

# 基于堆叠式双向 LSTM 的心电图自动识别算法



王文刀 王润泽 魏鑫磊 漆云亮 马义德

兰州大学信息科学与工程学院 兰州 730000

(897675693@qq.com)

**摘要** 针对日趋增长的心电图数据分析需求,提出了一种新的心电图分类算法。首先对原始数据进行截断固定长度、样本均衡、求取信号的瞬时频率和光谱熵等预处理操作,数据经过预处理后模型能够更好地从其中提取特征进行学习;在训练过程中采用两个双向 LSTM(BILSIM)网络堆叠组成的模型,堆叠式的双向 LSTM(BILSIM)模型是一种改进的循环神经网络模型,相较于卷积神经网络,循环神经网络更加适合用来处理像心电图这样的序列数据。该模型在 Windows 下的 MATLAB2018b 上进行训练和测试,CUDA 版本为 9.0,采用分类准确率作为衡量模型性能的指标在两个数据集上进行了测试,一个是 2017 年生理信号挑战赛的数据(下文简称 2017 数据集),该模型在此数据集上最终分类准确率为 97.4%;另一个是 2018 年生理信号挑战赛的数据(下文简称 2018 数据集),最终分类准确率为 77.6%,并在所属的 MATLAB 组获得了第三名的成绩。该算法与传统 LSTM 网络的结果相比,在 2017 数据集上提升了 5.6% 的准确率,在 2018 数据集上提升了 7.6% 的准确率;与单层的双向 LSTM 网络的结果相比,在 2017 数据集上提升了 4.2% 的准确率,在 2018 数据集上提升了 5.7% 的准确率,这充分验证了该算法的可行性和优势。

**关键词**:堆叠式双向 LSTM 网络;心律失常;心电图分类;深度学习

中图分类号 TP274

## Automatic Recognition of ECG Based on Stacked Bidirectional LSTM

WANG Wen-dao, WANG Run-ze, WEI Xin-lei, QI Yun-liang and MA Yi-de

School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China

**Abstract** For the growing demand of ECG data analysis, a new ECG classification algorithm is proposed. Firstly, the original data are truncated by fixed length, sample equilibrium is obtained, and the pre-processing operations such as instantaneous frequency and spectral entropy of the signal are obtained. After the data is preprocessed, the model can better extract features from the data for learning. In training progress, a two-way LSTM network stacking model is adopted. The stacked two-way LSTM model is an improved cyclic neural network model. Compared with convolutional neural networks, the cyclic neural network is more suitable for processing sequence data such as electrocardiogram. The experiment is conducted using MATLAB2018b under Windows for training and testing. The CUDA version is 9.0. The classification accuracy rate is used as an indicator to measure the performance of the model. The model is tested on two data sets, one is the data of the 2017 Physiological Signal Challenge (hereinafter referred to as the 2017 dataset), the final classification accuracy rate is 97.4%; the other is the data of the 2018 Physiological Signal Challenge (hereinafter referred to as the 2018 dataset), and the final classification accuracy rate is 77.6% on this dataset. The MATLAB group to which it belongs has achieved the third place. This algorithm improves the accuracy of 5.6% in the 2017 dataset and 7.6% in the 2018 dataset compared to the results of the traditional LSTM network. Compared to the results of a single-layer bidirectional LSTM network, in the 2017 data set, the accuracy rate improves 4.2%, and the accuracy rate improves 5.7% in the 2018 data set, which fully verifies the feasibility and advantages of the algorithm.

**Keywords** Stacked bidirectional LSTM network, Arrhythmia, Electrocardiogram classification, Deep learning

## 1 引言

常是心血管疾病中一类常见的疾病。心电图是一种重要的心电信号监测方法<sup>[2]</sup>,是专业医生进行辅助诊断和病理分析的重要工具。一种可靠、准确、快速的心电图分类方法<sup>[3]</sup>可以极

心血管疾病被称为危害人类健康的头号杀手<sup>[1]</sup>,心律失常

收到日期:2019-06-26 返修日期:2019-09-20 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(61175012,60872109,60572011);甘肃省自然科学基金(18JR3RA288);中央高校基本科研业务费专项资金(lzujbky-2017-it72)

This work was supported by the National Natural Science Foundation of China (61175012,60872109,60572011), Natural Science Foundation of Gansu Province, China(18JR3RA288) and Fundamental Research Funds for the Central Universities of Ministry of Education of China (lzujbky-2017-it72).

通信作者:马义德(mayide510510@qq.com)

大地减轻医生的工作负荷,具有较高的价值<sup>[4]</sup>。本文提出了一种基于堆叠式的双向 LSTM<sup>[5]</sup>模型来进行心电图自动分类,在 2017 数据集上<sup>[6]</sup>对单导联心电图信号进行了二分类(区分正常和心房颤动(AFib)的信号);在 2018 数据集上对十二导联心电图信号进行了九分类<sup>[7]</sup>。利用深度学习<sup>[8-9]</sup>模型对心电图<sup>[10]</sup>信号进行自动分类<sup>[11]</sup>是一个较新的领域。Lipton 等<sup>[12]</sup>提出了一种用于多标签分类任务的 LSTM 网络,该网络可以从心电信号序列的所有时间点上提取特征。Zihlmann 等<sup>[13]</sup>采用卷积递归神经网络(CRNN)架构进行心电图信号分类。Rajpurkar 等<sup>[14]</sup>采用序列到序列学习的方法来进行训练,其中每个 ECG 信号被视为一段固定长度的短片段,并且每个信号的标签被视为每个片段的标签;还采用了 ResNET-34 深度卷积网络架构。本文方法使用的模型与 Lipton 等提出的模型属于同一个体系,但是在网络结构上存在几点不同,本文使用的 LSTM 网络为双向 LSTM 网络<sup>[15]</sup>,且使用了堆叠式的深层结构<sup>[16]</sup>。

## 2 LSTM 网络

### 2.1 循环神经网络

循环神经网络(RNN)是通过神经元循环赋值存储信息的神经网络,其中每个神经网络单元都会将信息传入下一个单元<sup>[17]</sup>。

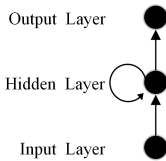


图 1 循环神经网络结构

Fig. 1 Cyclic neural network structure

图 1 给出了一个典型的循环神经网络结构,循环神经网络通常是用来处理序列数据的。循环神经网络指的是一个序列当前的输出与之前的输出有关,具体的表现形式为网络会对前面的信息进行记忆,将其保存在网络的内部状态中,并将其应用于当前输出的计算中,即隐含层之间的节点不再无连接而是有连接的,并且隐层的输入不仅包含输入层的输出还包含上一时刻隐层的输出。理论上,循环神经网络能够处理任何长度的序列数据,但是在实践中,对于长度超过 1000 的序列,预测结果非常糟糕。

循环神经网络在处理超长序列时遇到的最大难题是梯度消失和梯度爆炸。针对这个问题,Hochreiter 等提出了长短期记忆网络(LSTM)。

### 2.2 长短期记忆网络

长短期记忆网络是循环神经网络的改进版本。与标准 RNN 相比,LSTM 结构在捕获长期依赖性方面要好得多。在 LSTM 中,隐藏层中的每个传统节点都被存储器单元替换。存储器单元是 LSTM 中最重要的结构,可以避免梯度消失和梯度爆炸。LSTM 单元由存储器单元和它们包含的门单元共同组成<sup>[18]</sup>。乘法输入门单元是用来防止无关输入可能带来的负面影响。输入流经过输入门处理后进入存储器单元,存储器的输出流通过输出门到其他 LSTM 单元。图 2 给出了 LSTM 存储器单元的内部结构。

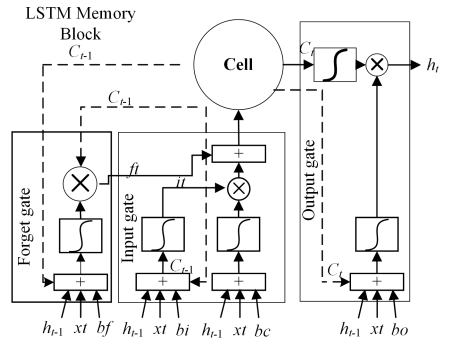


图 2 LSTM 存储器单元的内部结构

Fig. 2 Internal structure of LSTM memory unit

存储块结构中的遗忘门由简单的单层神经网络控制。遗忘门的激活函数如式(1)所示:

$$f_t = \sigma(w[x_t, h_{t-1}, c_{t-1}] + b_f) \quad (1)$$

其中, $x_t$ 是输入序列, $h_{t-1}$ 是前一个块的输出, $c_{t-1}$ 是前一个存储器单元, $b_f$ 是偏置矢量, $w$ 表示每个输入的单独权重向量, $\sigma$ 是逻辑 S 形函数。如果激活输出包含接近 0 的值,遗忘门将忘记之前输入的任务。输入门位于新存储器所在的部分,具有 tanh 激活功能和记忆阻滞效应,这些功能的实现方式如式(2)、式(3)所示:

$$i_t = \sigma(w[x_t, h_{t-1}, c_{t-1}] + b_i) \quad (2)$$

$$C_t = f_t * C_{t-1} + i_t * \tanh(w[x_t, h_{t-1}, C_{t-1}] + b_c) \quad (3)$$

从内存单元到网络其余部分的唯一输出来源于输出门,输出的计算方法如式(4)、式(5)所示:

$$o_t = \sigma(w[x_t, h_{t-1}, C_{t-1}] + b_o) \quad (4)$$

$$h_t = \tanh(C_t) * O_t \quad (5)$$

### 2.3 堆叠式双向长短期记忆网络

双向 LSTM 网络可以消除单向 LSTM 网络的一些限制,其可以看作是前向和后向两种不同的循环神经网络,这两个网络通过同一个输出层输出信息,通过这种结构可以评估时间范围内顺序输入的过去和未来的情况。

在自然语言处理<sup>[19]</sup>、语音识别<sup>[20]</sup>、序列分类<sup>[21]</sup>、机器翻译<sup>[22]</sup>等很多领域,BILSTM 的效果都超过了传统的单向 LSTM。

深度神经网络在预测分类问题上的成功通常归因于网络的深度。鉴于 LSTM 对序列数据进行操作,这意味着加深网络增加了提取输入信号特征的时间。目前堆叠式 BILSTM 是用于序列预测问题的较好模型。堆叠式 BILSTM 架构可以定义为由多个 BILSTM 层组成的 BILSTM 模型。上面的 BILSTM 层提供序列输出,而不是单个值输出到下面的 BILSTM 层。图 3 给出了一个堆叠式 BILSTM 的模型。

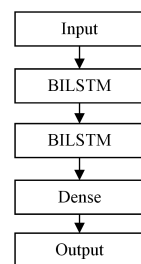


图 3 堆叠式 BILSTM 模型

Fig. 3 Stacked BILSTM model

### 3 数据集与方法

#### 3.1 数据集

本文使用了两个数据集,分别为 2017 数据集和 2018 数据集,2017 数据集的训练集包含 8528 个单导联心电图记录,持续时间为 30~60 s,每种类型的数量如表 1 所列,信号分为 4 个类型:正常(N)、心房纤颤(AFib)、其他节奏和噪声记录。对于 2017 数据集,本文只针对正常(N)和心房纤颤(AFib)进行分类,因此舍弃掉其他节奏和噪声信号,剩下的训练集数量为 5788 个心电图记录,测试集包含 3658 个相似长度的心电图记录。心电图记录采样频率为 300 Hz,它们还在 AliveCor 设备上进行了带通滤波。2018 数据集的训练集包含 6877 (女性 3178 名,男性 3699 名)个人的十二导联心电图记录,持续时间为 6~60 s,每种类型数量如表 2 所列,测试集包含 2954 个长度相同的心电图记录,对公众未公开。心电图记录采样频率为 500 Hz。

表 1 2017 生理信号挑战赛数据集分布

Table 1 2017 Physiological signal challenge dataset distribution

类型	记录数量
正常(Normal)	5050
心房纤颤(AF)	738

表 2 2018 生理信号挑战赛数据集分布

Table 2 2018 Physiological signal challenge dataset distribution

类型	记录数量
正常(Normal)	918
心房纤颤(AF)	1098
一级房室传导阻滞(I-AVB)	704
左束支传导阻滞(LBBB)	207
右束支传导阻滞(RBBB)	1695
早发性心房收缩(PAC)	556
室性早搏(PVC)	672
ST 段压低(STD)	825
ST 段升高(STE)	202

#### 3.2 数据的预处理

2017 数据集集中的 ECG 记录持续时间为 30~60 s,平均持续时间为 45 s;2018 数据集集中的 ECG 记录持续时间为 6~60 s,平均持续时间为 16 s。这两个数据集的数据都是在高采样率(300/500 Hz)下收集的,都存在着很长的序列(最长的序列长度达到 5 万以上),而 LSTM 的适宜范围大概为 500~1000 的序列长度,原封不动地输入数据会导致训练时间大大增长。另外,尝试在很长的序列中进行反向传播可能会导致梯度消失,从而削弱模型的可靠性,降低准确率,因此需要对数据进行一些预处理。本文方法对数据集集中的序列长度进行了统计:2017 数据集的序列长度大多集中在 9000 左右,统计结果如图 4 所示;2018 数据集的序列长度大多数集中在 5000 左右,统计结果如图 5 所示。本文算法将信号序列统一截断至数据集中最多的序列长度,同时,如果序列足够长则将其分成尽可能多的子序列,如在 2017 数据集中,长度为 18200 的序列可以分成两个长度为 9000 的子序列,长度不足 9000 的序列则舍去。此外,在两个数据集上都存在着样本不均衡的问

题。为了避免这种情况,本文算法通过复制数量少的类型,使所有类型数据的数量尽可能地接近。由于过采样可能导致拟合和数据冗余,本文算法在复制数据的同时,还对数据进行了一些变换,如调换序列前后的顺序等,然后将所有数据按 9:1 的比例随机分为训练集和测试集。从数据中提取特征有助于提高分类器的训练和测试精度,使用短时傅里叶变换在时间窗口中计算频谱图<sup>[23]</sup>,时间频率矩可以从时频图像中提取信息,每个时刻都可以作为一维特征输入 LSTM,同时基于信号的功率谱图计算瞬时频率和光谱熵<sup>[24]</sup>。将基于功率谱图得到的瞬时频率和光谱熵分别作为一维特征,然后将提取到的瞬时频率和光谱熵进行连接,这样每个导联的数据单元具有了二维特征。由于瞬时频率和光谱熵的数值几乎相差一个数量级,此外瞬时频率可能存在均值大高的问题,因此这可能导致 LSTM 网络无法进行有效的学习。当网络适合具有较大平均值和大范围值的数据时,大输入可能会减慢网络的学习和收敛速度。本文采用使用训练集平均值和标准差来标准化训练集和测试集的方法来解决这两个问题。

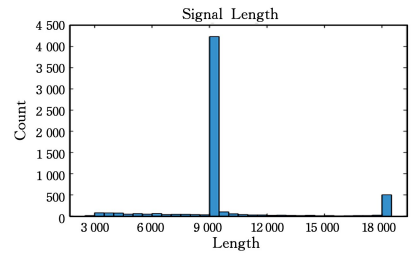


图 4 2017 数据集序列长度统计

Fig. 4 2017 dataset sequence length statistics

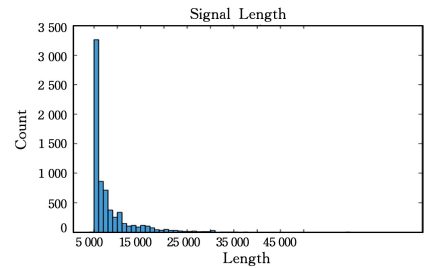


图 5 2018 数据集序列长度统计

Fig. 5 2018 dataset sequence length statistics

标准化,即 z 评分,是在训练中提高网络性能的一种流行方式,通过求取数据平均值和标准差来标准化训练集和测试集,具体算法如式(6)所示:

$$x^* = \frac{x - \bar{x}}{\delta} \quad (6)$$

其中, $\bar{x}$ 为数据的均值, $\delta$ 为数据的标准差。

#### 3.3 模型结构

本文算法使用的模型是基于堆叠式的双向 LSTM 网络构建的,因为在两个数据集上进行的分类任务有所不同,所以本文算法针对不同的任务设计了不同的模型。数据经过预处理后,2017 数据集的数据转换为了 2 维向量序列,2018 数据集的数据转换为了 24 维向量序列,因此模型的序列输入层的输入大小分为 2 和 24。在序列输入层后是两个双向 LSTM

单元,单向的 LSTM 是根据前面的信息推出后面的信息<sup>[25]</sup>,但有时只看前面的信息是不够的,有些预测可能需要由前面的若干输入和后面的若干输入共同决定,这样预测结果会更加准确<sup>[11]</sup>。双向 LSTM 网络的隐藏层要保存两个值,即  $A$  和  $A'$ ,  $A$  参与正向计算,  $A'$  参与反向计算。最终的输出值  $y$  取决于  $A$  和  $A'$ 。双向 LSTM 中每个 LSTM 单元对的一部分处理输入序列的一半,对于普通 LSTM 来说,双向 LSTM 在每个 LSTM 单元对上减少了一半的数据量,大大减轻了每个单元的负担。每个双向 LSTM 单元由前向和后向流组成,并且在每个时间步骤中组合两个流的输出,然后将最后一步前向和后向流的输出连接起来并馈送到分类层。同时,为了防止过拟合,在每个双向 LSTM 单元后加入了 Dropout 层<sup>[26]</sup>(丢弃概率为 20%),分类层使用 Softmax 激活函数来获得最终的 2/9 维输出。模型结构及参数分别如图 6、图 7 所示。

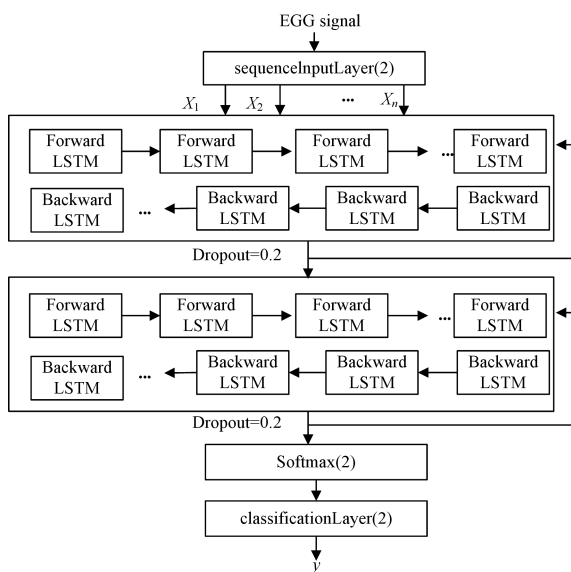


图 6 在 2017 数据集上使用的模型结构

Fig. 6 Model structure used on 2017 dataset

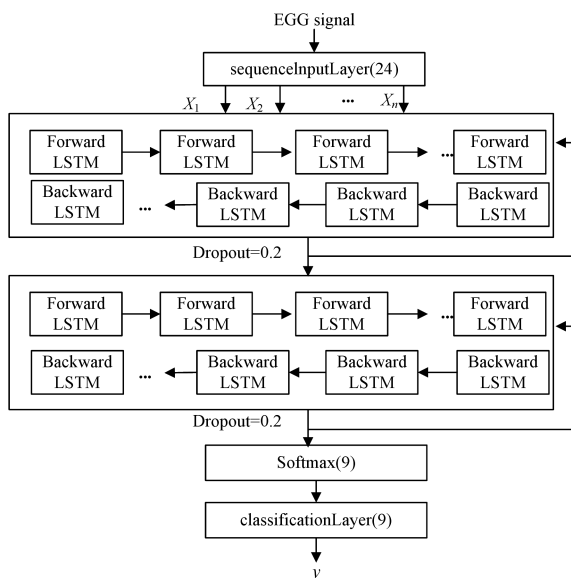


图 7 在 2018 数据集上使用的模型结构

Fig. 7 Model structure used on 2018 dataset

### 3.4 训练参数设置

本文在训练时使用了 Adam<sup>[27]</sup>求解器。与 Sgdm 求解的默认随机梯度下降相比, LSTM 这样的递归神经网络(RNN)在训练时使用 Adam 求解器的效果更好。使用 He 初始化方案<sup>[28]</sup>初始化权重,初始学习率设置为 0.01, *Mini-BatchSize* 设置为 150, *MaxEpochs* 设置为 100。

## 4 实验结果与讨论

### 4.1 训练

为了训练网络模型,按 9:1 的比例将数据集划分为训练集和验证集。按照 2.2 节的方法,通过复制的手段进行数据增强,使所有类别样本数量尽可能相近,在 2017 数据集上使用 8876 个数据(正常 4438 个,心房纤颤 4438 个)进行训练,将剩下的 980 个数据(正常 480 个,心房纤颤 480 个)作为测试集。在 2018 数据集上使用 6190 个数据进行训练,由于异常种类较多,因此未采取复制的手段进行数据增强。本文算法采取在评价指标中通过计算类频率加权平均值来判断模型的训练效果。此外,对于 2018 数据集,本文算法先将十二导联的心电图数据拆分成 12 个单导联的数据,经过截断处理后共有 101136 个数据,将剩下的 11232 个数据作为测试集,每个子数据的标签继承父数据的标签。将数据求取瞬时频率和光谱熵进行标准化后,将处理后的子序列重新合并成十二导联数据(共 24 维),此时用于训练的数据为 8428 个,剩下的 936 个数据作为测试集。实验发现,对于十二导联心电图, LSTM 网络需要结合全部维度的特征才能获得好的训练效果。

### 4.2 验证集上的结果

用训练好的网络对验证集进行预测,同时为了证明堆叠式双向 LSTM 网络的优势,本文算法按照 3.1 节的训练方法来训练堆叠式 LSTM 和单层的双向 LSTM 网络以进行测试,三者验证集得到的结果证明了堆叠式双向 LSTM 相比未改进版本的 LSTM 模型的优势。

### 4.3 结果和分析

#### 4.3.1 评价指标

本文通过计算每个类别的  $F_1$  和总体的  $F_1$  得分,在验证集上评估模型的性能。每个类别  $F_1$  的计算方式如式(7)所示:

$$F_{1i} = \frac{2 * N_{ii}}{\sum_{j=1}^9 (N_{ij} + N_{ji})} \quad (7)$$

如表 3 所列,  $N_{ij}$  代表第  $i$  类在验证集上被分为类别  $j$ 。本文采用两种不同的方式来计算总体的  $F_1$  得分。第一种是采用每个类别  $F_1$  的均值,如式(8)、式(9)所示:

$$F_1 = \frac{\sum_{i=1}^9 F_{1i}}{2} \quad (8)$$

$$F_1 = \frac{\sum_{i=1}^9 F_{1i}}{9} \quad (9)$$

表3 变量的命名规则

Table 3 Naming rules for variables

		预测								
		Normal	AF	I-AVB	LBBB	RBBB	PAC	PVC	STD	STE
标签	Normal	N <sub>11</sub>	N <sub>12</sub>	N <sub>13</sub>	N <sub>14</sub>	N <sub>15</sub>	N <sub>16</sub>	N <sub>17</sub>	N <sub>18</sub>	N <sub>19</sub>
	AF	N <sub>21</sub>	N <sub>22</sub>	N <sub>23</sub>	N <sub>24</sub>	N <sub>25</sub>	N <sub>26</sub>	N <sub>27</sub>	N <sub>28</sub>	N <sub>29</sub>
	I-AVB	N <sub>31</sub>	N <sub>32</sub>	N <sub>33</sub>	N <sub>34</sub>	N <sub>35</sub>	N <sub>36</sub>	N <sub>37</sub>	N <sub>38</sub>	N <sub>39</sub>
	LBBB	N <sub>41</sub>	N <sub>42</sub>	N <sub>43</sub>	N <sub>44</sub>	N <sub>45</sub>	N <sub>46</sub>	N <sub>47</sub>	N <sub>48</sub>	N <sub>49</sub>
	RBBB	N <sub>51</sub>	N <sub>52</sub>	N <sub>53</sub>	N <sub>54</sub>	N <sub>55</sub>	N <sub>56</sub>	N <sub>57</sub>	N <sub>58</sub>	N <sub>59</sub>
	PAC	N <sub>61</sub>	N <sub>62</sub>	N <sub>63</sub>	N <sub>64</sub>	N <sub>65</sub>	N <sub>66</sub>	N <sub>67</sub>	N <sub>68</sub>	N <sub>69</sub>
	PVC	N <sub>71</sub>	N <sub>72</sub>	N <sub>73</sub>	N <sub>74</sub>	N <sub>75</sub>	N <sub>76</sub>	N <sub>77</sub>	N <sub>78</sub>	N <sub>79</sub>
	STD	N <sub>81</sub>	N <sub>82</sub>	N <sub>83</sub>	N <sub>84</sub>	N <sub>85</sub>	N <sub>86</sub>	N <sub>87</sub>	N <sub>88</sub>	N <sub>89</sub>
	STE	N <sub>91</sub>	N <sub>92</sub>	N <sub>93</sub>	N <sub>94</sub>	N <sub>95</sub>	N <sub>96</sub>	N <sub>97</sub>	N <sub>98</sub>	N <sub>99</sub>

由于 2018 数据集未进行样本不均衡的处理,通过计算类加权平均值得到总体的  $F_1$  得分,如式(10)所示:

$$F_1 = \frac{N_i}{N} * F_{1i} \quad (10)$$

其中,  $N$  是验证集的总数,  $N_i$  表示验证集中第  $i$  类的数目。

#### 4.3.2 实验结果

表 4 列出了模型在 2017 数据集和 2018 数据集上得到的  $F_1$  分数。可以看出,在 2017 数据集上,模型对于心房颤颤的分类效果略好于正常的分类。在 2018 数据集上,模型同样对心房颤颤的分类效果最好,在 ST 段压低(STD)上表现最差。

表 4 模型在 2017 数据集和 2018 数据集上每种类型的  $F_1$  得分Table 4 Model for each type of  $F_1$  score on 2017 and 2018 data sets

类型	2017 数据集	2018 数据集
	$F_1$ 得分	$F_1$ 得分
正常 (Normal)	0.973	0.809
心房颤颤 (AF)	0.975	0.855
一级房室传导阻滞 (I-AVB)	/	0.741
左束支传导阻滞 (LBBB)	/	0.767
右束支传导阻滞 (RBBB)	/	0.768
早发性心房收缩 (PAC)	/	0.758
室性早搏 (PVC)	/	0.751
ST 段压低 (STD)	/	0.737
ST 段升高 (STE)	/	0.739

表 5 列出了在两个数据集上使用两种评估体系得到总体的  $F_1$  得分。从加权  $F_1$  得分可以看出,模型对每种类型心电图的判别能力并未有很大的偏差。

表 5 模型在 2017 数据集和 2018 数据集上的  $F_1$  总得分Table 5 Model  $F_1$  total score on 2017 and 2018 data sets

	均值 $F_1$ 得分	加权 $F_1$ 得分
2017 数据集	0.974	0.974
2018 数据集	0.769	0.777

将本文模型与堆叠式 LSTM 模型和单层双向 LSTM 模型进行对比实验。从表 6 可以看出,本文算法使用的堆叠式双向 LSTM 模型在性能上优于堆叠式 LSTM 模型和单层双向 LSTM 模型。与传统 LSTM 网络的结果相比,本文模型在 2017 数据集上准确率提升了 5.6%,在 2018 数据集上准确率提升了 7.6%;与单层的 LSTM 网络的结果相比,本文模型在 2017 数据集上准确率提升了 4.2%,在 2018 数据集上准确率提升了 5.7%,这充分验证了本文算法的可行性和优势。

表 6 3 种网络在两个数据集上的加权  $F_1$  得分Table 6 Weighted  $F_1$  scores for three networks on two data sets

	堆叠式 LSTM	单层的 LSTM	堆叠式双向 LSTM
2017 数据集加权 $F_1$ 得分	0.918	0.932	0.974
2018 数据集加权 $F_1$ 得分	0.701	0.720	0.777

为了更好地评估模型的性能,本文还引入了混淆矩阵对模型性能进行评估<sup>[29]</sup>。混淆矩阵的每一列代表着测试数据集中数据样本的真实类型,每一列的总和是该类型数据样本的真实数量;混淆矩阵的每一行代表着测试数据集中数据样本的预测结果,每一行的总和是该类型数据样本的预测数量。图 8 和图 9 给出了模型在两个数据集上得到的混淆矩阵,其中绿色表格代表着正确分类的样本,红色表格代表着错误分类的样本。九种心电图类型的英文缩写分别为正常-1, AF-2, IAVB-3, LBBB-4, RBBB-5, PAC-6, PVC-7, STD-8, STE-9。从图 8 所示的混淆矩阵的第一列可以看出,所有标签为心房颤颤的数据均得到了正确的识别,这表明本文模型对于心房颤颤类数据的特征学习的效果较好;第二列上 5.1% 的正常信号被错分为心房颤颤,可以看到有部分标签为正常的信号具有与异常信号相似的特征,这就对 LSTM 网络造成了一定的干扰,迫使模型做出了错误的预测。

Testing Accuracy Confusion Matrix

Output Class N	A	490 50.0%	25 2.6%	95.1% 4.9%
	N	0 0.0%	465 47.4%	100% 0.0%
	A	100% 0.0%	94.9% 5.1%	97.4% 2.6%
		A	N	Target Class

图 8 2017 生理信号挑战赛数据集结果的混淆矩阵(电子版为彩色)

Fig. 8 Results on 2017 physiopath challenge dataset

Testing Accuracy Confusion Matrix

Output Class	1	76 8.1%	4 0.4%	10 1.1%	0 0.0%	4 0.4%	7 0.7%	2 0.2%	4 0.4%	2 0.2%	69.7% 30.3%
	2	1 0.1%	117 12.5%	4 0.4%	1 0.1%	5 0.5%	10 1.1%	4 0.4%	8 0.9%	0 0.0%	78.0% 22.0%
	3	13 1.4%	3 0.3%	71 7.6%	0 0.0%	5 0.5%	4 0.4%	2 0.2%	5 0.5%	2 0.2%	67.7% 32.4%
	4	0 0.0%	0 0.0%	2 0.2%	32 3.4%	0 0.0%	1 0.1%	0 0.0%	0 0.0%	0 0.0%	91.4% 8.6%
	5	1 0.1%	2 0.2%	3 0.3%	0 0.0%	165 17.6%	5 0.5%	4 0.4%	3 0.3%	1 0.1%	89.7% 10.3%
	6	13 1.4%	9 1.0%	4 0.4%	0 0.0%	4 0.4%	73 7.8%	7 0.7%	11 1.2%	2 0.2%	59.3% 40.7%
	7	2 0.2%	1 0.1%	1 0.1%	0 0.0%	1 0.1%	5 0.5%	96 7.4%	2 0.2%	2 0.2%	87.3% 12.7%
	8	6 0.6%	4 0.4%	1 0.1%	0 0.0%	4 0.4%	0 0.0%	3 0.3%	69 7.4%	1 0.1%	78.4% 21.6%
	9	1 0.1%	1 0.1%	2 0.2%	0 0.0%	1 0.1%	0 0.0%	0 0.0%	0 0.0%	27 2.9%	84.4% 15.6%
		67.3% 32.7%	83.0% 17.0%	72.4% 27.6%	97.0% 3.0%	87.3% 12.7%	69.5% 30.5%	81.4% 18.6%	67.6% 32.4%	73.0% 27.0%	77.6% 22.4%
		1	2	3	4	5	6	7	8	9	Target Class

图 9 2018 生理信号挑战赛数据集结果的混淆矩阵(电子版为彩色)

Fig. 9 Results on 2018 Physiopath Challenge Dataset

在 2018 数据集上同样存在这样的情况,如图 10、图 11 所示。因为模型在 2017 数据集上表现较好,下面对 2018

数据集的几个例子进行分析,可以看出,由于存在噪声(如 A5906)或类似于病理的模式(用黑色矩形标记),这些信号被错误地预测。

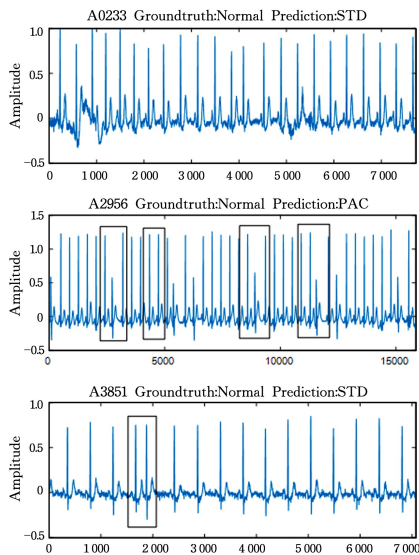


图 10 预测为异常的正常 ECG 信号的一些示例 1

Fig. 10 Some examples of normal ECG signals predicted to be abnormal 1

Testing Accuracy Confusion Matrix

	A	N		
Output Class	A	490 50.0%	25 2.6%	95.1% 4.9%
	N	0 0.0%	465 47.4%	100% 0.0%
		100% 0.0%	94.9% 5.1%	97.4% 2.6%
	A	N	Target Class	

图 11 预测为异常的正常 ECG 信号的一些示例 2

Fig. 11 Some examples of normal ECG signals predicted to be abnormal 2

图 12 所示的示例中 A3621 的 QRS 复合波类似于正常 ECG;而 A5085 的信号表现出了与 AF 相似的 RR 间期变化。

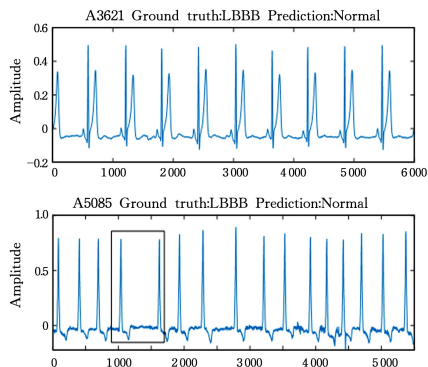


图 12 错误分类的 LBBB 信号示例

Fig. 12 Example of misclassified LBBB signal

本文模型对类型为 ST 段异常的心电图记录检测表现不佳。尽管有明确的病理学视觉证据,但很多属于 STD 和 STE

这两个类别的心电图记录被误分类为正常。部分例子如图 13、图 14 所示。

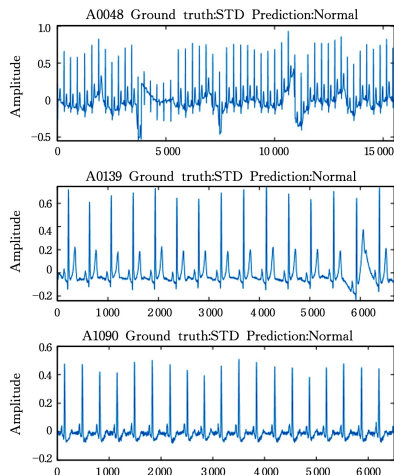


图 13 分类为正常的 STD 异常的例子 1

Fig. 13 Example 1 classified as a normal STD anomaly

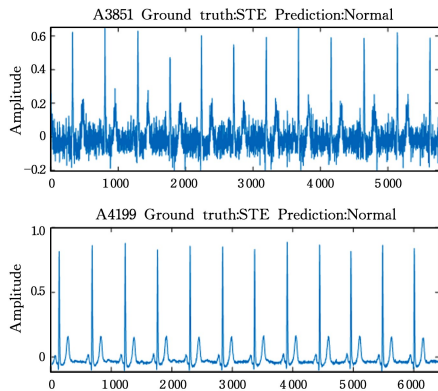


图 14 分类为正常的 STE 异常的例子

Fig. 14 Example of normal STE exception

**结束语** 本文算法基于堆叠式的双向 LSTM 创建了一个模型,用于检测单导联和多导联心电信号中的异常。该模型在 2017 数据集的验证集上的均值  $F_1$  得分为 0.974,在 2018 数据集的验证集上的均值  $F_1$  得分为 0.769,加权  $F_1$  得分为 0.777。由该模型在 2017 数据集上得到的结果可知,对于单导联信号的二分类问题,该模型已经表现得足够优秀。而在 2018 数据集上,模型还有一定的提升空间,对于 ST 段异常的判断不够准确,但每类之间的总体准确率相差不大。未来的工作重点是数据的预处理,因为直接对数据进行过采样可能会带来数据冗余,所以考虑使用权重平衡法代替过采样来处理数据不平衡的问题,即增加稀缺类别样本计算损失时的权重,其次可能在网络结构上堆叠更深的双向 LSTM,或者将序列通过预处理转换为图像的形式,用卷积神经网络进行分类。

### 参考文献

[1] MEMBERS W G, MOZAFFARIAN D, BENJAMIN E J, et al. Executive Summary: Heart Disease and Stroke Statistics--2016 Update: A Report from the American Heart Association[J]. Circulation, 2016, 127(1): 143-152.

- [2] 何方田. 临床心电图详解与诊断[M]. 杭州: 浙江大学出版社, 2010.
- [3] ENGIN M. ECG Beat Classification Using Neuro-Fuzzy Network[J]. Pattern Recognition Letters, 2004, 25(15): 1715-1722.
- [4] YU S N, CHEN Y H. Electrocardiogram Beat Classification Based on Wavelet Transformation and Probabilistic Neural Network[J]. Pattern Recognition Letters, 2007, 28(10): 1142-1150.
- [5] GRAVES A. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [6] TEIJEIRO T, GARCIA C A, CASTRO D, et al. Arrhythmia Classification from the Abductive Interpretation of Short Single-Lead ECG records [J/OL]. <http://www.cinc.org/archives/2017/pdf/166-054.pdf>.
- [7] 陈新. 临床心律失常学[M]. 北京: 人民卫生出版社, 2009.
- [8] LECUNY Y, BENGIO Y, HINTON G. Deep Learning[J]. Nature, 2015, 521(7553): 365-436.
- [9] MA Y H. Review of in-depth study [J]. Reading Abstracts (Middle), 2017, 804(6): 594-628.
- [10] OUYANG K. Biomedical Signal Processing [J]. Beijing Biomedical Engineering, 2006, 445(9): 655-720.
- [11] JIN L P, DONG J. Deep Learning Algorithms for Clinical ECG Analysis [J]. Chinese Science: Information Science, 2015, 45(3): 398-416.
- [12] LIPTON Z C, KALE D C, ELKAN C, et al. Learning to Diagnose with Lstm Recurrent Neural Networks[J]. Computer Science, 2015, 123(4): 223-268.
- [13] ZIHLMANN M, PEREKRESTENKO D, TSCHANNEN M. Convolutional Recurrent Neural Networks for Electrocardiogram Classification[J]. Computer Science, 2017, 57(12): 148-205.
- [14] RAJPURKAR P, HANNUN A Y, HAGHPANAHI M, et al. Cardiologist - level Arrhythmia Detection with Convolutional Neural Networks [J]. Computer Science, 2017, 225(5): 308-374.
- [15] GRAVES A, SCHMIDHUBER J. Framewise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures[J]. Neural Networks, 2005, 18(5/6): 602-610.
- [16] PRAKASH A, HASAN S A, LEE K, et al. Neural Paraphrase Generation with Stacked Residual Lstm Networks [J]. 2016, 65(8): 439-512.
- [17] GRAVES A. Supervised Sequence Labelling with Recurrent Neural Networks [J]. Studies in Computational Intelligence, 2008, 552(35): 385-443.
- [18] DONGHYUN. Long Short-Term Memory Recurrent Neural Network - based Acoustic Model Using Connectionist Temporal Classification on a Large-scale Training Corpus[J]. China Communications, 2017, 14(9): 23-31.
- [19] MIKOLOV T, ZWEIG G. Context Dependent Recurrent Neural Network Language Model[C]// Spoken Language Technology Workshop. 2013.
- [20] GRAVES A, MOHAMED A R, HINTON G. Speech Recognition with Deep Recurrent Neural Networks[C]// 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2013.
- [21] YUANFANG R, CHONG T, FEI L I, et al. Relation Classification Via Sequence Features and Bi-directional LSTMs[J]. Wuhan University Journal of Natural Sciences, 2017, 22(6): 489-497.
- [22] SENNRICH R, HADDOW B, BIRCH A. Neural Machine Translation of Rare Words with Subword Units[J]. Computer Science, 2015, 27(5): 224-245.
- [23] PONS J, LIDY T, SERRA X. Experimenting with Musically Motivated Convolutional Neural Networks[C]// 2016 14th International Workshop on Content-Based Multimedia Indexing (CBMI). IEEE, 2016.
- [24] WANG D L. Deep Learning Reinvents the Hearing Aid [J]. IEEE Spectrum, 2017, 54(3): 32-37.
- [25] GRAVES A. Supervised Sequence Labelling with Recurrent Neural Networks [J]. Studies in Computational Intelligence, 2008, 385(6): 57-128.
- [26] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting[J]. Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
- [27] MULVEY J M, RUSZCZYNSKI A. A New Scenario Decomposition Method for Large-scale Stochastic Optimization[J]. Operations Research, 1995, 43(3): 477-490.
- [28] HE K, ZHANG X, REN S, et al. Delving Deep Into Rectifiers: Surpassing Human-level Performance on Imagenet Classification [J]. Computer Science, 2015, 246(12): 450-480.
- [29] TOWNSEND J T. Theoretical Analysis of an Alphabetic Confusion Matrix[J]. Perception & Psychophysics, 1971, 9(1): 40-50.



**WANG Wen-dao**, born in 1995, post-graduate. His main research interests include in-depth study and processing of medical signals.



**MA Yi-de**, born in 1963, Ph.D, professor, is a member of China Computer Federation. His main research interests include digital image processing, biomedical engineering and artificial neural network.