地减少了计算量, 提高了训练速率, 使特征数据更具有代表性。为了验证本文方法的有效性和可靠性, 利用“PAS-

CAL 挑战数据部分数据集”进行了实验。实验结果表明, 本文提出的方法具有较高的诊断精度和较快的诊断速度。

参考文献:

- [1] RUBIN, ABREU J, GANGULI R, et al. Recognizing abnormal heart sounds using deep learning [J]. ArXiv, 2017: 8.
- [2] XIAO B, XUY Q, BI X L, et al. Heart sounds classification using a novel 1-D convolutional neural network with extremely low parameter consumption [J]. Neurocomputing, 2020, 392: 153-159.
- [3] SURATKAR S, BHIUNGADE S. Deep-fake video detection approaches using convolutional-recurrent neural networks [J]. 控制与决策学报(英文), 2023, (2): 198-214.
- [4] LI T, YIN Y B, MA K N. Lightweight end-to-end neural network model for automatic heart sound classification [J]. Information, 2021, 12 (2): 54.
- [5] ALKHODARI M, FRAIWAN L. Convolutional and recurrent neural networks for the detection of valvular heart diseases in phonocardiogram recordings [J]. Computer Methods and Programs in Biomedicine, 2021, 200: 105940.
- [6] CHEN Y C, WEI S S, ZHANG Y T. Classification of heart sounds based on the combination of the modified frequency wavelet transform and convolutional neural network [J]. Medical & Biological Engineering & Computing, 2020, 58 (9): 2039-2047.
- [7] ZEINALI Y, TAGHI S, NIAKI A. Heart sound classification using signal processing and machine learning algorithms [J]. Machine Learning with Applications, 2022, 7: 100206.
- [8] DENG M Q, MENG T T. Heart sound classification based on improved MFCC features and convolutional recurrent neural networks [J]. Neural Networks: the Official Journal of the International Neural Network Society, 2020, 130: 22-32.
- [9] EMMANUEL, BABATUNDE S. A review of signal processing techniques for heart sound analysis in clinical diagnosis [J]. Journal of Medical Engineering & Technology, 2012, 36 (6): 303-307.
- [10] MORTEZA Z, RAD, BAHRAMI A, et al. Heart sound anomaly and quality detection using ensemble of neural networks without segmentation [C] //2016 Computing in Cardiology Conference (CINC), 2016, 43: 613-616.
- [11] SCHMIDT, SAMUEL E. Acoustic features for the identification of coronary artery disease [J]. IEEE Transactions on Biomedical Engineering, 2015, 62 (11): 2611-2619.
- [12] JAMAL N, IBRAHIM N. Automated heart sound signal segmentation and identification using abrupt changes and peak finding detection [J]. Procedia Computer Science, 2021, 179: 260-267.
- [13] CHEN Q Y, ZHANG W B. Automatic heart and lung sounds classification using convolutional neural networks [C] //Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, 2016: 1-4.

(下转第 215 页)