2. 数据与方法

2.1 数据收集

在本研究中，我们采用开源且匿名化后的数据集Covid19-Cough [1] 训练咳嗽音分类模型。该数据集依托一个呼叫中心和Telegram消息机器人，对1324名受试者进行咳嗽音的收集，且询问受试者是否目前是Covid19阳性患者。在这批受试者中，有682人报告为阳性，其中382人具有PCR检测结果, 295为有症状患者。该数据集总共收集了总时长为58分钟的咳嗽音录音，每个录音样本平均时长为xxx. 因为我们的目标是根据咳嗽音判断患者肺部健康状况，我们将具有covid19症状的样本作为阳性样本，其他样本作为阴性样本。

同时，由于我们的应用在实际使用时可能会接收到各种声音干扰，我们需要使用一个模型来检测一段录音是否具有咳嗽声。Covid19-Cough数据集中的所有样本都是具有咳嗽音的。我们采用URBANSOUND8K数据集作为没有咳嗽音的样本。该数据集包含8732个具有标注的音频片段，记录了城市中可能会出现的各种声音，如空调声，汽车鸣笛，音乐，施工声，小孩嬉戏等。我们将Covid19-Cough数据集与URBANSOUND8K数据集组成的新的数据集叫做Cough classification Dataset 1 (CoughCLS-1)数据集.

2.3 数据预处理

当我们设备接收到一段录音之后，我们首先会对录音文件进行加载和重采样为16kHz的波形数据(numpy格式)，我们将波形数据统一padding为15 \* 16000 的numpy array。波形数据通过torchaudio工具包进行短时傅里叶变换(Short-Time Fourier Transform, STFT) 得到频谱表征(window size: 2048; stride: 302), 然后将频谱表征投影到20Hz 到24kHz的128组梅尔滤波器组上，得到梅尔频谱图。在训练时，我们对训练样本的音频采用添加噪声的方式(采用torchaudio实现)进行数据增强(data augmentation).

2.2 模型架构

如下图所示，我们主要采用模型框架进行咳嗽音分类。我们的模型框架主要有以下四个步骤：(a)首先将波形数据转化为梅尔频谱; (b) 梅尔频谱作为单通道图输入预训练的卷积神经网络中进行特征提取; (c) 卷积网络输出的特征图(feature map)经过集合平均池化，转化为特征向量; (d) 特征向量经过一个两层的全连接层进行池化，输出标签预测。我们的模型采用深度学习框架PyTorch进行开发。

作为对比，我们也对比了广泛在音频分析领域应用的特征提取+机器学习分类的方法[3]，具体步骤如下图：(a)首先对波形数据进行特征提取, 最后每个音频由一个特征向量表示；（b）特征向量通过机器学习模型得到分类结果。我们采用目前最强大的机器学习模型XgBoost[4]，LightGBM[5]， CatBoost[6]和 SVM作为分类器。

However, the most often  
used features are the *Mel-Frequency Cepstral Coefficients*(MFCC), computed as the discrete cosine transform of the logspectrogram aggregated into Mel-scale frequency bins. Analysis of the pairwise Mutual Information in [15] concluded that  
the commonly used acoustic features complement each other  
in cough detection tasks. Other engineered features used in the  
studies include spectral spread and centroid, Sample Entropy,  
which measures the complexity and self-similarity of the timeseries, non-Gaussianity statistics, and Linear Predictive Coding  
features, [16, 17, 18], and [19].

下面我们具体介绍模型架构的细节，以及模型的表现。

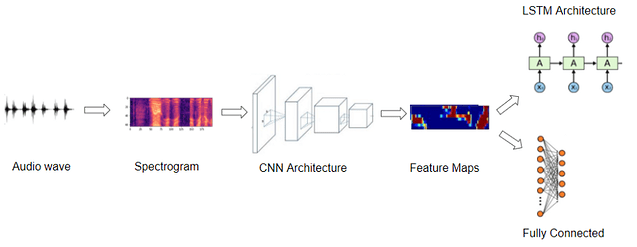


图1. 基于谱图与卷积神经网络的咳嗽因分类模型框架 （xxx, 重画， 采用实际样本替换图标）

图2. 基于音频特征提取的机器学习模型。 （补充架构图）

2.3 模型主干

本研究中我们采用卷积神经网络[7]从梅尔频谱图中提取特征。卷积网络的主干由多个卷积层，激活函数和池化层组成。卷积层可以产生一组平行的特征图，它通过在输入图像上滑动不同的卷积核并执行一定的运算而组成。此外，在每一个滑动的位置上，卷积核与输入图像之间会执行一个元素对应乘积并求和的运算以将感受野内的信息投影到特征图中的一个元素。卷积神经网络在大型图像处理任务上获得了巨大成功[8]，并引领了第二次人工智能浪潮。

本研究采用预训练+微调的范式，这也是目前表现最为优异的方法。经过预训练的网络已经学习了大量图片中蕴含的语义知识，且具有良好的参数初始化，有利于整个模型在下游任务上的快速优化。我们采用在ImageNet上预训练的卷积神经网络作为主干。目前较为流行的网络主干有MobileNet[9], EfficientNet[10], ResNet[11], ResNeXt[12], 以及NFNET. 其中ResNet[11]是最为广泛使用的网络结构，其提出的残差链接有利于更深的神经网络训练而不会有梯度消失问题，目前是各种神经网络必备的模块。MobileNet和EfficientNet在ResNet基础上采用更为轻量化的网络设计，更加利于端侧的模型部署与应用。NfNet[13]将网络中的标准化操作舍弃，使得模型更加容易在下游任务上进行参数适配。

2.4 快速集成方法

一般采用卷积网络作为主干进行特征提取后，我们会使用一个多层全连接层进行分类结果预测。在众多研究中，模型集成被大量使用。[14] 但是一般的模型集成需要采用多个不同的模型进行分类，这样不仅在训练过程中会消耗大量的显卡资源，而且在模型部署后推理过程中显著拖慢整个应用的速度。我们在本研究中开发了一种快速高效的模型集成方法，我们叫做快速集成方法（Fast-ensemble method）。快速集成方法的实现如下：我们在一个卷积网络主干后接上5个不同的全连接层作为分类头。在训练时，每个分类头接收到的特征向量由于dropout的随机性而不同，这样训练后分类头可以学习到不同的参数，相当于通过不同的视角理解图像的特征向量，获得多种不同的logits结果。模型输出的logits是由这五个分类头的logits求平均后的结果。

快速集成方法只在模型主干上增加了几个分类层，相比于模型主干，moxingtianjia

图3. 快速集成方法示意图. (待补充)

2.4 模型评估

2.4.1 咳嗽检测模型

2.4.2 基于咳嗽音的肺部健康检测模型

我们采用在Covid19-Cough数据集上进行5折交叉验证的方式进行模型评估，评估的主要指标为5折交叉验证平均测试AUC分数。同时，我们还汇报以0.5为阈值时的平均真阳性率和平均真阴性率。模型的表现对比见下面的表1. 在本研究中，我们默认采用预训练后的卷积网络。我们也对比了完全随机初始化的网络的表现，这些模型会标识” 未预训练”。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 模型架构 | Auc(%) | TPR(%) | TNR(%) |
| 机器学习模型 | XgBoost |  |  |  |
|  | LightGBM |  |  |  |
|  | CatBoost |  |  |  |
|  | LogisticRegression |  |  |  |
| 深度学习模型 | ResNet |  |  |  |
|  | ResNeXt |  |  |  |
|  | EfficientNet-v3 |  |  |  |
|  | MobileNet-v3  (未预训练) |  |  |  |
|  | MobileNet-v3 |  |  |  |
|  | MobileNet-v3  + Fast Ensemble |  |  |  |
|  | NfNet (未预训练) | 72.3 | 75.3 | 59.2 |
|  | NfNet | 82.1 | 80.1 | 63.8 |
|  | NfNet + Fast Ensemble | 83.5 | 72.8 | 79.8 |

表1. 各个不同的模型在Covid19-Cough数据集上进行5折交叉验证的表现。

我们可以看到，基于特征提取的机器学习模型在咳嗽音分类任务上表现明显差于基于深度卷积网络的模型，显示出深度卷积网络的表达能力和特征提取能力。在各个卷积网络主干的比较中，NfNet网络获得了最高的平均AUC分数。我们发现，MobileNet-v3在AUC分数上要稍低于NfNet，但是MobileNet-v3在模型体量上要明显小于NfNet，且推理速度要快于比NfNet快约十倍，非常适合在非GPU环境进行部署。所以我们最后在设备上采用MobileNet-v3模型。

[1] <https://github.com/covid19-cough/dataset>

[2] <https://urbansounddataset.weebly.com/urbansound8k.html>

[3] project achoo paper:

[4] XgBoost;

[5] LightGBM;

[6] CatBoost;

[7] 卷积神经网络：

[8] AlexNet

[9] MobileNet

[10] EfficientNet

[11] ResNet

[12] ResNeXt

[13]