

# 基于集成知识蒸馏的肺病分析

李坚强, 王 成, 黄志超, 陈 杰<sup>†</sup>

(深圳大学 计算机与软件学院, 广东 深圳 518060)

**摘要:** 新冠肺炎以来, 肺部疾病引起了人们更大的关注. 肺音的特征与诊断是肺病理学中重要的组成部分. 现有的肺音分析工作主要是对肺音的类型进行分类, 分类肺病的研究较少. 另外, 单个分类模型无法在保护隐私的前提下融合多方数据, 复杂的模型也难以保证分类的实时性. 针对这些问题, 本文提出集成知识蒸馏的肺疾病分类模型. 首先从肺音音频中提取梅尔频谱特征, 然后建立多个二分类卷积神经网络模型作为教师模型, 最后通过集成知识蒸馏技术, 将多个教师模型的知识集成到一个精简的多分类学生模型上. 实验表明, 该模型能够在预测准确率达95%的情况下, 参数量比教师模型减少79%, 预测速度提升20%. 在同等条件下, 耗时仅有MobileNet-v3-small模型的6%, 实现实时性肺疾病分析.

**关键词:** 肺疾病; 卷积神经网络; 集成知识蒸馏; 教师模型; 学生模型

**引用格式:** 李坚强, 王成, 黄志超, 等. 基于集成知识蒸馏的肺病分析. 控制理论与应用, 2020, 37(x): xxx – xxx

DOI: 10.7641/CTA.2020.xxxxx

## Lung disease analysis using ensemble knowledge distillation

LI Jian-qiang, WANG Cheng, HUANG Zhi-chao, CHEN Jie<sup>†</sup>

(College of Computer Science and Software Engineering, Shenzhen University, Shenzhen Guangdong 518060, China;)

**Abstract:** Since the outbreak of COVID-19, lung diseases have attracted more attention. The characteristics and diagnosis of lung sounds become an important part of pulmonary pathology. The existing works for lung sound analysis mainly aim to classify the types of abnormal lung sounds. There are few studies focusing on the classification of lung diseases. Moreover, a single classification model cannot take advantage of the train data from multiple sources due to privacy leakage concerns, and it is difficult for complex models to classify in real time. Therefore, this paper proposes a model for classifying lung diseases based on ensemble knowledge distillation. Firstly, Mel-spectrum features were extracted from lung sound, and then multiple binary convolutional neural network models were established as teacher models. Finally, a simplified multi-class student model will learn the knowledge of multiple teacher models through the technology of ensemble knowledge distillation. Our experiments show that the student model reduces 79% of the parameters and improves the prediction efficiency by 20% than teacher model while achieving a predictive accuracy of 95%. Under the same condition, the student model only incurs 6% of the time that is used by the state-of-the-art MobileNet-v3-small model. Thus, our model has potential to be deployed in real world for real-time diagnosis of the lung diseases.

**Key words:** lung diseases; convolutional neural network; ensemble knowledge distillation; teacher model; student model

**Citation:** LI Jianqiang, WANG Cheng, HUANG Zhichao, et al. Lung disease analysis using ensemble knowledge distillation. *Control Theory & Applications*, 2020, 37(x): xxx – xxx

## 1 引言

2019年12月以来, 新型冠状病毒肺炎(Corona Virus Disease 2019, COVID-19)陆续在世界各地爆发.

由于该病毒具有较强的传染性, 医护人员必须身穿多层防护服, 这给医护人员的肺音听诊工作带来较大的困难. 因此, 如果肺音能在采集之后, 传递给后台进行

收稿日期: 2020-04-30; 录用日期: 2020-xx-xx.

<sup>†</sup>通信作者. E-mail: chenjie@szu.edu.cn.

本文责任编辑:

国家自然科学基金项目(U1713212, 61836005, 61702341),

广东省杰出青年基金项目(2019B151502018),

深圳市技术研究基金项目(JSJG20180507182904693)和深圳市公共技术平台基金项目(GGFW2018021118145859)资助.

Supported by the National Natural Science Foundation of China (U1713212, 61836005, and 61702341) and

the Natural Science Foundation of Guangdong Province-Outstanding Youth Program(2019B151502018) and

the Technology Research Project of Shenzhen City(JSJG20180507182904693) and Public Technology Platform of Shenzhen City(GGFW2018021118145859).

智能化的分析,辅助医疗人员的诊断,就可以有效的减少医患接触频次,降低医护人员感染风险。

预计到2050年,中国60岁以上老年人口将达到4.98亿<sup>[1]</sup>,随着年龄的增长,肺纤维化、慢性阻塞性肺疾病的发病率也逐步上升,这意味着我国患肺部疾病的老年人的护理问题日趋严峻。近年来,我国患有呼吸系统疾病的人数在逐渐递增,呼吸系统疾病成为我国人口主要的死因之一<sup>[2]</sup>。

呼吸音是衡量肺部生理及病理的一项重要指标,其特征和诊断是肺病理学的重要组成部分。呼吸系统听诊已被证实是鉴别和观察不同类型肺部疾病的重要诊断手段,可以有效地发现肺部疾病,并且对于不同年龄或者发病程度存在差异的患者都能得到很好的应用<sup>[3]</sup>。

使用听诊器人工听诊的传统方式可能由于医生的经验不足或人耳对肺音低频段敏感性低等因素,导致诊断结果不够准确。在诊断一些异常的患者时,肺音的平稳性降低,也会使得医生的诊断变得更加困难<sup>[4]</sup>。听诊通常需要在安静的环境中进行,最好是病人处于静止状态,这样的环境要求也就限制了听诊的时间和灵活性。此外,由于呼吸系统症状可能发生在居家环境、公共场所,甚至夜间,这种由医护人员才能进行听诊的方法无法得到普遍的使用。

与主观听诊相比,对肺音的计算机分析可以进行可重复的肺音量化<sup>[5]</sup>。随着计算机技术在各个领域的发展与应用<sup>[6-7]</sup>,通过计算机技术分析肺音也受到了众多研究者的关注,已有很多学者专注于研究肺音分析和检测中的机器学习技术。机器学习算法应用于肺音诊断上,使得听诊工具的准确率得到很大的提升。研究者们提出了很多基于计算机肺部声音分析的机器学习算法,如人工神经网络、隐马尔科夫模型、K近邻算法、高斯混合模型、遗传算法和模糊逻辑等<sup>[8-10]</sup>。

由于肺部声音信号不稳定,因此常规的频率分析方法在诊断分类中效果欠佳,学者提出一种使用小波变换分析肺部声音信号以及使用人工神经网络进行分类的方法<sup>[9]</sup>,但是这种方法需要人工标记肺音的周期。异常肺音本身是一种非周期信号,其可能不会在整个记录中出现,Datta等人开发一种自动肺音分析仪,该工具在不需要额外的流量信号或标记呼吸周期的情况下,就能识别新的时频域肺部声音特征<sup>[11]</sup>,但是这种方法主要是区分正常肺音和异常肺音,如果用于肺音的细类区分,仍存在一定的局限性。听诊时可能出现各种各样的不确定音和噪声干扰肺音诊断,Yamashita等人基于隐马尔科夫模型提出“异常呼吸置信度”检测的两步分类方法<sup>[10]</sup>,但是该工作也只是关注单一类型的肺音。并且,如果通过少量的患者来简化问题,从小数据集学习到的特征并不具有泛化性,

为此,Chamberlain等人利用深度神经网络,将半监督深度学习应用于自动识别肺部声音<sup>[12]</sup>。减少评估呼吸音时固有的主观性,能够提高呼吸音分辨的准确率,Aykanat等人使用支持向量机和卷积神经网络(convolutional neural network, CNN),开发了一种性价比高、使用方便的电子听诊器,该电子听诊器能够通过一种非侵入性的方法对录音软件中所记录的呼吸声音进行分类<sup>[13]</sup>。

然而,现有的很多工作主要聚焦在异常呼吸音的分类上,如区分喘鸣音(wheeze)或者破裂音(crackle)<sup>[14]</sup>,直接利用肺音对肺疾病进行分类的研究较少。文献[15]使用CNN对肺疾病进行分类,但是复杂的模型不利于实时性的决策<sup>[16]</sup>,除此之外,由于医疗数据存在隐私保护问题,其也无法胜任多方数据不共享的场景。

本文为了解决上述的这些问题,受文献[17]的启发,提出集成知识蒸馏的肺疾病分类模型。该模型利用分布在各个单位的本地数据建立二分类的教师模型进行肺疾病的异常检测,通过多教师知识蒸馏模型,指导多分类学生模型学习,从而充分利用隐私受保护的多方肺音数据。同时,知识蒸馏兼顾学生模型规模与预测准确性,可以有效精简多分类学生模型,从而提高肺疾病分类的实时性。

## 2 集成蒸馏架构

本文提出集成知识蒸馏的肺疾病分类方法,其蒸馏架构图如图1所示。该模型假设不同医院间的模型相互独立,并且可利用医院本地数据集训练二分类(正常与某类肺疾病)模型作为教师模型。

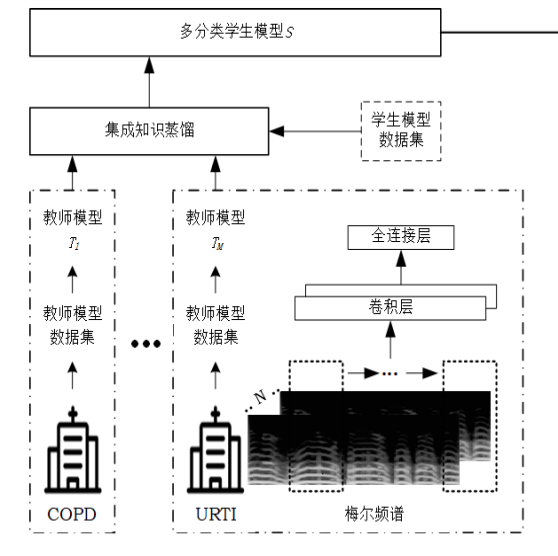


图1 集成蒸馏模型

Fig. 1 Ensemble knowledge distillation model

对于 $M$ 种疾病,本文将蒸馏架构中的多个教师模型定义为 $T_i, i \in \{1, \dots, M\}$ ,而医疗研究基础较为落后的单位在拥有少量的本地数据条件下,可通过蒸

馏多个二分类教师模型的知识, 训练精简的多分类学生模型 $S$ , 这样学生模型 $S$ 不仅能从输入的数据中学习, 也能向教师模型学习. 肺疾病分类任务是提取肺音音频中的特征, 建立特征与肺疾病之间的联系, 而由此对不同的肺疾病进行分类. 如图 2所示, 该模型将肺疾病分类任务分为音频数据增强, 提取梅尔频谱(Mel-spectrogram)特征, 数据归一化, 建立教师模型, 建立学生模型与知识蒸馏.

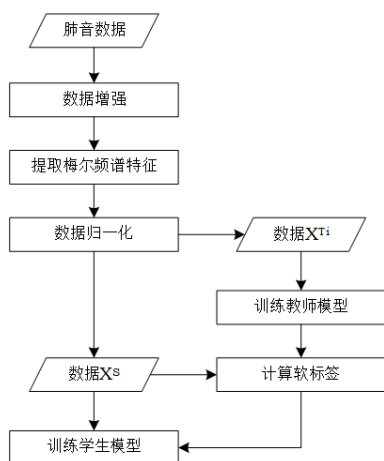


图 2 集成知识蒸馏流程

Fig. 2 Flow chart of Ensemble knowledge distillation

### 3 数据预处理

#### 3.1 数据定义

由于医院之间数据互不共享的限制, 无法直接的集中分析这些本地数据, 因此, 由不同的教师模型学习不同来源的数据集. 对于第 $i$ 种肺疾病, 本文利用教师模型 $T_i$ 学习该疾病的本地数据 $X^{T_i}$ ,  $X^{T_i}$ 数据对应的标签定义为 $Y^{T_i}$ . 而由学生模型 $S$ , 学习数据集 $X^S$ ,  $X^S$ 对应的标签为 $Y^S$ , 并且利用知识蒸馏技术将教师模型的知识集中到学生模型上, 这也就间接地利用了各个医院的数据资源.

#### 3.2 数据增强

适当的数据增强有利于提高模型的性能. 原始数据进行增强后能够生成更多的数据, 提高模型的泛化能力和鲁棒性. 下面将介绍本文的3种数据增强的方法: 时间变换、音高变换和 $\mu$ 律( $\mu$ -law)压缩.

时间变换旨在通过改变参数因子而影响音频的采样速率. 而音高变换就是以半音为单位, 对音频的音调进行调整. 在本文中, 时间变换和音高变换使用和文献[15]同样的变换参数, 时间变换4个取值分别为: 0.5, 0.7, 1.2, 1.5; 音高变换4个取值分别为: -2, -1, 1, 2.

$\mu$ 律压缩是一种动态范围压缩算法<sup>[18]</sup>, 对于输入的信号 $s$ ,  $-1 \leq s \leq 1$ , 其计算 $\mu$ 律压缩如公式1所示:

$$y_s = \text{sgn}(s) \frac{\ln(1 + \text{abs}(s)\mu)}{\ln(1 + \mu)}. \quad (1)$$

其中,  $\text{sgn}(s)$ 表示符号函数,  $\text{abs}(s)$ 表示取信号 $s$ 的绝对值,  $\mu$ 为压缩参数, 一般取值为2的幂减1. 本文数据增强中 $\mu$ 有4个取值: 7, 15, 31, 63.

#### 3.3 提取梅尔频谱特征

本文将采集到的肺音音频分成音频片段, 然后分别提取音频片段中的梅尔频谱特征, 梅尔频谱是一种比较典型的特征表示, 也被应用在其他音频分类任务中<sup>[19-20]</sup>. 图 3为从6类肺音音频中提取特征得到的梅尔频谱图, 这6类肺音分别为慢性阻塞性肺疾病(Chronic Obstructive Pulmonary Disease, COPD), 健康(Healthy), 上呼吸道感染(Upper Respiratory Tract Infection, URTI), 支气管扩张(Bronchiectasis), 肺炎(Pneumonia), 细支气管炎(Bronchiolitis).

#### 3.4 数据归一化处理

对肺音音频的分析问题, 通过提取音频的梅尔频谱之后, 转换为类似图片分类问题, 本文希望评估不同归一化的方法, 找到更加适合处理肺音音频数据的方法, 提高模型的性能. 在数据的归一化上, 除了评估文献[15]中提到的3种归一化方法: 均方根归一化(Root Mean Square Normalization), 峰值归一化(Peak Normalization)和EBU Standard R128归一化(European Broadcasting Union, EBU), 还评估了零均值归一化(Zero Mean Normalization).

这里使用的均方根归一化方法是以音频信号振幅的均方根(Root Mean Square, RMS)为参考值, 然后对音频进行缩放, 该方式可以比较好的衡量出信号振幅的强度. 均方根归一化的方式如公式2所示:

$$s_{\text{rms}} = \frac{s}{\sqrt{(s \cdot s) / |s|}}. \quad (2)$$

其中,  $|s|$ 表示信号长度,  $s_{\text{rms}}$ 表示信号进行RMS归一化后的结果.

峰值归一化即是以信号 $s$ 中振幅绝对值最大的为参考值, 然后将其他所有的振幅值按照该参考值进行缩放. 峰值归一化计算如公式3所示:

$$s_{\text{peak}} = \frac{s}{\max(\text{abs}(s))}. \quad (3)$$

其中,  $s_{\text{peak}}$ 表示信号 $s$ 进行峰值归一化后的结果.

EBU Standard R128归一化, 以下简称EBU R128归一化, 该方法是依据欧洲广播联盟提出的一个音频响度标准对音频进行处理. EBU Standard R128规定音频的响度水平为23 LUFS, LUFS(Loudness Units referenced to Full Scale)是一种衡量响度的单位.

零均值归一化是音频和图像处理任务中比较常用的归一化方法<sup>[21]</sup>. 对于信号 $s$ , 零均值归一化就是减去信号振幅的均值之后, 再除以振幅的标准差, 如公



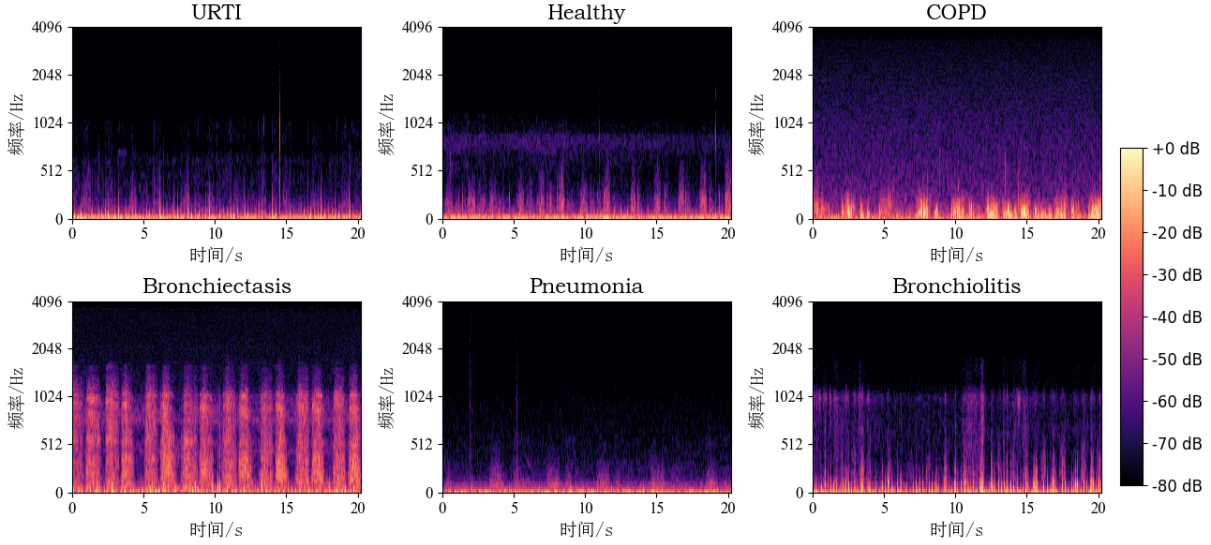


图3 梅尔频谱特征提取

Fig. 3 Mel-spectrogram feature extraction

式4所示:

$$s_z = \frac{s - \bar{s}}{\sigma}. \quad (4)$$

其中,  $\bar{s}$ 表示振幅的均值,  $\sigma$ 表示振幅的标准差.

## 4 模型定义

### 4.1 教师模型

本文的教师模型使用的是一个CNN模型, 模型共有6层 $l_1, \dots, l_6$ , 第 $l_1$ 层是有24个滤波器的卷积层, 卷积核大小为(5, 5), 步幅为(4, 2). 第 $l_2$ 层同样是卷积层, 包含48个滤波器, 卷积核大小、步幅与 $l_1$ 相同. 第 $l_3$ 层是一个最大池化层, 其pool size为(2, 2). 第 $l_4$ 层卷积层的滤波器数量和卷积核大小与 $l_2$ 相同. 第 $l_5$ 层全连接层有64个隐藏单元, 最后一层为输出层.

在训练教师模型的过程中, 从音频片段提取得到的梅尔频谱特征数据的维度将被重塑, 以匹配模型的输入维度. 每个教师模型分别对应一种肺疾病, 拟合来自不同医院的数据, 训练目标为最小化教师模型的预测结果和真实标签之间的交叉熵(cross-entropy)损失, 对于教师模型 $T_i$ , 定义其损失函数如公式5所示:

$$LossT_i = -\frac{1}{|X^{T_i}|} \sum_{n=1}^{|X^{T_i}|} y_n^{T_i} \log f^{T_i}(x_n^{T_i}). \quad (5)$$

这里的 $|X^{T_i}|$ 表示 $X^{T_i}$ 中元素数量,  $x_n^{T_i}$ 是 $X^{T_i}$ 中的一个数据点,  $y_n^{T_i}$ 是与 $x_n^{T_i}$ 相对应的类别标签,  $f^{T_i}$ 表示 $T_i$ 的预测函数.

### 4.2 知识蒸馏

来自多方互不共享的数据, 是无法像文献[22]的工作那样直接将教师-学生模型同时训练, 在训练的过程中传递知识. 对于每一个训练好的教师模型而言, 其已经学习了某一类疾病的数据特征, 如果把学生模

型遇到的问题, 先交由教师模型解答, 然后让学生模型在有参考的前提下进行训练, 同样可以学习到教师模型的现有知识.

对于数据集 $X^S$ 中的一个数据点 $x_n^S$ , 通过训练好的教师模型 $T_i$ 计算其软标签时, 使用文献[23]中的定义, 如公式6所示:

$$q_j^{(i)}(x_n^S) = \frac{\exp(z_j^{(i)}(x_n^S)/t)}{\sum_{j=0}^1 \exp(z_j^{(i)}(x_n^S)/t)}. \quad (6)$$

其中,  $z_j^{(i)}(\cdot)$ 表示教师模型 $T_i$ 对应类别 $j \in \{0, 1\}$ 时logits层的输出,  $t$ 为温控超参数,  $t$ 取值越大, 软标签的分布越平滑.

由于教师模型是一个二分类的模型, 其logits层的输出只有2个类别, 而对于另外的 $M-1$ 个类别, 本文采用取0值的方式, 以此将多个二分类模型的软标签转换为一个多分类的软标签 $\pi_j(x_n^S, y_n^S)$ , 与学生模型的类别维度统一,  $\pi_j(x_n^S, y_n^S)$ 的计算如公式7所示:

$$\pi_j(x_n^S, y_n^S) = \begin{cases} q_0^{(\xi(y_n^S))}(x_n^S), & j = 0, \\ q_1^{(\xi(y_n^S))}(x_n^S), & j = \xi(y_n^S), \\ 0, & j \neq 0, j \neq \xi(y_n^S). \end{cases} \quad (7)$$

其中,  $j \in \{0, \dots, M\}$ ,  $y_n^S$ 是与 $x_n^S$ 相对应的类别标签, 对于任意的一个 $y_n^S \in \{0, \dots, M\}$ , 如果 $y_n^S = 0$ , 定义0表示健康类别, 那么函数 $\xi(y_n^S)$ 的返回值为集合 $\{1, \dots, M\}$ 中的任意一个元素, 且满足每个元素被选中的概率相等; 否则, 函数 $\xi(y_n^S)$ 直接返回 $y_n^S$ .

学生模型是在教师模型的基础上进行简化的一个模型, 模型共有6层 $l_1, \dots, l_6$ , 第 $l_1$ 层是有6个滤波器的卷积层, 卷积核大小为(5, 5), 步幅为(4, 2). 第 $l_2$ 层

同样是卷积层, 包含18个滤波器, 卷积核大小、步幅与 $l_1$ 相同. 第 $l_3$ 层是一个最大池化层, 其pool size为(2, 2). 第 $l_4$ 卷积层有24个滤波器, 卷积核大小与 $l_2$ 相同. 第 $l_5$ 层全连接层有32个隐藏单元, 然后连接输出层.

学生模型的软标签计算与教师模型类似, 区别在于学生模型需要考虑 $M + 1$ 个类别, 如公式8所示:

$$p_j(x_n^S) = \frac{\exp(g_j(x_n^S)/t)}{\sum_{j=0}^M \exp(g_j(x_n^S)/t)}. \quad (8)$$

其中,  $g_j(\cdot)$ 表示学生模型对应类别 $j \in \{0, \dots, M\}$ 时logits层的输出, 0表示健康类别.

除此之外, 学生模型的学习目标和教师模型也不同, 学生模型有两个学习目标, 一方面需要使得输出的预测 $f^S(x_n^S)$ 尽量接近真实标签 $y_n^S$ , 另外一方面需要使得自身的软标签 $p_j(x_n^S)$ 接近教师模型的软标签 $\pi_j(x_n^S, y_n^S)$ . 与真实标签相比, 软标签 $\pi_j(x_n^S, y_n^S)$ 具有更为丰富的信息, 其包含样本中类与类之间的关联, 这些关联由复杂的教师模型学习得到. 学生模型由于自身结构比较简单, 对样本的学习能力比教师模型弱, 但如果在训练过程中, 让学生模型拟合教师模型学习到的关联信息, 能够对自身的学习起到辅助作用. 然而, 教师模型也是存在一定的出错概率, 与仅向教师模型学习相比, 添加真实标签可以减少学生模型学习错误知识的可能性. 另外, 模型最后的输出相比软标签更为离散, 各个类别之间的关联性不强, 因而本文让学生模型的软标签接近教师模型的软标签, 而不是让学生模型的输出接近教师模型的软标签. 学生模型的损失如公式9所示:

$$Loss_S = -\frac{1}{|X^S|} \sum_{n=1}^{|X^S|} \left\{ \lambda y_n^S \log f^S(x_n^S) + (1 - \lambda) \sum_{j=0}^M \pi_j(x_n^S, y_n^S) \log(p_j(x_n^S)) \right\}. \quad (9)$$

这里的 $|X^S|$ 表示 $X^S$ 中元素数量,  $f^S$ 表示 $S$ 的预测函数,  $\lambda \in [0, 1]$ 是一个常数, 用于学生模型选择侧重真实标签, 还是侧重软标签.

## 5 实验结果与分析

本文实验分为两个部分, 第一部分为归一化方法评估实验, 用于验证不同的数据归一化方法, 对肺音音频数据的梅尔频谱特征的作用效果; 第二部分为集成知识蒸馏评估实验, 验证精简的学生模型在集成多个性能优异的教师模型的知识后, 对其性能的提升作用.

### 5.1 数据集描述

实验数据集使用的是文

献[24]的公开数据集, 该数据集包含8类音频: COPD, Healthy, URTI, Bronchiectasis, Pneumonia, Bronchiolitis, Asthma和Lower Respiratory Tract Infection (LRTI), 其中Asthma与LRTI这两类数据量很小, 不具有统计意义和实验价值, 因此只选用了其他6种类型的音频用于实验. 这6类肺音原有音频文件共917个, 每个音频文件采集时长为20s, 在对其中的5种类型按照本文阐述的数据增强方式进行增强之后得到音频文件2653个, 每种类型的音频数量如表1所示.

实验中将没有进行数据增强的原始数据随机打乱, 并分成70%的训练数据 $Train_1$ 和30%的测试数据 $Test$ . 而进行过增强的数据 $Train_2$ 与 $Train_1$ 共同作为数据增强后的训练数据. 实验中, 梅尔频谱特征的抽取基于一个开源的python包librosa<sup>[25]</sup>实现, 抽取出来的特征矩阵将被重塑为模型的输入维度(128, 128, 1).

表 1 原始数据量和增强后数据量  
Table 1 Original and augmented data size

类型	原始数据量	增强后数据量
COPD	793	793
Healthy	35	525
URTI	23	345
Bronchiectasis	16	240
Pneumonia	37	555
Bronchiolitis	13	195
合计	917	2653

### 5.2 归一化方法评估实验

该实验使用的模型是将教师模型结构用于多分类问题上. 模型训练使用单个Tesla P100 GPU, 批处理大小设置为128, epoch设置为20. 该实验对比了原始数据(Original), 进行数据增强(Augmented), 数据增强并且使用峰值归一化处理(Augmented Peak), 数据增强并使用均方根归一化处理(Augmented RMS), 数据增强并使用EBU R128归一化处理(Augmented EBU), 数据增强并使用零均值归一化处理(Augmented Zmean)这6组实验. 表2为对比实验的结果, 使用的评估指标有准确率(Accuracy, 表中Acc), 精确率(Precision, 表中Pre), 召回率(Recall, 表中Rec), F1分数(F1-score, 表中F1).

据表2所示, 通过对原始数据进行零均值归一化处理之后, 模型的准确率, 精确率, 召回率, F1分数均达到了98%, 这说明该归一化方式比其他3种方法更加适合处理肺音音频的梅尔频谱特征数据, 对模型性能的提高作用比较明显. 在使用Peak归一化和RMS归一化对数据进行处理之后, 各个评价指标有所下降, 而使用EBU R128归一化方法进行处理也没有明显的

提升效果. 另外, 从分类原始数据和分类增强后的数据来看, 通过对原始数据进行适当的数据增强能够较好的提高模型的性能.

表2 归一化方法评估实验

Table 2 Evaluation experiment of normalization method

方法	Acc	Pre	Rec	F1
Original	0.64	0.73	0.64	0.63
Augmented	0.93	0.94	0.93	0.93
Augmented Peak	0.87	0.90	0.87	0.87
Augmented RMS	0.91	0.92	0.91	0.91
Augmented EBU	0.93	0.94	0.93	0.93
<b>Augmented Zmean</b>	<b>0.98</b>	<b>0.98</b>	<b>0.98</b>	<b>0.98</b>

### 5.3 集成知识蒸馏评估实验

该实验从 $Train_1 + Train_2$ 中随机抽取第 $i$ 类疾病数据和健康数据的70%作为 $X^T$ , 剩下30%的数据作为 $X^S$ , 数据均使用零均值归一化处理, 批处理大小为128.

表3为5个二分类教师模型在测试集 $Test$ 上的评估结果,  $T_1$ 至 $T_5$ 依次表示对应疾病Bronchiectasis, Bronchiolitis, COPD, Pneumonia, URTI的教师模型. 该实验结果显示, 每个二分类教师模型在各个评价指标上均达到99%(或99%以上), 表现优异.

表3 教师模型实验结果

Table 3 Evaluation experiment of teacher model

模型	Acc	Pre	Rec	F1
$T_1$	1.00	1.00	1.00	1.00
$T_2$	1.00	1.00	1.00	1.00
$T_3$	0.99	0.99	0.99	0.99
$T_4$	1.00	1.00	1.00	1.00
$T_5$	1.00	1.00	1.00	1.00

而在训练学生模型时, 对于参数的选择, 本文使用网格搜索<sup>[26]</sup>的方法, 设置epoch的搜索范围为{25, 50, 75, 100}, 温度参数 $t$ 的搜索范围为{2.0, 5.0, 7.0, 10.0},  $\lambda$ 的搜索范围为{0.1, 0.3, 0.5, 0.7, 0.9}, 该算法会将所有参数的组合逐一枚举验证, 得出最佳参数. 依据网格搜索的结果, 实验将epoch设置为50,  $t$ 设置为2.0,  $\lambda$ 设置为0.5. 在模型的对比上, 除了测试本文的几个模型外, 还对比了最近的几个轻量化的模型: Shufflenet-v2<sup>[27]</sup>, Mobilenet-v3-large和Mobilenet-v3-small<sup>[28]</sup>, 这几个模型都是专门为小型设备设计的高效神经网络模型.

表4为集成知识蒸馏的实验结果, 其中评价指标"Par"表示参数量(parameters), 数量级为 $10^6$ , "Time"表示模型预测耗时, 单位为

秒, Net1为Shufflenet-v2模型, Net2为Mobilenet-v3-large模型, Net3为Mobilenet-v3-small模型,  $T$ -Mul为零均值评估实验中的多分类模型,  $S$ -No为没有进行知识蒸馏, 只使用真实标签训练的多分类学生模型,  $S$ -Dis为本文进行蒸馏的学生模型, 同样在准确率, 精确率, 召回率, F1分数4个指标上对各个模型进行了评估. 相比模型 $T$ -Mul, 学生模型 $S$ -Dis在各项指标上仅差0.03、0.02、0.03、0.03, 而模型 $S$ -No在各指标上比模型 $T$ -Mul均差8%. 该实验结果表明, 将性能好的教师模型的知识集成到精简的学生模型上, 能够在一定程度上提高学生模型的分分类能力.

表4 集成知识蒸馏实验结果

Table 4 Ensemble knowledge distillation experiment

模型	Acc	Pre	Rec	F1	Par	Time
Net1 <sup>[27]</sup>	0.95	0.95	0.96	0.95	4.02	98.08
Net2 <sup>[28]</sup>	0.94	0.94	0.95	0.94	4.23	140.51
Net3 <sup>[28]</sup>	0.93	0.93	0.94	0.93	1.54	41.46
$T$ -Mul <sup>[15]</sup>	0.98	0.98	0.98	0.98	0.28	3.12
$S$ -No	0.90	0.90	0.90	0.90	0.06	2.26
<b><math>S</math>-Dis</b>	<b>0.95</b>	<b>0.96</b>	<b>0.95</b>	<b>0.95</b>	<b>0.06</b>	<b>2.48</b>

为了更加全面的评估精简后的学生模型, 本文还测试了模型的预测速度. 实验中使用CPU单核处理, 对比了各个模型在同等条件下预测8000条数据的耗时. 经简化之后, 学生模型 $S$ -Dis的参数仅有0.06M, 比模型 $T$ -Mul的参数减少了79%, 耗时为2.48s, 比模型 $T$ -Mul节省了20%时间. 虽然Net1、Net2和Net3在准确率上和 $S$ -Dis相近, 但是 $S$ -Dis的参数量, 仅有Net1和Net2的1%, Net3的3%. 进行相同的预测,  $S$ -Dis的耗时仅有Net1的3%, Net2的2%, Net3的6%.

学生模型 $S$ -Dis虽然在准确率, 精确率, 召回率, F1分数4个指标上比模型 $T$ -Mul略差一点, 但是其模型更加精简, 分类预测的速度有着显著的提升, 这也更加符合肺疾病实时性检测的需求.

## 6 结论

本文提出集成知识蒸馏的肺疾病分类算法, 在现有文献对肺疾病进行简单分类的基础上, 评估了零均值归一化方法在处理肺音数据上表现. 同时, 通过多个教师模型学习多方数据, 在保护医疗数据隐私的前提下, 间接融合了多方数据的知识, 由教师模型集成到学生模型上, 有效的提高了学生模型的准确率, 解决了小数据集带来的泛化问题. 而蒸馏得到的学生模型是一个比教师模型更为精简的模型, 参数量减少了79%, 预测速度比教师模型提升了20%, 能够更快的对肺疾病做出分析, 其预测准确率达95%, 在基本接近教师模型准确率的前提下, 能更好的满足实际场景中实时诊断的需求.

## 参考文献:

- [1] YANG Yukun, ZHU Xiangqing, RUAN Guangping, et al. Association of aging with pulmonary fibrosis and chronic obstructive pulmonary disease and treatment with mesenchymal stem cells. *Chinese Journal of Tissue Engineering Research*, 2019, 23(17): 2746-2752. (杨育坤, 朱向情, 阮光萍, 等. 衰老与肺纤维化和慢性阻塞性肺疾病以及间充质干细胞的治疗作用. *中国组织工程研究*, 2019, 23(17): 2746-2752.)
- [2] ZHANG Ran, QIN Yi, GAO Dongping, et al. Analysis of the Changes in the Main Causes of Death of Chinese Residents in the Last 70 Years. *Journal of Medical Informatics*, 2019, 40(8): 9-14. (张冉, 秦奕, 高东平, 等. 近70年我国居民主要死因变化情况分析. *医学信息学杂志*, 2019, 40(8): 9-14.)
- [3] ABBAS A, FAHIM A. An automated computerized auscultation and diagnostic system for pulmonary diseases. *Journal of Medical Systems*, 2009, 34(6): 1149-1155.
- [4] SENGUPTA N, SAHIDULLAH M, SAHA G. Lung sound classification using cepstral-based statistical features. *Computers in Biology and Medicine*, 2016, 75(8): 118-129.
- [5] TAPLIDOU S A, HADJILEONTIADIS L J. Wheeze detection based on time-frequency analysis of breath sounds. *Computers in Biology and Medicine*, 2017, 37(8): 1073-1083.
- [6] BAKIBILLAH A S M, HASAN M, RAHMAN M M. et al. Predictive car-following scheme for improving traffic flows on urban road networks. *Control Theory and Technology*, 2019, 17(4): 325 - 334.
- [7] XIANG G, HUANG Y, YU J. et al. Intelligence evolution for service robot: An ADRC perspective. *Control Theory and Technology*, 2018, 16(4): 324 - 335.
- [8] PALANIAPPAN R, SUNDARAJ K, AHAMED N U. Machine learning in lung sound analysis: A systematic review. *Biocybernetics and Biomedical Engineering*, 2013, 33(3): 129-135.
- [9] KANDASWAMY A, KUMAR C S, RAMANATHAN R P, et al. Neural classification of lung sounds using wavelet coefficients. *Computers in Biology and Medicine*, 2004, 34(6): 523-537.
- [10] YAMASHITA M, MATSUNAGA S, MIYAHARA S. Discrimination between healthy subjects and patients with pulmonary emphysema by detection of abnormal respiration. *International Conference on Acoustics, Speech and Signal Processing*. Prague: IEEE, 2011: 693-696.
- [11] DATTA S, CHOUDHURY A D, DESHPANDE P, et al. Automated lung sound analysis for detecting pulmonary abnormalities. *39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Seogwipo: IEEE, 2017: 594-4598.
- [12] CHAMBERLAIN D, KODGULE R, GANELIN D, et al. Application of semi-supervised deep learning to lung sound analysis. *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Orlando: IEEE, 2016: 804-807.
- [13] AYKANAT M, KLC O, KURT B, et al. Classification of lung sounds using convolutional neural networks. *EURASIP Journal on Image and Video Processing*, 2017, 2017(1): 1-9.
- [14] FALAH A H, JONDRI J. Lung sounds classification using stacked autoencoder and support vector machine. *7th International Conference on Information and Communication Technology (ICoICT)*. Kuala Lumpur: IEEE, 2019: 1-5.
- [15] TARIQ Z, SHAH S K, LEE Y. Lung disease classification using deep convolutional neural network. *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. San Diego: IEEE, 2019: 732-735.
- [16] ZHANG X, ZHOU X, LIN M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices. *Conference on Computer Vision and Pattern Recognition, CVPR*. Salt Lake: IEEE, 2018: 6848-6856.
- [17] SHIN B, YANG H, CHOI J D. The pupil has become the master: Teacher-student model-based word embedding distillation with ensemble learning. *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*. Macao: IJCAI, 2019: 3439-3445.
- [18] SKLAR B. *Digital Communications: Fundamentals and Applications*. Prentice-Hall, Inc., New Jersey, USA, 1988.
- [19] CHI Z, LI Y, CHEN C. Deep convolutional neural network combined with concatenated spectrogram for environmental sound classification. *7th International Conference on Computer Science and Network Technology (ICCSNT)*. Dalian: IEEE, 2019: 251-254.
- [20] PHAYE S S R, BENETOS E, WANG Y. Subspectralnet using subspectrogram based convolutional neural networks for acoustic scene classification. *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Brighton: IEEE, 2019: 825-829.
- [21] GURS J. *Data Science from Scratch: First Principles with Python*. O'Reilly Media, Sebastopol, CA. 2015.
- [22] HU Z, MA X, LIU Z, et al. Harnessing deep neural networks with logic rules. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. Berlin: ACL, 2016: 2410-2420.
- [23] HINTON G E, VINYALS O, DEAN J. Distilling the knowledge in a neural network. Montreal: *NIPS 2014 Deep Learning Workshop*, 2015(2015-03-09)[2020-02-30]. <https://arxiv.org/abs/1503.02531>.
- [24] ROCHA B, FILOS D, MENDES L. et al. A respiratory sound database for the development of automated classification. *Precision Medicine Powered by pHealth and Connected Health*, 2018, 66(1): 33-37.
- [25] MCFEE B, RAFFEL C, LIANG D, et al. librosa: Audio and Music Signal Analysis in Python. In *Proceedings of the 14th Python in Science Conference (2015)*. Texas: SciPy, 2015: 18-24.
- [26] LAVALLE S M, BRANICKY M S, LINDEMANN S R. On the relationship between classical grid search and probabilistic roadmaps. *The International Journal of Robotics Research*, 2004, 23(7): 673-692.
- [27] MA N, ZHANG X, ZHENG H, et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design. *ECCV 2018*. Munich: Springer, 2018: 122-138.
- [28] HOWARD A, SANDLER M, CHEN B, et al. Searching for mobilenetv3. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul: IEEE, 2019: 1314-1324.

## 作者简介:

李坚强 博士, 教授, 博士生导师, 目前研究方向为人工智能、物联网、移动医疗、复杂系统等, E-mail: lijq@szu.edu.cn;

王成 硕士研究生, 目前研究方向为移动医疗, E-mail: wangcheng2019@email.szu.edu.cn;

黄志超 硕士研究生, 目前研究方向为医疗影像分析, E-mail: 2170276044@email.szu.edu.cn;

陈杰 博士, 副研究员, 目前研究方向为计算性能可扩展、可自动学习、可解释以及隐私数据可受保护的机器学习, E-mail: chenjie@szu.edu.cn.