

本文引用格式: 林家荣,郑凯文,甘兆明,等.基于知识蒸馏的心肺音分离模型[J].自动化与信息工程,2022,43(5):13-16,29.

LIN Jiarong, ZHENG Kaiwen, GAN Zhaoming, et al. Cardiorespiratory sound separation method based on knowledge distillation[J]. Automation & Information Engineering, 2022,43(5):13-16,29.

基于知识蒸馏的心肺音分离模型

林家荣 郑凯文 甘兆明 谢仕宇 郑君锐

(广东工业大学, 广东 广州 510006)

摘要: 针对基于双向门控循环单元(BiGRU)的心肺音分离模型存在的参数量大、算法繁杂、训练成本高、硬件算力要求高等问题,提出基于知识蒸馏的心肺音分离模型。该模型基于 BiGRU,教师网络、学生网络分别采用 3 层、1 层 BiGRU 网络。实验结果表明,采用“教师—学生”知识蒸馏方法可提高学生网络性能,且蒸馏后的“学生+KD”心肺音分离网络相比教师网络,模型更小、算法更简单、训练成本更低,为部署到边缘设备以及资源受限的设备提供了理论依据。

关键词: 心肺音分离;双向门控循环单元;知识蒸馏;教师网络;学生网络

中图分类号: R318; TN912.3; TP183 文献标识码: A 文章编号: 1674-2605(2022)05-0003-05

DOI: 10.3969/j.issn.1674-2605.2022.05.003

Cardiorespiratory Sound Separation Method Based on Knowledge Distillation

LIN Jiarong ZHENG Kaiwen GAN Zhaoming XIE Shiyu ZHENG Junrui

(Guangdong University of Technology, Guangzhou 510006, China)

Abstract: In order to address the problems of the existing cardiorespiratory sound separation model based on bidirectional gated recurrent unit, such as huge parameters, complex algorithms, high training costs and high requirements for hardware computing power, this paper applies the knowledge distillation method to the cardiorespiratory sound separation, and proposes a cardiorespiratory sound separation model based on knowledge distillation. The model is based on BiGRU. The teacher network uses three-layer BiGRU network, and the student network uses one-layer BiGRU network. The experimental results show that the performance of student network can be improved by using the "teacher student" knowledge distillation method. The distilled "student +KD" cardiopulmonary sound separation network is smaller than the whole teacher model, with simpler algorithm and lower training cost, which provides a theoretical basis for deploying to edge devices and resource constrained devices.

Keywords: cardiorespiratory sound separation; bidirectional gated recurrent unit; knowledge distillation; teacher network; student network

0 引言

心音和肺音可用于心血管、呼吸系统和睡眠呼吸暂停综合征等疾病的诊断^[1-3]。然而现实采集的心音和肺音是混合信号,需进行分离,以准确识别相关疾病。

心音信号的频率范围一般为 10~320 Hz,肺音信号的频率范围一般为 60~600 Hz,其频率范围相互重叠。传统的带通滤波器仅适用于分离频带相互分离的信号,无法有效分离心音和肺音;自适应滤波法、小

波变换法依赖模板和参数,难以在实际应用中推广。

近几年,国内外学者虽然提出了很多心音信号和肺音信号的分离方法^[1-4],但这些方法都有一定的缺陷,如自适应滤波法收敛速度较慢;小波变换法难以选取小波基,不同的小波基分析结果也不相同等。随着深度学习的发展,采用深度学习网络模型分离心肺音的研究也越来越多,如长短时记忆网络(long short-term memory, LSTM)、门控循环单元(gated recurrent unit,

GRU)、双向门控循环单元(bidirectional gated recurrent unit, BiGRU)等循环神经网络(recurrent neural network, RNN)的变形,因可以较好地处理序列问题^[4]而应用于心肺音分离。

RNN 随着训练次数、层数增加,其参数呈指数增长,计算时间也相应增加,这对边缘设备的硬件要求较高。

为此,本文提出基于知识蒸馏的心肺音分离模型。该模型的基本框架为 BiGRU,将知识蒸馏应用于该模型,使模型更小、算法更简单、训练成本更低。

1 心肺音混合数学模型

本文研究的心肺音分离模型仅限于单通道心肺音信号,其混合数学模型为^[5]

$$x(m) = \sum_{\pi} x_{\pi}(m) + \eta(m) \quad (1)$$

式中:

x ——离散的心肺音混合信号;

m ——时间采样点, $m = 1, 2, \dots, M$;

x_{π} ——源信号, $x \in \{c, r\}$, c 和 r 分别表示心音信号、肺音信号;

η ——高斯白噪声。

为简单起见,假设只考虑安静无噪声环境下的心肺音混合信号,即仅含心音信号和肺音信号,公式(1)可简化为

$$x(m) = x_c(m) + x_r(m) \quad (2)$$

2 BiGRU

基于知识蒸馏的心肺音分离模型采用了 BiGRU 网络结构。BiGRU 主要由输入层、前向传递层、后向传递层和输出层组成^[6],结构如图 1 所示。

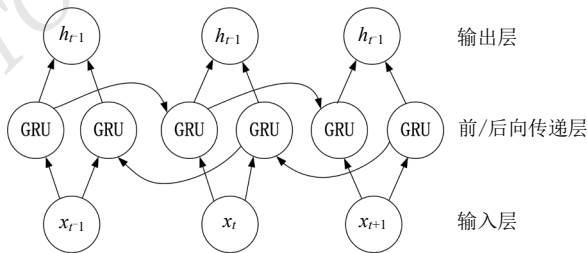


图 1 BiGRU 结构

BiGRU 的当前状态由当前时刻的输入、前一时刻传递层正向状态的输出以及反向状态的输出决定。 t 时刻 BiGRU 的状态 h^t 通过前向传递层正向状态 $\tilde{h}^t$ 和传递层反向状态 $\bar{h}^t$ 的加权求和获得,如公式(3)~(5)所示。

$$\tilde{h}^t = GRU(x_t, \tilde{h}^{t-1}) \quad (3)$$

$$\bar{h}^t = GRU(x_t, \bar{h}^{t-1}) \quad (4)$$

$$h^t = w^t \tilde{h}^t + v^t \bar{h}^t + b^t \quad (5)$$

式中:

x_t —— t 时刻的特征输入;

w 、 v ——网络权重;

b ——网络偏差。

在心肺音混合信号中,某个时刻的信息与历史信息、未来信息均有关联。单向的 GRU 网络只能感知按时间顺序传递的信息,无法反向获取未来信息与当前信息的关联。BiGRU 通过训练 2 个方向相反的单向 GRU 网络,既考虑数据时间正序、逆序的关联,又可充分提取心肺音混合信号的深层特征信息。BiGRU 在结构上简化为 2 个门:更新门和重置门,较少的参数使模型训练更快收敛。

3 基于知识蒸馏的心肺音分离方法

3.1 心肺音分离模型的知识蒸馏框架

知识蒸馏是一种网络轻量化方法,由多伦多大学的 HINTON 提出^[7],其核心思想是通过知识迁移,教师网络采用训练好的大网络,学生网络采用运行速度更快、参数更小的网络;将复杂的教师网络输出的软标签作为知识传递给学生网络,提升学生网络性能。本文心肺音分离网络的知识蒸馏框架如图 2 所示,其中 Teacher Model 表示教师网络,Student Model 表示学生网络。

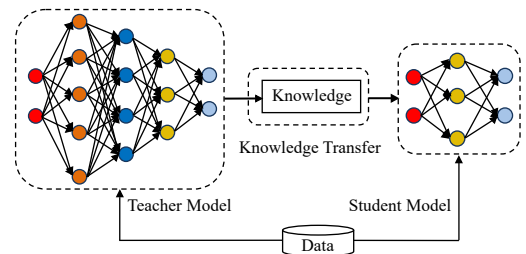


图 2 知识蒸馏框架

知识蒸馏网络的损失函数为蒸馏损失函数和学生网络在真实标签监督下的损失函数的加权和, 计算公式为

$$L_{\text{total}}(\theta) = \alpha L_{CE}(\theta) + (1 - \alpha) L_{KD}(\theta) \quad (6)$$

式中:

L_{CE} ——在真实标签监督下的损失;

L_{KD} ——蒸馏损失;

θ ——学生模型;

α ——加权系数。

知识蒸馏训练框架和学生网络的测试框架如图 3 所示。利用教师网络和学生网络同时训练, 训练出分离效果最好、总损失最小的学生网络, 并使用测试数据集进行测试。

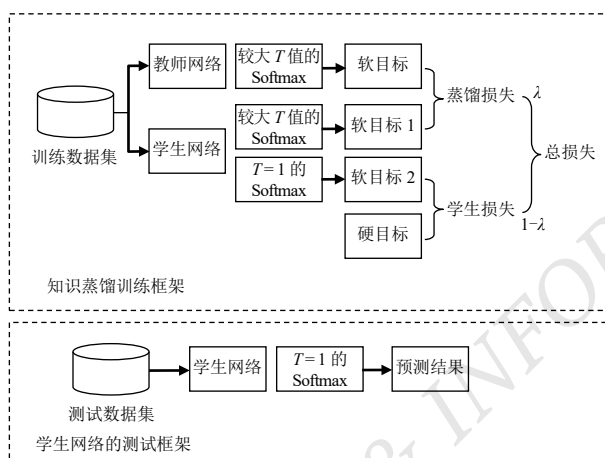


图 3 知识蒸馏训练框架和学生网络的测试框架

3.2 教师网络与学生网络

目前, 主流的基于深度学习的心肺音分离模型训练步骤为: 首先, 利用短时傅里叶变换 (short-time Fourier transform, STFT) 将心肺音混合信号转换成时频谱; 然后, 经过分离网络得到心音和肺音的时频掩码; 最后, 利用逆短时傅里叶变换乘以时频掩码, 得到分离后的心音信号和肺音信号^[5]。

本文教师网络框架如图 4 所示, 以 BiGRU 为主体架构, 基本训练步骤与主流的基于深度学习的心肺音分离模型一致。

学生网络性能依赖于教师网络。因此, 在利用知

识蒸馏训练学生网络时, 先将教师网络训练到最好效果。为尽可能地压缩学生网络规模, 本文设计的学生网络在教师网络的基础上减少 2 层 BiGRU, 从而减少网络的规模和参数数量, 其余部分与图 4 相同。

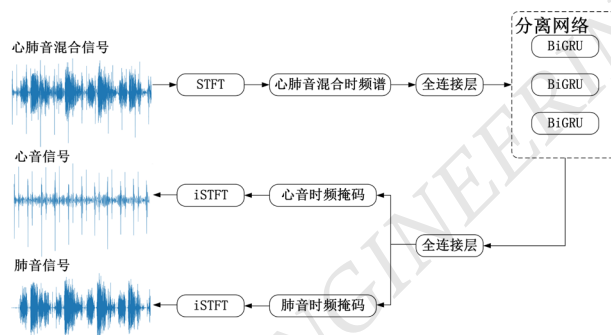


图 4 教师网络框架图

3.3 代价函数

心肺音分离的效果可通过分离后的心音信号和肺音信号的信噪比来衡量, 信噪比越高说明分离效果越好^[8]。本文取心音信号和肺音信号的信噪比之和, 并取负数作为损失函数, 则心肺音分离网络的代价函数为

$$a = \alpha \lg \frac{\|y_{\pi}\|_2^2}{\|y_{\pi} - \hat{y}_{\pi}\|_2^2} \quad (7)$$

$$b = (1 - \alpha) \lg \frac{\|y_T\|_2^2}{\|y_T - \hat{y}_{\pi}\|_2^2} \quad (8)$$

$$\min J(\omega) = - \sum_{\pi \in \{c, r\}} (a + b) \quad (9)$$

式中:

ω ——网络所有参数;

$y_{\pi} \in R^{1 \times L}$ ——干净的心音信号或肺音信号;

$\hat{y}_{\pi} \in R^{1 \times L}$ ——模型估计的心音信号或肺音信号;

$y_T \in R^{1 \times L}$ ——教师网络分离的心音信号和肺音信号;

$\hat{y}_{\pi} = f(\omega; x)$, $f(\cdot)$ ——分离模型;

L ——采样点个数;

x ——混合的时域信号。

其中, 公式(7)为学生网络在真实标签监督下的心

肺音信噪比;公式(8)为学生网络在教师网络伪标签监督下的心肺音信噪比。

3.4 训练流程

基于知识蒸馏的心肺音分离模型的训练流程为:

- 1) 搭建教师网络和学生网络,教师网络复杂、参数量较大;学生网络简单、参数量少;
- 2) 在公开数据集上训练教师网络,得到一个效果最好的网络;
- 3) 将教师网络分离后的心音信号和肺音信号作为伪标签,让学生网络在真实标签和伪标签的监督下训练,得到效果最好的学生网络。

4 实验和结果分析

4.1 实验设置

实验采用的心音信号和肺音信号分别来自公开数据集 PhysioNet 和 Welch Allyn^[9-10]。其中,心音信号 102 条,肺音信号 95 条,信号时长为 2~70 s,信号采样率为 4 kHz 或 44.1 kHz。为方便计算,统一对信号进行降采样,将信号频率降至 2 kHz,并按照固定的能力比 0 dB 和 10 dB 进行混合;以信噪比(signal-to-noise ratio, SNR)作为心肺音分离实验的评价指标。

$$SNR = 10 \lg \frac{\|y\|_2^2}{\|y - \hat{y}\|_2^2} \quad (10)$$

式中:

$\hat{y}$ ——模型分离的心音信号或肺音信号的时域信号;

y ——混合前的心音信号和肺音信号的时域信号。

教师网络设置 3 层 BiGRU,每一层有 128 个传递层单元。第一层的全连接层含有 33 个神经元,最后一层的全连接层神经元也为 33 个。学生网络设置 1 层 BiGRU,含有 64 个传递层单元。学生网络的全连接层也设置了前置全连接层和后置全连接层,与教师网络一致,包含 33 个神经元。在训练中,设置 batchsize 为 64, epoch 为 100 个,前 60 个 epoch 的学习率为 5×10^{-4} ,后 40 个 epoch 的学习率为 5×10^{-5} ,采用 Adam 算法优化。

4.2 实验结果及分析

为验证知识蒸馏方法对心肺音分离网络的有效性,对比基于 NMF 模型的心肺音分离网络、基于全连接 LSTM 的心肺音分离网络、教师网络、学生网络、学生+KD 网络(学生+KD 网络为采用知识蒸馏结构训练的学生网络)的分离效果,其信噪比如表 1 所示。

表 1 不同分离网络分离出心音的信噪比

分离方法	0 dB	10 dB
NMF	12.13	9.34
LSTM	15.46	12.32
教师网络	15.83	12.69
学生网络	12.92	10.48
学生+KD	13.42	10.69

由表 1 可以看出,基于 LSTM 的分离方法的分离效果比基于 NMF 的分离方法好;本文采用的 BiGRU 网络(即教师网络)的分离效果比基于 LSTM 的分离方法高约 0.4 dB,说明采用 BiGRU 网络不仅减少了网络参数,还提高了分离效果。

学生网络的网络参数比教师网络少,SNR 与教师网络相比下降约 2 dB,分离效果仍然优于基于 NMF 的分离方法。

经过知识蒸馏后,教师网络将知识迁移到学生网络,虽然学生+KD 网络分离效果比教师网络差,却比普通的学生网络提高了约 0.35 dB,说明知识蒸馏方法可以提高学生网络的心肺音分离精度。

5 结语

本文针对现有的心肺音分离模型参数量较大、算法繁杂、训练成本高等问题,将知识蒸馏方法应用于心肺音分离,提出基于知识蒸馏的心肺音分离模型。通过实验表明,使用深度学习的分离方法比传统的 NMF 分离方法效果好,且使用“教师—学生”知识蒸馏方法可提高学生网络性能。蒸馏后的“学生+KD”心肺音分离网络相比教师网络虽然 SNR 减小,但整个模型更小,算法更简单、训练成本更低,为部署到边缘设备、资源受限设备提供了可能。

(下转第 29 页)