

基于 HSMM 与 CRNN 的心音分类研究

陈涛, 周升超, 刘晓军

(中南民族大学 生物医学工程学院, 湖北 武汉 430074)

摘要: 心音是可以反映人体心脏及血管瓣膜状态的一种体征信号, 正常与异常心音分类在心脏类疾病的研究中具有重要作用。目前的深度学习心音分类研究大多采用无分割的方法, 本研究通过 PhysioNet 心音数据库获取的心音信号, 进行一系列预处理, 基于 HSMM (Hidden semi-Markov Model) 分割心音结合 CRNN (Convolutional Recurrent Neural Network) 分类, 与未进行心音分割的分类方法对比, 本研究采用的方法在 F1 分数和非异常类的敏感度有所提高, 能够根据心音信号对正常、异常心音分类。

关键词: CRNN; HSMM; 心音分割; 心音分类

中图分类号: R318.04

文献标识码: A

文章编号: 1674-1064 (2022) 03-010-03

DOI: 10.12310/j.issn.1674-1064.2022.03.004

心音信号中蕴含了大量的心脏生理信息, 其中包含心脏各心腔、瓣膜以及血管的运行状态^[1], 是诊断心血管疾病常用的医学信号之一, 可以揭示许多病理性心脏状况, 如心律失常、瓣膜疾病、心力衰竭等。

心音给心脏疾病的评估提供重要线索, 可以作为初步诊断的依据, 在心血管疾病的早期检测中发挥重要作用。通过心音听诊可以分析心脏状态, 只有拥有丰富听诊经验的医生, 才能得出较为准确的诊断结果。据《2020年中国心血管健康与疾病报告概要》数据显示, 我国心血管患病率处于持续上升阶段。据我国疾病死亡人数最新统计报告显示, 我国每年发生心源性猝死的人数约为54.4万。随着心脏疾病患者不断增多, 快速并准确地诊断心脏疾病患者是相关医疗人员面临的重大问题。因此, 通过计算机辅助诊断有助于心脏类疾病患者的早期诊断与治疗。

由于以往的研究在进行心音分类时很少进行分割预处理, 会导致精确度不高、可靠性低等问题。因此, 本研究采用 HSMM 分割心音结合 CRNN 对心音信号分类。

1 心音分割与特征提取

1.1 基于 HSMM 的心音分割

HMM 是一个双重随机过程, 一个随机过程描述的是状态之间的转移, 另一个随机过程描述的是状态与观测值之间的统计对应关系。在 HMM 中不能直接观测到真实状态, 只能通过观测值推断出隐藏状态。隐马尔可夫模型能够很好地描述动态短时平稳信号的平稳性和可变性, 其动态时间序列建模能力很强, 在训练和识别时计算量小。因此, HMM 符合心音序列的统计模型, 能够较好地描述心音序列的短时平稳性及整体的非平稳性^[2]。

一般来说, HMM 可以简化为公式 (1):

$$\lambda = \{\pi, A, B\} \quad (1)$$

其中, π 为初始状态概率分布, A 为状态转移概率矩阵, B 为观察值概率矩阵。隐藏状态序列由 π 和 A 决定, B 决定了观察序列。

在 HMM 中, 模型在一个状态停留的时间 d 的概率为:

$$P_i(d) = a_{ii}^{d-1}(1 - a_{ii}) \quad (2)$$

其中, a_{ii} 为转移概率。由于上述公式是指数分布, 概率 P 随着时间的增长呈指数下降, 这显然不适用于心音分析。

当利用 HMM 分割心音信号时, 给定一段心音信号, 则心音信号或者心音信号的特征值可作为观测序列, 隐藏状态序列则是 $W = \{W_1, W_2, W_3, W_4\}$ 。 W_1 是 S1 (第一心音), W_2 为收缩期, W_3 是 S2 (第二心音), W_4 为舒张期。

假设时刻的状态为 S_t , 观测序列为 $O = \{O_1, O_2, O_3, \dots, O_t\}$, 则状态转移矩阵 $A = \{a_{ij}\}$, a_{ij} 是 W_i 转移到 W_j 的概率:

$$a_{ij} = P(S_{t+1} = W_j | S_t = W_i) \quad (3)$$

心音的状态必定是按照 S1 → 收缩期 → S2 → 舒张期 → S1 转移, 所以, $a_{12} = a_{23} = a_{34} = a_{41} = 1$ 。

观察值概率矩阵 $B = \{\beta_j(O_t)\}$, $\beta_j(O_t)$ 是 j 状态在 t 时刻输出 O_t 的概率, 初始状态概率 π 是模型在初始时刻某个状态出现的概率, 在心音采集上各状态 $\pi_i = 0.25$ ($1 \leq i \leq 4$)。

通常来说, HMM 会使用维特比算法求解心音信号的最优状态序列, $\delta_t(j)$ 是 t 时刻隐藏状态为 j 所有可能的状态转移路径 j_1, j_2, \dots, j_t 中的概率最大值, 通过公式 (4) 可以计算得到 $\delta_t(j)$, 通过公式 (5) 可以得到 $\delta_t(j)$ 的最大值。

$$\delta_t(j) = [\max_{1 \leq i \leq N} \delta_{t-1}(i) a_{ij}] \beta_j(O_t) \quad (4)$$

$$\Psi_t(j) = \arg \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] \quad (5)$$

计算出时刻 T 最大的 $\delta_T(j)$, 即最优隐藏状态序列出现的概率, 时刻 T 最大的 $\Psi_T(j)$ 即时刻 T 最优的隐藏状态。如公式 (8) 所示, 利用局部状态 $\Psi(i)$ 回溯最终得到最优隐藏状态序列 i_1^* 。

基金项目: 中南民族大学中央高校基本科研业务费专项资金资助 (CZY18028); 中南民族大学教学研究项目 (JYX19081)。

通讯作者: 刘晓军, 博士后, 副教授。

$$P^* = \max_{1 \leq i \leq N} \delta_{\tau}(i) \quad (6)$$

$$i_{\tau}^* = \arg \max_{1 \leq i \leq N} [\delta_{\tau}(i)] \quad (7)$$

$$i_t^* = \psi_{t+1}(i_{t+1}^*), \quad t \in [T-1, T-2, \dots, 1] \quad (8)$$

可以由HMM（隐马尔可夫模型）扩展而来的HSMM在Markov链中加入状态驻留概率分布 $p = \{P_i(d)\}$ ， p 表示在持续时间 d 内状态为 i 的概率，可以把预测的状态由某一时间点扩展到一个时间段，有效解决用HMM来分割心音的局限性。因此，可以记HSMM为：

$$\lambda = \{\pi, A, B, P\} \quad (9)$$

则适合HSMM的维特比解码算法需要加入持续时间密度，表达式如公式（10）所示。

$$\delta_i(j) = \max_d [\max_{i'} [\delta_{i'd}(i) a_{ij}] p_j(d) \prod_{s=0}^{d-1} \beta_j(O_{t+s})] \quad (10)$$

1.2 心音分割

MB Malarvili等人提出通过参考心电信号标注^[3]提高分割精度，然而，当心电信号异常的时候^[4]，使得依靠ECG信号标注分割心音的方法不可行。虽然通过人工标注心音的方法，其精度会略低于参考ECG信号的方法，但是其操作简单，标记的误差在容忍的范围之中，所以成为本研究的首选标注方法。

训练分割模型需要大量的心音数据。本研究选取心音数据库中（PhysioNet）的1 000条正常心音数据与1 000条异常心音数据共2 000条心音信号。将原始数据集按照8 : 1 : 1的比例拆分为训练集、验证集和测试集，分别用于训练和评估心音分割的模型。

从训练集中的PCG记录中得到同态包络，希尔伯特包络，功率谱密度包络，离散小波变换包络等特征值，并对训练集PCG记录中的第一心音（S1）和第二心音（S2）位置进行了人工标注，经过训练得到HSMM分割模型。

测试心音分割结果如图1、图2所示。

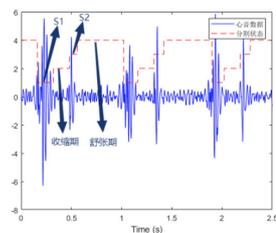


图1 心音分割结果（正常）

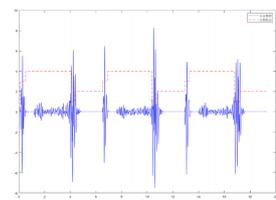


图2 心音分割结果（异常）

1.3 FBank特征提取

随着卷积神经网络（Convolutional neural networks, CNN）^[5-6]逐渐应用到音频领域，研究表明，人耳对声音频率的感知是非线性的，具有对低频音频信号敏感度高、对于高频信号敏感度低的特点，Mel刻度更符合人耳的听觉特性，

将心音信号的频域变换到人耳感知频域中，可以更好地模拟出人耳的效果。

其中，频率与Mel刻度的转换如公式（11）所示：

$$f_{mel} = 2595 * \lg(1 + \frac{f}{700}) \quad (11)$$

FBank是模拟人耳听觉特性提出的参数，随着深度学习的发展，作为二维特征的FBank频谱特征逐渐成为能够利用的音频特征，与梅尔频率倒谱系数（Mel-Frequency Cepstral Coefficients, MFCCs）相比，Fbank没有进行离散余弦变换（discrete cosine transform, DCT）。因为DCT造成音频信号的高度非线性成分丢失严重，所以FBank特征包含更多信息，并且提取的FBank特征更符合声音信号的本质，可以作为心音分类的特征使用。实验已经证明，MFCCs在对高度相关信息不敏感的神经网络中的效果不如FBank。

因此，本研究选用FBank特征替代传统的MFCCs作为卷积神经网络的输入特征。FBank特征的提取流程如图3所示。

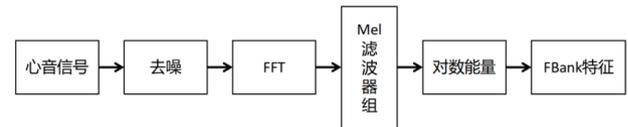


图3 FBank特征的提取流程

2 卷积神经网络架构

2.1 CRNN

CNN是由多个卷积层、池化层以及全连接层组成的前馈神经网络。CNN模型网络在处理图像信息上具有很好的效果。近几年，使用频谱图作为深度学习的输入已经得到了广泛的应用^[7]。

由于心音信号本身是一种时间序列，FBank频谱图含有时域的前后相关的一些特征，其横轴为时间轴，纵轴表示该时间上不同的FBank特征值，CNN中的卷积层可视为对感受野区域的特征提取，采用 $n * 1$ 维的卷积核可对同一时间内的特征进行提取，随后的池化层采用最大池化层，由于心音信号相邻的两层相差较小，采用最大池化层不仅可以减少网络的参数，增加训练速度。

由于使用CNN处理频谱图时不能完全挖掘出与时间相关的信息，在CNN之后加上RNN网络结构可以解决这个问题。RNN模型网络结构适合处理时序相关信息，当前序列会受到之前序列的影响，考虑到了位置信息，因此，可以通过RNN网络结构分析CNN的输出，充分利用FBank频谱图中的信息，完成CRNN分类模型建立。心音分类网络结构图如图4所示。

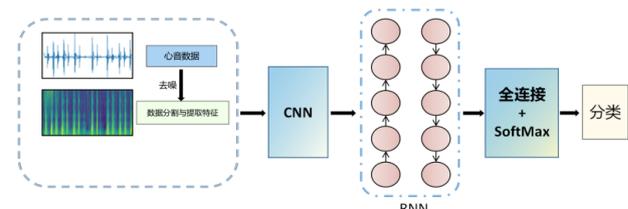


图4 心音分类模型图

2.2 数据预处理

一个完整的心动周期不会超过5 s。因为PhysioNet数据库中的心音长度不一,为了保证以固定维数的数据输入网络,在分类前要切分心音。本研究对已经分割的心音按照5 s间隔划分,并提取FBank特征值。

正常与异常心音的FBank特征如图5、图6所示。

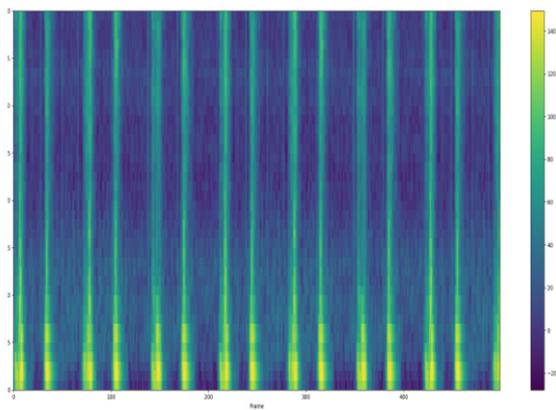


图5 正常心音FBank特征

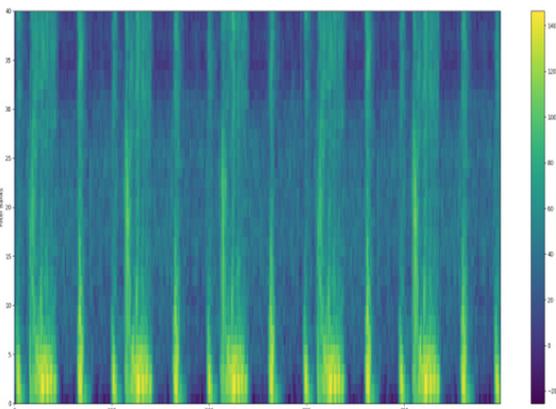


图6 异常心音FBank特征

3 实验设计

3.1 实验评价指标

本研究使用灵敏度 (Sensitivity)、精确率 (Precision) 和F1分数作为评价指标。TP表示正常信号预测为正常的样本数量。TN表示异常信号预测为异常的样本数量。FP表示异常信号预测为正常的样本数量。FN表示正常信号预测为异常的样本数量。敏感度表示所有正常信号被正确分类的比例,其值等同于召回率 (Recall),精确率表示所有被预测为正常信号的样本中,被正确分类为正常信号的比例,F1分数是用来衡量二分类模型精确度的指标。

召回率、精确率和F1分数计算如公式(12)、公式(13)、公式(14)所示。

$$\text{Sensitivity}=\text{Recall}=\frac{\text{TP}}{\text{TP}+\text{FN}}\times 100\% \quad (12)$$

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}}\times 100\% \quad (13)$$

$$\text{F1}=\frac{2*\text{Precision}*\text{Recall}}{\text{Precision}+\text{Recall}}\times 100\% \quad (14)$$

3.2 分类算法对比

对经过分割后数据集按照训练集、测试集4:1的比例随机分配并提取FBank特征,然后使用以下分类方法比较:第一,CNN分类;第二,CRNN分类;第三,基于HSMM心音分割的CRNN分类。分类方法的比较如表1所示。

表1 分类方法性能对比

分类方法	F1分数	敏感度
基于CNN的分类方法	86.4%	89.2%
基于CRNN的分类方法	88.1%	90%
本文的方法	93.8%	96%

4 结果与讨论

心音分类在心脏疾病诊断方向有着重要的作用,是心脏疾病辅助诊断领域的研究热点。针对现有的基于深度学习的心音分类算法很少使用分割预处理,本研究通过构建HSMM心音分割模型,采用具有时频域信息的FBank特征频谱图作为CNN的特征输入,提出了一种结合HSMM分割心音与CRNN模型的心音分类方法。通过与单独使用CNN或CRNN模型分类方法对比,该方法具有更高的F1分数与敏感度,获得了更好的分类效果,也为计算机辅助心音诊断研究提供了一个很有潜力的方向。

参考文献

- [1] 成谢锋,李伟.基于心音窗函数的心音图图形化处理方法的研究[J].物理学报,2015(5):393-403.
- [2] 许春冬,周静,应冬文,等.基于DHMM的低心率变异性心音的分割方法[J].数据采集与处理,2019,34(4):605-614.
- [3] MALARVILI M B,KAMARULAFIZAM I,HUSSAIN S,et al.Heart sound segmentation algorithm based on instantaneous energy of electrocardiogram[C].Computers in Cardiology.Thessaloniki Chalkidiki,2003:327-330.
- [4] 孙树平,吴越,黄婷婷,等.基于STMHT算法的心音分割研究[J].中国医学物理学杂志,2020,37(12):1553-1559.
- [5] Lecun Y ,Boser B ,Denker J , et al. Backpropagation Applied to Handwritten Zip Code Recognition[J].Neural Computation,2014,1(4):541-551.
- [6] 林景栋,吴欣怡,柴毅,等.卷积神经网络结构优化综述[J].自动化学报,2020,46(1):24-37.
- [7] Cummins N,Amiriparian S,Hagerer G,et al.An Image-based Deep Spectrum Feature Representation for the Recognition of Emotional Speech[C].ACM,2017:478-484.