

基于双向长短时记忆网络和自注意力机制的心音分类

卢官明^{1,2}, 李齐健^{1,2}, 卢峻禾³, 戚继荣⁴, 赵宇航⁴, 王 洋⁴, 魏金生^{1,2}

(1. 南京邮电大学通信与信息工程学院, 南京 210003; 2. 智能信息处理与通信技术省高校重点实验室(南京邮电大学), 南京 210003; 3. 南京邮电大学计算机学院, 南京 210023; 4. 南京医科大学附属儿童医院心胸外科, 南京 210008)

摘 要: 心音听诊是早期筛查心脏病的有效诊断方法。为了提高异常心音检测性能, 提出了一种基于双向长短时记忆(Bi-directional long short-term memory, Bi-LSTM)网络和自注意力机制(Self-attention mechanism, SA)的心音分类算法。对心音信号进行分帧处理, 提取每帧心音信号的梅尔频率倒谱系数(Mel-frequency cepstral coefficients, MFCC)特征; 将MFCC特征序列输入Bi-LSTM网络, 利用Bi-LSTM网络提取心音信号的时域上下文特征; 通过自注意力机制动态调整Bi-LSTM网络各时间步输出特征的权重, 得到有利于分类的更具鉴别性的心音特征; 通过Softmax分类器实现正常/异常心音的分类。在PhysioNet/CinC Challenge 2016心音数据集上对所提出的算法使用10折交叉验证法进行了评估, 得到0.942 5的灵敏度、0.943 7的特异度、0.836 7的精度、0.886 5的 F_1 得分和0.943 4的准确率, 优于对比的典型算法。实验结果表明, 该算法在无需进行心音分段的基础上就能有效实现异常心音检测, 具有潜在的临床应用前景。

关键词: 心音分类; 梅尔频率倒谱系数; 双向长短时记忆网络; 自注意力机制

中图分类号: TP391.4

文献标志码: A

Heart Sound Classification Using Bi-LSTM and Self-attention Mechanism

LU Guanming^{1,2}, LI Qijian^{1,2}, LU Junhe³, QI Jirong⁴, ZHAO Yuhang⁴, WANG Yang⁴, WEI Jinsheng^{1,2}

(1. School of Communications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China; 2. Jiangsu Key Laboratory of Intelligent Information Processing and Communication Technology, Nanjing University of Posts and Telecommunications, Nanjing 210003, China; 3. School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China; 4. Department of Cardiothoracic Surgery, Children's Hospital of Nanjing Medical University, Nanjing 210008, China)

Abstract: Heart sound auscultation is an effective diagnostic method for early screening of heart disease. In order to improve the performance of abnormal heart sound detection, this paper proposes a heart sound classification algorithm based on bi-directional long short-term memory (Bi-LSTM) network and self-attention mechanism (SA). Firstly, the heart sound signal is partitioned into frames, and the Mel-

基金项目: 国家自然科学基金(72074038); 江苏省卫生健康委员会重点项目(K2023036); 南京市留学人员科技创新项目(NJKC-ZYZZ2023-04); 南京邮电大学引进人才科研启动基金(NY223030)。

收稿日期: 2024-03-30; **修订日期:** 2024-04-29

frequency cepstral coefficients (MFCC) features are extracted from each frame of the heart sound signal. Next, the MFCC feature sequence is input into the Bi-LSTM network to extract the temporal contextual features of the heart sound signals. Then, the weights of the features output from the Bi-LSTM network at each time step are dynamically adjusted through self-attention mechanism, and more discriminative heart sound features that are conducive to classification are obtained. Finally, the Softmax classifier is used to classify normal/abnormal heart sounds. The proposed algorithm is evaluated using 10-fold cross-validation on the heart sound dataset provided by PhysioNet/CinC Challenge 2016, and achieves sensitivity of 0.942 5, specificity of 0.943 7, accuracy of 0.836 7, F_1 score of 0.886 5, and accuracy of 0.943 4, respectively, which are superior to typical comparative algorithms. Experimental results show that the proposed algorithm can effectively detect abnormal heart sounds without the need for heart sound segmentation, and has potential clinical application prospects.

Key words: heart sound classification; Mel-frequency cepstral coefficients (MFCC); bi-directional long short-term memory (Bi-LSTM); self-attention mechanism

引 言

目前,心血管疾病在我国城乡居民疾病死亡构成比中占据首位^[1]。心音是由心脏肌群收缩、瓣膜开合和血液流动碰撞心脏壁引发的机械性振动而产生的声音^[2],能够反映心房、心室、瓣膜和动脉血管壁等结构的工作状况。在心血管疾病的早期筛查中,心音听诊是最为经济有效的辅助诊断方法^[3]。然而,心音听诊结果往往依赖于医务人员的听诊经验和诊断水平,受主观因素影响较大。随着数字听诊器的出现和模式识别等技术的发展,使得心音信号的数字化采集、存储、分析和先心病筛查的智能化成为可能^[4-5]。心音自动分类技术可在一定程度上减轻医务人员临床诊断的工作压力,提升心血管疾病的诊断水平。

在现有的大多数心音分类算法中,首先要对心音信号进行心音分段,即确定每个心动周期中第一心音(S1)、心脏收缩期(Systole)、第二心音(S2)和心脏舒张期(Diastole)4个状态的位置区间,为心音特征提取与心音分类提供定位基准。心音分段是对心音信号进行分类的前提和基础,心音分段的精度对心音分类的性能具有至关重要的影响。然而,对于一些受病理性杂音及噪音干扰严重的异常心音来说,由于心动周期的结构并非是有规则的,所以准确地对心音信号进行分段的难度较大。为了避免受心音分段的定位精度不高而造成心音分类性能下降的影响,本文提出一种无需进行心音分段的心音分类算法。

1 相关工作

心音信号是准周期性的,每个心音信号的周期称之为心动周期(Cardiac cycle)。每个心动周期含有S1、收缩期、S2和舒张期4种状态。在心音自动分析技术中,国内外研究人员提出了大量的心音分类方法,主要分为基于传统机器学习的分类方法和基于深度学习的分类方法两大类。

基于传统机器学习的心音分类方法先提取人工设计的心音特征,然后再采用基于支持向量机(Support vector machine, SVM)、朴素贝叶斯或隐马尔可夫模型(Hidden Markov model, HMM)的分类器进行心音分类。Deng等^[6]提出了一种基于自相关特征和扩散映射的心音分类方法,首先利用离散小波变换(Discrete wavelet transform, DWT)从心音信号的每个子带包络中提取自相关特征,然后通过扩散映

射将各个子带的自相关特征融合成一个特征向量,最后将特征向量输入到SVM分类器进行心音分类。Milani等^[7]首先提取心音信号中S1和S2的时域和频域特征,然后使用人工神经网络(Artificial neural network, ANN)进行正常/异常心音分类,从PhysioNet/CinC Challenge 2016数据集中选择150条心音记录进行了实验测试,在使用时域、频域和时频域组合特征时分别取得了90%、83.33%和93.33%的准确率。Guyen等^[8]首先将一个心音信号划分成若干个时长为5 s的心音片段,然后从心音片段中提取短期特征,从整个心音信号提取长期特征,并将短期特征与长期特征组合起来创建一个称为长短期特征的特征库,最后分别使用决策树、SVM、朴素贝叶斯、K近邻算法和集成方法进行正常/异常心音分类,从PhysioNet/CinC Challenge 2016数据集中选择2 435条心音记录进行了实验测试,分别取得了87.8%、91.2%、83.3%、89.9%和92.7%的准确率。以上基于传统机器学习的心音分类性能往往依赖于人工设计的心音特征,然而人工设计有效的特征是一件困难的事情。

基于深度学习的心音分类方法采用深度神经网络自动提取心音特征并完成心音分类。Chen等^[9]提出了一种基于改进的频率切片小波变换(Modified frequency slice wavelet transform, MFSWT)和CNN的心音分类方法。首先采用HMM确定S1、S2、收缩期和舒张期的位置,然后使用MFSWT将心音信号转换为二维时频图像,最后利用CNN进行正常/异常心音分类,在PhysioNet/CinC Challenge 2016数据集上进行了实验,取得了0.94的平均准确率、0.95的灵敏度和0.93的特异度。Zhu^[10]等提出了一种基于残差网络(Residual network, ResNet)和长短时记忆网络(Long short-term memory, LSTM)的心音分类算法。首先采用隐马尔可夫模型将心音信号分为多个心动周期片段,接着分别从心动周期片段中提取滤波器组特征(Filterbank, Fbank)、梅尔频率谱系数(Mel-frequency spectrum coefficients, MFSC)特征、梅尔频率倒谱系数(Mel-frequency cepstral coefficients, MFCC)特征,然后将提取到的特征分别输入到融合ResNet和LSTM网络的分类模型中进行正常/异常心音分类。在PhysioNet/CinC Challenge 2016数据集上的实验结果表明,MFSC特征优于其他特征,取得了84.3%的 F_1 得分、84.4%的准确率、84.3%的灵敏度和85.6%的特异度。Li等^[11]首先采用HMM确定心音信号中每个心动周期中S1、S2、收缩期和舒张期的位置,再以S1为起点截取时长为3 s的心音片段,然后提取心音片段的MFCC和对数梅尔(Log-Mel)谱作为二维梅尔(Mel)尺度特征,输入到CNN进行正常/异常心音分类。在PhysioNet/CinC Challenge 2016数据集上的实验结果表明,以Log-Mel谱为特征的模型实现了89.6%的未加权平均召回率、89.5%的灵敏度和89.7%的特异度。

上述心音分类方法均需要在特征提取前有效地完成心音分段,即定位心动周期的S1、收缩期、S2、舒张期4种状态。然而,心音分段一直是心音自动分析技术中极具挑战性的难题,因此一部分研究人员开始研究无需分段的心音分类方法,从而规避因分段不佳给最终分类结果带来的影响。

Singh等^[12]首先对经过预处理后的时长为5 s的未分段心音信号进行连续小波变换,生成二维小波谱图,然后利用预训练的AlexNet网络进行正常/异常心音分类。在PhysioNet/CinC Challenge 2016数据集上进行了实验,取得了90.00%的灵敏度、90.00%的特异度和90.00%的准确率。Zhang等^[13]提出了一种基于时间准周期特征和LSTM的心音分类方法。首先,使用短时傅里叶变换来提取心音信号的频谱图;然后,根据不同频带的频谱图,通过平均幅度差函数(Average magnitude difference function, AMDF)计算心音信号的时间准周期特征,再使用LSTM获得时间准周期特征内的依赖关系,得到有利于心音分类的鉴别特征;最后,通过神经网络进行正常/异常心音分类。在PhysioNet/CinC Challenge 2016数据集上进行了实验,取得了 $0.942\ 2 \pm 0.024\ 1$ 的灵敏度、 $0.904\ 8 \pm 0.017\ 4$ 的特异度和 $0.923\ 5 \pm$

0.010 9的总体得分。Krishnan等^[14]首先将心音信号划分成6 s的心音片段,然后通过Savitzky-Golay滤波器对心音片段进行滤波后,分别输入到一维卷积神经网络(1D-CNN)和前馈神经网络中进行正常/异常心音分类,选取PhysioNet/CinC Challenge 2016数据集中1 081条心音记录进行实验测试,结果表明具有5个隐层的前馈神经网络模型有较好的性能,取得了0.856 5的总体精度、0.867 3的灵敏度和0.847 5的特异度。Li等^[15]首先对经过预处理后的时长为3 s的未分段心音信号进行短时傅里叶变换,得到频谱图,然后利用CNN进行正常/异常心音分类,在PhysioNet/CinC Challenge 2016数据集上进行了实验,得到了87%的灵敏度、85%的特异度、85%的准确率和86%的总体精度。张俊飞等^[16]提出了一种基于卷积块注意力模块(Convolutional block attention module, CBAM)和ResNet的心音分类方法。首先提取心音信号的MFSC谱图,然后采用融合CBAM的ResNet网络进行正常/异常心音分类,在PhysioNet/CinC Challenge 2016数据集上进行了实验,取得了90.25%的灵敏度、91.63%的特异度和90.90%的准确率。

2 基于Bi-LSTM网络 and 自注意力机制的心音分类

鉴于Bi-LSTM网络对时间序列具有良好的建模能力,提出了一种基于Bi-LSTM网络和自注意力机制的心音分类算法,其原理框图如图1所示。

首先,对心音信号进行分帧处理,提取每帧心音信号的MFCC特征;接着,将MFCC特征序列输入Bi-LSTM网络,利用Bi-LSTM网络提取心音信号的时域上下文特征;然后,在此基础上再通过自注意力机制动态调整Bi-LSTM网络各个时间步输出特征的权重,得到有利于分类的更具鉴别性的心音特征;最后,通过Softmax分类器实现正常/异常心音的分类。

2.1 心音MFCC特征提取

心音信号是一种一维时序信号,尽管通过傅里叶变换将其映射至频域能够呈现信号的频谱分布,然而这种变换会导致时域信息的丧失,无法反映频谱分布随时间的动态变化。为了解决该问题,很多时频分析方法应运而生,其中,短时傅里叶变换(Short-time Fourier transform, STFT)是最经典的时频分析方法^[17]。在处理相对较长的时序信号时,短时傅里叶变换首先对其进行分帧、加窗,然后对每一帧进行快速傅里叶变换。频率的单位是赫兹(Hz),但人耳对赫兹这种标度单位的感知并非线性关系。梅尔(Mel)标度是一种基于人耳感知频率的非线性标度,在Mel标度下,频率的变化更符合人耳感知的相对变化。Mel标度的频率 $\text{Mel}(f)$ 与赫兹标度的频率 f 之间的映射关系为

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

梅尔频率倒谱和倒谱的区别在于,梅尔频率倒谱的频带划分是在梅尔标度上等距划分的,比用于正常的对数倒谱中的线性间隔的频带更能反映人耳的感知能力与声音频率之间的非线性关系。相关实验验证了MFCC能够有效表示心音信号的特征^[18-20]。

心音MFCC特征的提取流程主要包括分帧、加窗、快速傅里叶变换、Mel标度转换、计算对数能量和离散余弦变换等操作,具体步骤如下:

(1)分帧、加窗。首先对每个心音片段进行分帧处理,使得每帧的心音信号近似为平稳信号;紧接着对每帧心音信号进行加窗操作,以避免频谱泄漏现象。但加窗使得每一帧两端的信号变弱,为了解决该问题,帧与帧之间有一定的重叠,相邻两帧的起始位置的时间差叫做帧移。本文选择汉明(Hamming)窗,窗长与帧长相同。设每个心音片段包含的帧数为 T ,分帧、加窗后的第 t 帧信号为 $x_t(n)$,其中 t 表示帧序号, $t = 0, 1, \dots, T-1$; n 表示第 t 帧信号的采样点序号, $n = 0, 1, \dots, N-1$, N 为每帧心音信号

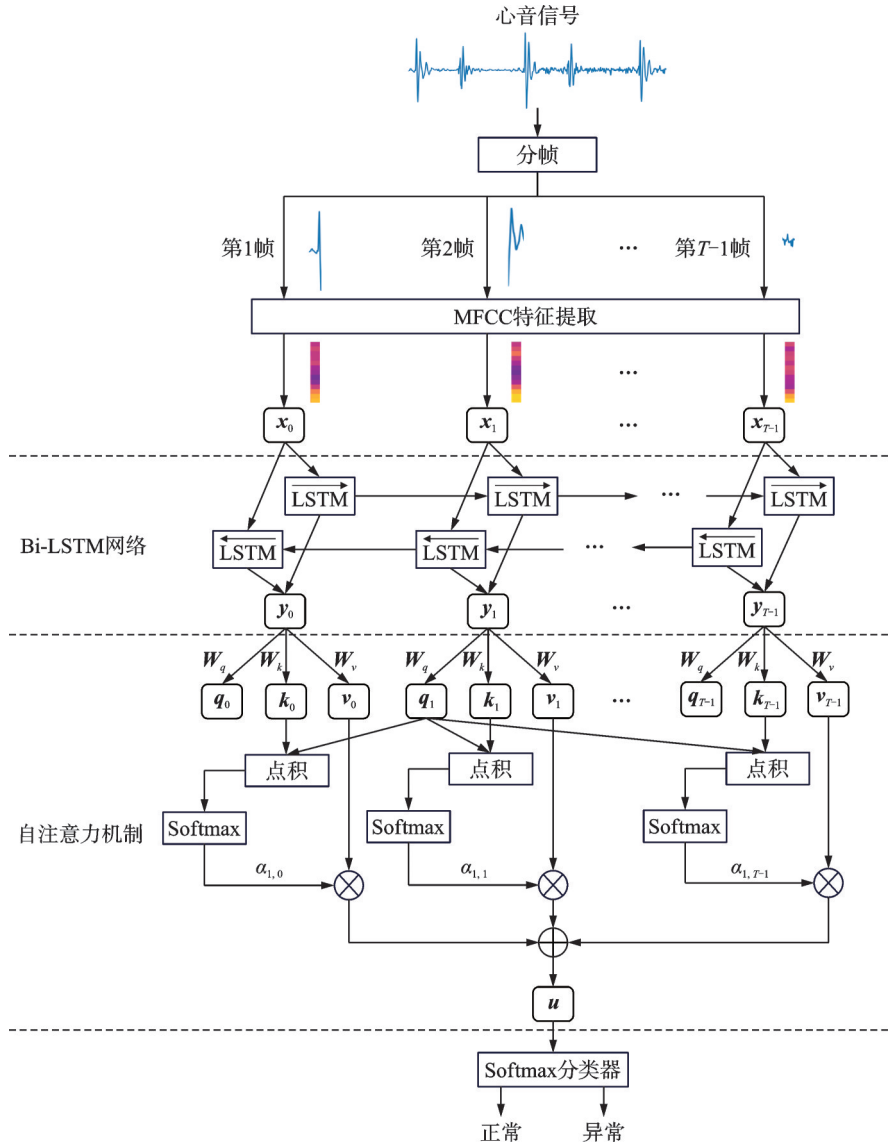


图1 基于Bi-LSTM网络和自注意力机制的心音分类原理框图

Fig.1 Block diagram of heart sound classification based on Bi-LSTM and self-attention mechanism

的采样点总数。

(2)快速傅里叶变换。对加窗操作后的每帧心音信号进行快速傅里叶变换,得到各帧心音信号的频谱,其中第 t 帧信号频谱 $X_t(k)$ 的计算公式为

$$X_t(k) = \sum_{n=0}^{N-1} x_t(n) e^{-j2\pi nk/K} \quad k = 0, 1, \dots, K-1 \quad (2)$$

式中: k 为频域中的频率索引, K 为快速傅里叶变换的点数。在此基础上对频谱取模平方可得到对应的功率谱 $|X_t(k)|^2$,将时域信号转换为频域上的能量分布。

(3)Mel标度转换。用Mel滤波器组对功率谱进行滤波和标度转换处理。Mel滤波器组的频率响应为

$$H_m(q) = \begin{cases} 0 & q < f(m-1) \\ \frac{q-f(m-1)}{f(m)-f(m-1)} & f(m-1) \leq q \leq f(m) \\ \frac{f(m+1)-q}{f(m+1)-f(m)} & f(m) \leq q \leq f(m+1) \\ 0 & q > f(m+1) \end{cases} \quad (3)$$

式中: m 为滤波器序号, $m=1, 2, \dots, M$, M 为 Mel 滤波器组中滤波器的个数, 研究中 M 一般取 13^[21]; $f(m)$ 为第 m 个滤波器的中心频率; q 为频率; $H_m(q)$ 满足等式 $\sum_{m=1}^M H_m(q) = 1$ 。

(4) 计算对数能量。对 Mel 标度转换的结果进行取对数运算, 第 t 帧心音信号的对数能量 $E_t(m)$ 的计算公式为

$$E_t(m) = \ln \left(\sum_{k=0}^{K-1} |X_t(k)|^2 H_m(q) \right) \quad m=1, 2, \dots, M \quad (4)$$

取对数运算可实现对心音频谱动态范围的压缩, 放大低能量下的能量差异。

(5) 离散余弦变换。为了减少滤波器重叠造成的能量相关性, 对第 t 帧心音信号的对数能量 $E_t(m)$ 进行离散余弦变换, 得到第 t 帧信号的 MFCC, 计算公式为

$$C_t(l) = \sum_{m=1}^M E_t(m) \cos(\pi l(m-0.5)/M) \quad l=1, 2, \dots, M \quad (5)$$

式中 l 为第 t 帧信号 MFCC 的阶数索引。

这样, 从每一帧心音信号中提取到一个 M 维的 MFCC 特征向量, 将每一帧的 MFCC 特征向量按顺序排列起来, 得到每个心音片段的 MFCC 特征向量序列 $x_0, x_1, \dots, x_t, \dots, x_{T-1}$, 其中 x_t 表示第 t 帧心音信号的 MFCC 特征向量。

2.2 Bi-LSTM 网络

循环神经网络(Recurrent neural network, RNN)在处理较长输入序列时存在梯度消失或梯度爆炸的问题, 这导致其难以捕捉输入序列中的长期依赖关系。针对这一问题, LSTM 网络通过引入遗忘门、输入门和输出门 3 种门控单元作为内部机制来调节信息流, 从而有效地处理较长输入序列并捕捉长期依赖关系。但 LSTM 网络只能从前向后获取输入序列信息, 不能充分利用输入序列中的下文信息。而 Bi-LSTM 网络由两个结构相同、方向相反的 LSTM 层组成, 即包含前向 LSTM 层和后向 LSTM 层, 如图 2 所示。前向 LSTM 层从前向后依次计算隐藏层的状态, 后向 LSTM 层从后向前依次计算隐藏层的状态。Bi-LSTM 网络可学习到过去和未来的上下文信息, 在时间序列建模方面比 LSTM 网络更具优势。因此, 采用 Bi-LSTM 网络提取心音信号的时域上下文特征。

对于每个心音片段, 将 T 帧心音信号的 MFCC 特征序列 $x_0, x_1, \dots, x_t, \dots, x_{T-1}$ 作为 Bi-LSTM 网络的输入。设第 t 个时间步当前神经元的输入为 x_t , 前向 LSTM 层隐藏状态 $\vec{h}_t$ 和后向 LSTM 层隐藏状态 $\overleftarrow{h}_t$ 分别提取正、反向时域特征, 分别如式(6)和式(7)所示。

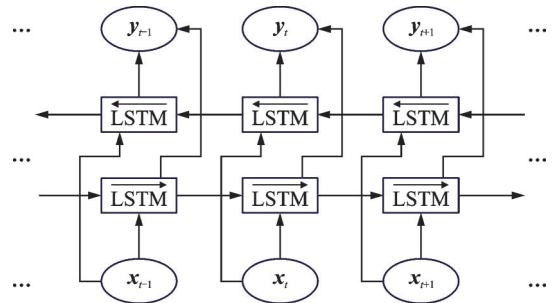


图2 Bi-LSTM网络结构示意图

Fig.2 Bi-LSTM network architecture

$$\vec{h}_t = \overrightarrow{\text{LSTM}}(x_t, \vec{h}_{t-1}) \quad t = 0, 1, \dots, T-1 \quad (6)$$

$$\tilde{h}_t = \overleftarrow{\text{LSTM}}(x_t, \tilde{h}_{t-1}) \quad t = T-1, T-2, \dots, 1, 0 \quad (7)$$

式中: $\vec{h}_t$ 和 $\vec{h}_{t-1}$ 分别为前向 LSTM 层中当前神经元和前一个神经元的隐藏状态向量, $\tilde{h}_t$ 和 $\tilde{h}_{t-1}$ 分别为后向 LSTM 层中当前神经元和前一个神经元的隐藏状态向量。

将正、反两个方向隐藏层提取的特征向量进行拼接, 得到第 t 个时间步的时域上下文特征向量 y_t , 如式(8)所示。

$$y_t = g(\vec{h}_t, \tilde{h}_t) \quad (8)$$

式中: $g(\cdot)$ 表示对特征向量进行拼接的函数。

Bi-LSTM 网络 T 个时间步的时域上下文特征向量组成输出特征序列 $y_0, y_1, \dots, y_t, \dots, y_{T-1}$ 。

2.3 自注意力机制

自注意力机制允许输入序列中的每个元素都与序列中的其他所有元素进行交互, 通过计算每个元素对其他所有元素的注意力权重, 然后将这些权重应用于对应元素的表示, 从而得到一个加权求和的输出表示。自注意力机制不依赖于外部信息或先前的隐藏状态, 完全基于输入序列本身^[22]。自注意力机制通过生成查询、键和值向量, 计算并归一化注意力分数, 最终对值向量进行加权求和, 从而得到输入序列中每个位置的加权表示。

本文通过自注意力机制来动态调整 Bi-LSTM 网络各个时间步输出特征的权重, 得到有利于分类的更具鉴别性的心音特征。将 Bi-LSTM 网络输出的特征序列 $y_0, y_1, \dots, y_t, \dots, y_{T-1}$ 作为自注意力层的输入。自注意力层首先为特征序列中的每个输入向量 y_t 分别生成一个查询向量 q_t 、一个键向量 k_t 和一个值向量 v_t , 其中 $t = 0, 1, \dots, T-1$, y_t, q_t, k_t, v_t 的维度均为 Bi-LSTM 网络隐藏状态维度 d ; 接着计算每个查询向量 q_t 与键向量 k_t 之间的点积, 其中 $t = 0, 1, \dots, T-1$, 得到一个 $T \times T$ 大小的注意力分数矩阵 QK^T , 该矩阵中的每个元素表示每个查询向量 q_t 与键向量 k_t 之间的相关性分数, 其中, $Q = [q_0 \ q_1 \ \dots \ q_{T-2} \ q_{T-1}]^T$, $K = [k_0 \ k_1 \ \dots \ k_{T-2} \ k_{T-1}]$; 然后使用 $\text{Softmax}(\cdot)$ 函数对注意力分数矩阵进行归一化, 在应用 Softmax 之前, 通常会除以一个缩放因子 $\sqrt{d}$ 来稳定梯度; 最后使用归一化后的注意力权重对值向量 v_t 进行加权求和, 加权求和的结果是自注意力机制的输出, 包含了输入序列中所有位置的加权信息。自注意力层最终的输出特征向量 u 为

$$u = \sum_{t=0}^{T-1} \text{Softmax}\left(\frac{QK^T}{\sqrt{d}}\right) \cdot v_t \quad (9)$$

2.4 Softmax 分类器

使用 Softmax 分类器对自注意力层输出的特征向量 u 进行分类, 将心音信号分为正常心音或异常心音两种类别。设 c 为心音信号所属类别的编号, $c = 0$ 表示正常心音, $c = 1$ 表示异常心音, 任意一个心音信号以一定的概率被分类为正常心音或异常心音。

将 Softmax 分类器定义为

$$h(u; \theta) = \begin{bmatrix} p(c=0|u; \theta_0) \\ p(c=1|u; \theta_1) \end{bmatrix} = \frac{1}{\sum_{j=0}^1 e^{\theta_j^T u}} \begin{bmatrix} e^{\theta_0^T u} \\ e^{\theta_1^T u} \end{bmatrix} \quad (10)$$

式中: $h(u; \theta)$ 为一个 2 行 1 列的列向量, 每行的 $p(c=j|u; \theta_j)$ 表示特征向量 u 被分类为第 j 类的概率,

$j=0,1$; θ_j 为 Softmax 分类器待估计的参数,并且构成参数矩阵 θ , θ_j^T 为 θ_j 的转置。

对于 u , 选择 $p(c=j|u; \theta_j)$ 为最大值所对应的 j 作为 Softmax 分类器输出的心音类别编号 y , 即

$$y = \arg \max_j p(c=j|u; \theta_j) \quad (11)$$

在训练分类模型时,将 y 与训练样本标注的类别编号 c 做比较,如果一致,则分类正确,否则分类错误。常用的交叉熵损失函数为

$$L_{CE} = - \sum_{j=0}^1 I(c=j) \log p(c=j|u; \theta_j) \quad (12)$$

式中: $p(c=j|u; \theta_j)$ 表示特征向量 u 被分类为第 j 类的概率, $I(c=j)$ 为示性函数,当 $c=j$ 时, $I(c_m=j)=1$; 当 $c \neq j$ 时, $I(c_m=j)=0$ 。

在正常/异常心音二分类任务中,为了缓解心音数据集样本的类别不平衡问题,即正常心音的样本数目远远多于异常心音的样本数目的不平衡问题,本文采用类加权交叉熵损失函数来训练分类模型。

类加权交叉熵损失函数是对交叉熵损失函数的一种改进,通过为每个类别分配一个权重实现不同类别的加权处理,以调整不同类别的重要性。设 w_0 、 w_1 分别表示正常心音和异常心音的权重系数,类加权交叉熵损失函数的表达式为

$$L_{WCE} = - \sum_{j=0}^1 w_j I(c=j) \log p(c=j|u; \theta_j) \quad (13)$$

在本文中,为了加大异常心音被错误预测为正常心音的代价,设置 $w_0=1$, w_1 等于正常心音的训练样本数与异常心音的训练样本数之比值。

3 实验结果与分析

3.1 心音数据集

为了验证所提出的心音分类算法的有效性,使用 PhysioNet/CinC Challenge 2016 数据集^[23]进行了实验。该数据集包含了从健康受试者和各种心血管疾病患者身上采集的 3 240 份 5~120 s 的心音记录,其中正常心音记录 2 575 条,异常心音记录 665 条。对这些心音样本使用巴特沃斯滤波器降噪处理后进行重采样,采样频率调整为 2 000 Hz。

由于数据集中的每一条心音记录时长不同,本文对每一条心音记录进行切分以截取固定时长的心音片段为研究对象。为了保证切分后的心音片段至少包含一个完整心动周期,将数据集中的每一条心音记录切分为多个时长为 2 s 的心音片段,并丢弃时长不足 2 s 的部分,最终得到 34 490 个心音片段,其中包含 26 410 个正常心音片段,8 080 个异常心音片段。

3.2 实验环境与设置

实验基于 TensorFlow 深度学习框架,实验硬件环境为搭载 Linux 内核的台式机,硬件配置为 NVIDIA GeForce GTX1080TI GPU。实验中,使用类加权交叉熵损失函数和 Adam 优化算法对分类模型进行训练,初始学习率设置为 0.001,衰减率为 $1e-6$,批量大小为 32,训练轮次为 20,丢弃率设置为 0.5。按照 10 折交叉验证的方法对数据集进行划分,将数据集按 9:1 的比例划分为训练集与测试集,每次选择一个不同的测试集,进行 10 次实验。每次实验使用的训练集包含 23 769 个正常心音片段,7 272 个异常心音片段;测试集包含 2 641 个正常心音片段,808 个异常心音片段。对这 10 次实验的灵敏度、特异度、精度、 F_1 得分和准确率取平均作为心音分类模型的性能指标。

3.3 性能指标

使用灵敏度 S_e (Sensitivity)、特异度 S_p (Specificity)、精度 P_+ (Precision)、 F_1 得分、准确率 Acc (Accuracy) 评估模型分类性能, 其计算公式分别为

$$S_e = \frac{TP}{TP + FN} \quad (14)$$

$$S_p = \frac{TN}{TN + FP} \quad (15)$$

$$P_+ = \frac{TP}{TP + FP} \quad (16)$$

$$F_1 = 2 \frac{P_+ \cdot S_e}{P_+ + S_e} \quad (17)$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (18)$$

式中: TP 表示测试样本中标签为异常心音且被正确预测为异常心音的样本数目, 即真阳性; FN 表示测试样本中标签为异常心音但被错误预测为正常心音的样本数目, 即假阴性; FP 表示测试样本中标签为正常心音但被错误预测为异常心音的样本数目, 即假阳性; TN 表示测试样本中标签为正常心音且被正确预测为正常心音的样本数目, 即真阴性。

3.4 实验结果与分析

3.4.1 帧长和帧移的选取

心音信号是一种时变的非平稳信号, 具有短时平稳性, 一般认为在 10~30 ms 的时间内心音信号是平稳的^[24]。在提取心音信号 MFCC 特征时, 需要对心音片段进行分帧处理, 帧长和帧移是两个重要的参数, 需要合理地选择。为了确定合适的帧长, 在实验中, 以 5 ms 的时间间隔在 10~30 ms 范围内调整帧长的取值, 并取帧移为帧长的一半, 测试不同的帧长对灵敏度 S_e 、特异度 S_p 、精度 P_+ 、 F_1 得分和准确率 Acc 的影响, 其结果如图 3 所示。从图 3 可看出, 当帧长为 20 ms 时, 各项性能指标达到最高。

在确定帧长后, 再进一步选择合适的帧移。实验中, 固定帧长为 20 ms, 以 5 ms 的时间间隔在 5~15 ms 范围内调整帧移的取值, 测试不同的帧移对灵敏度 S_e 、特异度 S_p 、精度 P_+ 、 F_1 得分和准确率 Acc 的影响, 其结果如图 4 所示。从图 4 可看出, 当帧移为 10 ms 时, 各项性能指标达到最高。因此, 本文选定帧长为 20 ms, 帧移为 10 ms。

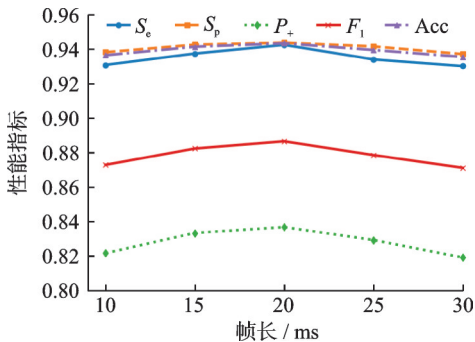


图3 帧长对分类性能的影响

Fig.3 Impact of frame length on classification performance

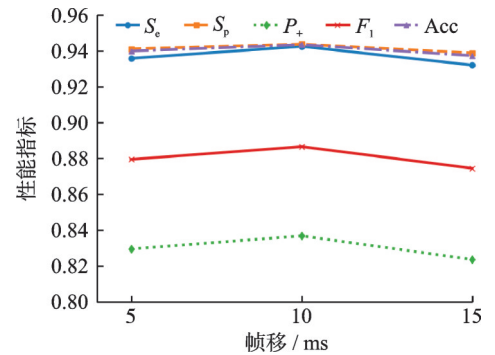


图4 帧移对分类性能的影响

Fig.4 Impact of frame shift on classification performance

3.4.2 使用不同网络模型和损失函数的实验对比

对使用 LSTM、Bi-LSTM 以及引入自注意力机制的 Bi-LSTM+SA 网络模型进行了对比实验,并在模型训练过程中分别使用交叉熵损失函数和类加权交叉熵损失函数,实验结果如表 1 所示。

表 1 使用不同网络模型和损失函数的分类性能对比

Table 1 Comparison of classification performance using different network models and loss functions

损失函数	网络模型	S_e	S_p	P_+	F_1	Acc
交叉熵损失函数	LSTM	0.809 9	0.921 2	0.758 8	0.783 5	0.895 2
	Bi-LSTM	0.852 4	0.932 4	0.794 2	0.822 3	0.913 7
	Bi-LSTM+SA	0.874 3	0.940 6	0.818 4	0.845 4	0.925 1
类加权交叉熵损失函数	LSTM	0.920 8	0.926 4	0.792 9	0.852 1	0.925 1
	Bi-LSTM	0.930 3	0.934 6	0.813 1	0.867 8	0.933 7
	Bi-LSTM+SA	0.942 5	0.943 7	0.836 7	0.886 5	0.943 4

从表 1 可看出,在使用同一种损失函数下,Bi-LSTM 相比 LSTM 网络,各项性能指标都得到了一定程度的提升,引入自注意力机制之后的 Bi-LSTM+SA 网络模型在各项性能指标上达到最高。实验结果表明,Bi-LSTM 比 LSTM 网络更适合对心音 MFCC 特征序列进行建模,Bi-LSTM 网络能够充分利用心音信号的上下文信息,捕获其中的长期依赖关系;融合自注意力机制的 Bi-LSTM 网络在训练时动态地调整各个时间步特征向量的权重,更加关注心音片段中重要帧的特征信息,以便更好地捕捉上下文长期依赖关系,进一步提升了模型的分类性能。

此外,从表 1 实验结果还可看出,相较于交叉熵损失函数,当使用类加权交叉熵损失函数时,无论是 LSTM、Bi-LSTM 还是 Bi-LSTM+SA 网络模型,各项性能指标都最高。图 5 给出了在 10 折交叉验证的 10 次实验中使用不同损失函数时 Bi-LSTM+SA 网络模型的混淆矩阵。从图 5 也可看出,当使用类加权交叉熵损失函数时,模型将 8 080 个标签为“异常心音”的样本正确预测为“异常

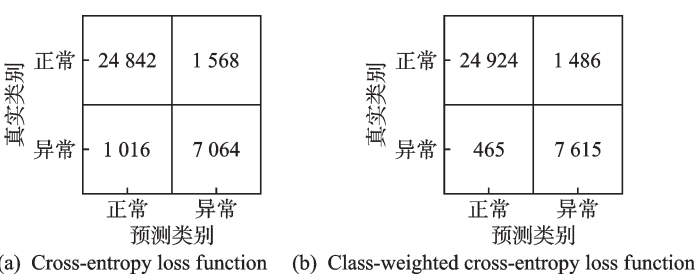


图 5 Bi-LSTM+SA 模型使用不同损失函数时的混淆矩阵

Fig.5 Confusion matrix of Bi-LSTM+SA model with different loss functions

心音”的样本数为 7 615 个,即真阳性的样本数为 7 615 个,多于使用交叉熵损失函数时的 7 064 个;同时,模型将 8 080 个标签为“异常心音”的样本错误预测为“正常心音”的样本数为 465 个,即假阴性的样本数为 465 个,少于使用交叉熵损失函数时的 1 016 个。这表明当使用类加权交叉熵损失函数时,模型更加关注异常心音样本。在临床上假阴性的风险远大于假阳性。假阳性可通过后续检查进一步排除,而假阴性往往会导致患者的病情被漏诊而耽误治疗,因此在心血管疾病的早期筛查中需要尽可能地减少假阴性病例。

通过分析,交叉熵损失函数常用于衡量标签类别和预测类别之间的差异,不关注训练样本的类别分布,因此当正常心音样本数远大于异常心音样本数时,使模型对异常心音的检测能力不足,从而导致更多的漏诊;类加权交叉熵损失函数通过调整正常/异常心音类别的权重,起到平衡正常和异常心音类别的作用,从而提升模型对异常心音检测的性能。

3.4.3 与现有其他心音分类算法的实验对比

为了评估本文心音分类算法的有效性,将本文算法与现有其他心音分类算法进行了对比实验,实验结果如表2所示。文献[7]首先提取心音信号中S1和S2的时域和频域特征,然后使用ANN进行正常/异常心音分类。文献[11]首先对心音信号进行分段,确定心动周期中S1的位置,再以S1为起点截取时长为3 s的心音片段,然后提取心音片段的Log-Mel谱特征,输入到CNN进行正常/异常心音分类。文献[15]首先对经过预处理后的时长为3 s的未分段心音信号进行短时傅里叶变换,得到频谱图,然后利用CNN进行正常/异常心音分类。文献[16]首先提取心音信号的MFSC特征,然后采用融合CBAM的ResNet网络进行正常/异常心音分类。

表2 本文算法与其他算法的分类性能对比

Table 2 Comparison of classification performance between the proposed algorithm and other algorithms

分类算法	特征提取	分类网络	S_e	S_p	P_+	F_1	Acc
文献[7]	时频域组合特征	ANN	0.880 2	0.942 8	0.824 8	0.851 6	0.928 1
文献[11]	Log-Mel谱	CNN	0.893 6	0.895 2	0.723 0	0.723 0	0.894 8
文献[15]	频谱图	CNN	0.872 3	0.851 1	0.640 4	0.738 6	0.856 1
文献[16]	MFSC	ResNet+CBMA	0.894 1	0.912 8	0.758 4	0.820 7	0.908 5
本文算法	MFCC	Bi-LSTM+SA	0.942 5	0.943 7	0.836 7	0.886 5	0.943 4

从表2实验结果可看出,本文算法的灵敏度为0.942 5,特异度为0.943 7,精度为0.836 7, F_1 得分为0.886 5,准确率为0.943 4,所有性能指标均高于对比文献中的心音分类算法。

4 结束语

提出的基于Bi-LSTM网络和自注意力机制的心音分类算法无需进行心音分段,可避免因心音分段的定位精度不高而造成心音分类性能下降的影响。该算法对未经分段的心音信号进行分帧处理,提取每帧心音信号的MFCC特征作为Bi-LSTM网络的输入,利用Bi-LSTM网络提取心音信号的时域上下文特征,通过自注意力机制动态调整Bi-LSTM网络各个时间步输出特征的权重,得到有利于分类的更具鉴别性的心音特征。在PhysioNet/CinC Challenge 2016心音数据集上的实验结果表明,融合自注意力机制的Bi-LSTM网络能够更好地捕捉上下文长期依赖关系,有效提升心音信号的分类性能。此外,针对心音数据集样本的类别不平衡问题,选用类加权交叉熵损失函数训练分类模型,能使模型更加关注异常心音样本,可有效提升异常心音检测的性能。总体来说,本文算法在无需进行心音分段的基础上就能有效实现异常心音检测,在心血管疾病的辅助诊断中具有潜在的临床应用前景。

参考文献:

- [1] 马丽媛,王增武,樊静,等.《中国心血管健康与疾病报告2022》要点解读[J].中国全科医学,2023,26(32):3975-3994.
MA Liyuan, WANG Zengwu, FAN Jing, et al. Interpretation of report on cardiovascular health and diseases in China 2022[J]. Chinese General Practice, 2023, 26(32): 3975-3994.
- [2] SHAHMOHAMMADI M, LUO H X, WESTPHAL P, et al. Hemodynamics-driven mathematical model of first and second heart sound generation[J]. PLoS Computational Biology, 2021, 17(9): 1009361.
- [3] 陈尧. 心音心电信号处理的神经网络方法[D]. 成都: 四川大学, 2021.
CHEN Yao. Research and applications on phonocardiogram and electrocardiogram with neural networks[D]. Chengdu: Sichuan University, 2021.
- [4] 胡晓静,王春宇,张成强,等. 新生儿先天性心脏病筛查技术的现状及智能化发展趋势[J]. 复旦学报(医学版), 2022, 49(1):

107-113.

HU Xiaojing, WANG Chunyu, ZHANG Chengqiang, et al. Current status and intelligent development trend of the screening technology for neonatal congenital heart disease[J]. Fudan University Journal of Medical Sciences, 2022, 49 (1): 107-113.

- [5] 徐玮泽,俞凯,徐佳俊,等. 先天性心脏病心音听诊筛查的人工智能技术应用现状[J]. 浙江大学学报(医学版), 2020, 49 (5): 548-555.

XU Weize, YU Kai, XU Jiajun, et al. Artificial intelligence technology in cardiac auscultation screening for congenital heart disease: Present and future[J]. Journal of Zhejiang University(Medical Sciences), 2020, 49 (5): 548-555.

- [6] DENG S W, HAN J Q. Towards heart sound classification without segmentation via autocorrelation feature and diffusion maps [J]. Future Generation Computer Systems, 2016, 60: 13-21.

- [7] MILANI M G M, ABAS P E, DE S L C, et al. Abnormal heart sound classification using phonocardiography signals[J]. Smart Health, 2021, 21: 100194.

- [8] GUVEN M, UYSAL F. A new method for heart disease detection: Long short-term feature extraction from heart sound data [J]. Sensors, 2023, 23 (13): 5835.

- [9] 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: 2039-2047.

- [10] ZHU L X, QIAN K, WANG Z H, et al. Heart sound classification based on residual shrinkage networks[C]//Proceedings of 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). Glasgow: IEEE, 2022: 4469-4472.

- [11] LI Z S, CHANG Y, SCHULLER B W. CNN-based heart sound classification with an imbalance-compensating weighted loss function[C]//Proceedings of 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). Glasgow: IEEE, 2022: 4934-4937.

- [12] SINGH S A, MAJUMDER S, MISHRA M. Classification of short unsegmented heart sound based on deep learning[C]//Proceedings of 2019 IEEE International Instrumentation and Measurement Technology Conference (I2MTC). [S.l.]: IEEE, 2019: 1-6.

- [13] ZHANG W J, HAN J Q, DENG S W. Abnormal heart sound detection using temporal quasi-periodic features and long short-term memory without segmentation[J]. Biomedical Signal Processing and Control, 2019, 53: 101560.

- [14] KRISHNAN P T, BALASUBRAMANIAN P, UMAPATHY S. Automated heart sound classification system from unsegmented phonocardiogram (PCG) using deep neural network[J]. Physical and Engineering Sciences in Medicine, 2020, 43: 505-515.

- [15] LI T, YIN Y, MA K, et al. Lightweight end-to-end neural network model for automatic heart sound classification[J]. Information, 2021, 12 (2): 54.

- [16] 张俊飞,张贵英. 改进残差神经网络实现心音分类[J]. 计算机工程与科学, 2022, 44 (8): 1514-1520.

ZHANG Junfei, ZHANG Guiying. An improved residual neural network for heart sound classification[J]. Computer Engineering & Science, 2022, 44 (8): 1514-1520.

- [17] DAUBECHIES I, LU J, WU H T. Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool[J]. Applied and Computational Harmonic Analysis, 2011, 30 (2): 243-261.

- [18] RUBIN J, ABREU R, GANGULI A, et al. Classifying heart sound recordings using deep convolutional neural networks and Mel-frequency cepstral coefficients[C]//Proceedings of 2016 Computing in Cardiology Conference (CinC). [S.l.]: IEEE, 2016: 813-816.

- [19] LI F, ZHANG Z, WANG L L, et al. Heart sound classification based on improved Mel-frequency spectral coefficients and deep residual learning[J]. Frontiers in Physiology, 2022, 13: 1084420.

- [20] ARORA V, LEEKHA R, SINGH R, et al. Heart sound classification using machine learning and phonocardiogram[J]. Modern Physics Letters B, 2019, 33 (26): 1950321.

- [21] OLIVEIRA J, NOGUEIRA D, RENNA F, et al. Do we really need a segmentation step in heart sound classification algorithms[C]//Proceedings of 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). [S.l.]: IEEE, 2021: 286-289.

- [22] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of 31st Conference on Neural Information Processing Systems. Long Beach: NIPS, 2017: 1-15.
- [23] LIU C, SPRINGER D, LI Q, et al. An open access database for the evaluation of heart sound algorithms[J]. *Physiological Measurement*, 2016, 37 (12): 2181.
- [24] 许春冬,周静,应冬文,等. 心音信号 MFCC 特征向量提取方法的优化[J]. *信号处理*, 2019, 35(3): 410-418.
XU Chundong, ZHOU Jing, YING Dongwen, et al. Optimization of the extraction method of MFCC feature vectors for heart sound signals[J]. *Journal of Signal Processing*, 2019, 35(3): 410-418.

作者简介:



卢官明(1965-),通信作者,男,博士,教授,博士生导师,研究方向:图像处理与多媒体通信、机器学习、计算机视觉,E-mail: lugm@njupt.edu.cn。



李齐健(1999-),男,硕士研究生,研究方向:音频处理、模式识别、机器学习。



卢峻禾(1996-),男,硕士研究生,研究方向:机器学习、模式识别、数据分析。



戚继荣(1973-),男,主任医师,教授,研究方向:先天性心脏病外科、小儿普胸外科、小儿微创心胸外科。



赵宇航(1998-),男,硕士研究生,研究方向:心胸外科、先天性心脏病、医学教育。



王洋(1998-),男,硕士研究生,研究方向:儿外科学、心胸外科。



魏金生(1995-),男,博士,讲师,研究方向:计算机视觉、情感计算、微表情识别。

(编辑:夏道家)