

基于跨语种声学分析的帕金森病检测方法

季薇^① 王传瑜^① 吴迪^① 李云^{*②} 郑慧芬^③

^①(南京邮电大学通信与信息工程学院 南京 210003)

^②(南京邮电大学计算机学院 南京 210023)

^③(南京医科大学附属老年医院 南京 210024)

摘要: 基于语音的帕金森病检测具有非介入式、成本较低和无创等优点。当前公开的帕金森病语音数据集大多来源于单一语种,存在数据容量不够大、受试者母语发音特点差异小等特点。单一语种数据集上训练的帕金森病检测模型在面对跨语种语音数据时,将出现性能下降。为避免语种差异带来的影响,提升模型在跨语种场景下的检测性能,该文引入对抗迁移学习和特征解耦的思想,提出一种帕金森病跨语种声学分析模型(CLSAM)。首先,将基于多头自注意力机制的Transformer编码块和多层神经网络级联,组成特征提取器模块,用于将从源域和目标域语音中提取的原始Fbank语音特征初步解耦为两个向量,即域不变病理信息表征向量和域信息表征向量;设计了目标任务不一致的双重对抗训练模块,显式地分离域不变病理信息和域信息;最终,提取跨语种语音数据中的域不变病理信息用于帕金森病检测。该文在公开的MaxLittle帕金森病语音数据集以及自采的帕金森病语音数据集上,采用十折交叉验证的方法验证了所提方法的有效性。实验结果表明:与传统机器学习方法以及现有的迁移学习算法相比,所提模型在跨语种场景中的检测准确率、敏感度和F1分数等性能均有明显提升。

关键词: 跨语种声学分析; 帕金森病; 对抗迁移学习; 特征解耦

中图分类号: TN911.7; TP391.4

文献标识码: A

文章编号: 1009-5896(2024)02-0546-09

DOI: 10.11999/JEIT230981

Parkinson's Disease Detection Method Based on Cross-Language Acoustic Analysis

JI Wei^① WANG Chuanyu^① WU Di^① LI Yun^{*②} ZHENG Huifen^③

^①(School of Communication and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

^②(School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China)

^③(Affiliated Geriatric Hospital of Nanjing Medical University, Nanjing 210024, China)

Abstract: The research on speech-based Parkinson's disease detection has the advantages of non-intrusive, low cost and non-invasive. The current publicly available speech datasets for Parkinson's disease mostly originate from single-language speech, which has the characteristics such as insufficient data capacity and small differences in the pronunciation characteristics of the subjects' mother tongue. The Parkinson's disease detection model trained on a single language dataset will experience performance degradation when faced with cross-language speech data. To avoid the impact of language differences and improve the detection performance of the model in cross-language scenarios, the ideas of adversarial transfer learning and feature decoupling is introduced and a Parkinson's disease Cross-Language Speech Analysis Model (CLSAM) is proposed in this paper. Firstly, the model cascades a multihead self-attention encoder and a multi-layer neural network to form a feature extractor module, which is used to decouple the original Fbank speech features extracted from the pronunciation characteristics of the source domain and target domain into two vectors, namely domain invariant pathological information representation vector and domain information representation vector.

收稿日期: 2023-09-07; 改回日期: 2023-12-04; 网络出版: 2023-12-13

*通信作者: 李云 liyun@njupt.edu.cn

基金项目: 江苏省高校基础科学(自然科学)重大项目(21KJA520003)

Foundation Item: The Basic Scientific (Natural Science) Major Program of the Higher Education Institutions of Jiangsu Province, China (21KJA520003)

Secondly, a dual adversarial training module with inconsistent target tasks is designed, which explicitly separates domain invariant pathological information and domain information. Finally, domain invariant pathological information is extracted from cross-language speech data for Parkinson's disease detection. This paper verifies the effectiveness of the proposed method using a ten-fold cross-validation method on both the publicly available MaxLittle Parkinson's disease speech dataset and the self-collected Parkinson's disease speech dataset. Experimental results show that compared with traditional machine learning methods and existing transfer learning algorithms, the proposed model significantly improves the accuracy, sensitivity and F1 scores in cross-language scenarios.

Key words: Cross-language speech analysis; Parkinson's disease; Adversarial transfer learning; Feature decoupling

1 引言

帕金森病(Parkinson's disease)患者往往患有不同程度的构音障碍^[1,2]。持续元音能够很好地反映受试者声带和声道中各种肌肉的协调能力,从而有效评估受试者的发音能力^[3,4]。然而,帕金森病患者的语音数据采集不易,标注成本高昂。已公开的帕金森病语音数据集大多来源于单一地域人群的语音,且存在数据容量不够大、受试者母语发音特点差异小等特点^[5]。当模型需要进行跨语种场景下的帕金森病检测时,由于采集环境的不同、受试者母语的差异,将导致不同语种的语音数据集服从不同的统计概率分布^[6],最终导致在单一语种的语音数据库训练的模型无法以同样高的准确率识别另一个语种的语音数据^[7]。

当前跨语种场景下的帕金森病检测研究还处于探索阶段。文献^[8]在韩语和美式英语的数据集上的实验发现:使用韩语和美式英语的帕金森病患者的语音在元音空间区域这一声学指标上存在差异。文献^[9]研究发现与健康的日语使用者相比,患有帕金森病的日语使用者的语速较慢。文献^[10]使用西班牙语、德语和捷克语数据集进行跨语种实验,通过在训练过程添加不同比例的多语种数据帮助模型学习跨语种分析的能力,结果发现越高比例的多语种语音数据加入到模型的训练过程中,越能够实现更加优秀的性能。文献^[11]提出了一种英语、韩语和泰米尔语的跨语种帕金森病检测方法,该方法充分考量了语言差异带来的影响,实现了模型在每种语言语音数据上检测准确率的提升。尽管已有一些研究者开始探索帕金森病的跨语种语音分析,然而结合汉语语音数据集的跨语种分析还很少,还需进一步借助人工智能技术挖掘能够独立于语言的声学特征,以实现跨语种的帕金森病检测^[12]。

迁移学习技术有助于解决传统有监督学习无法适应数据分布差异的问题^[13]。它将源域数据和目标域的数据分别映射至公共特征空间,并进行特征分布差异度量,最小化这一差异,即可使得源域数据

和目标域数据在公共特征空间实现分布对齐。文献^[14]提出域自适应网络(Domain Adaptive Network, DAN),引入最大均值差异(Maximum Mean Discrepancy, MMD)的概念,在公共特征空间中对特征差异进行显式距离度量,并通过最小化源域和目标域特征的MMD来实现跨域的差异弥合。文献^[15]在引入MMD进行差异度量的同时,也将标签信息纳入考虑,实现了更加细粒度的特征对齐,确保在每一个类别上对齐特征,帮助模型学得更好的可迁移特征。文献^[16]将对抗思想引入迁移学习,通过鉴别器识别数据的来源,以实现分布差异的隐式度量,隐式度量的方式能够在数据中自动学习适合于当前数据的分布差异度量函数。文献^[17]将条件对抗学习技术引入迁移学习中,实现了特征-类别的联合分布对齐。

基于对抗思想的迁移学习方法能够从数据中学习更适用于当前数据集的分布差异度量函数,因此更适用于帕金森病的跨语种检测领域。但是,由于帕金森病患者的语音样本有其自身的特殊性,原始数据中的病理信息与数据的域信息往往纠缠在一起,使得跨语种的语音数据包含了更多的与病理无关的信息。这里,病理信息通常为受病情影响导致的语音变化,表现为波形毛刺现象多、异常停顿、高频能量增加等;域信息则为可能对受试者语音样本产生影响的一些非病理性因素(如语言、采集环境等),域信息将使受试者语音在许多声学特征指标上出现分布差异,也会导致语音时域波形和频谱的变化等。如果不能将域不变病理信息和域信息进行完整地分离,模型仍会因受到域信息(非病理性因素)的影响而出现误判。

为避免语种差异带来的影响,提升模型在跨语种场景下的检测性能,本文引入对抗迁移学习^[16]和特征解耦^[17]思想,提出一种跨语种的帕金森病检测模型。首先,将具备丰富标签信息和语音数据的MaxLittle帕金森病语音数据库(受试者母语为英语)作为源域数据,自采的帕金森病语音数据集(受试者母语为

汉语)作为目标域数据。然后,通过一个参数共享的特征提取器将源域/目标域的数据进行高级特征提取,并进一步将其拆分为域不变的病理信息表征向量和域信息表征向量。接着,通过两个目标任务不同的对抗学习模块,对前述的两个向量进行解纠缠。随后,通过对解纠缠的两个向量进行正交约束,进一步促进信息的分离。最终,实现域不变病理信息和域信息的完整分离。通过利用表征域不变病理信息的特征向量,可以实现在跨语种场景下的帕金森病检测。本文在公开的MaxLittle帕金森病语音数据集以及自采的帕金森病语音数据集上,采用十折交叉验证的方法验证了所提方法的有效性。实验结果表明:所提模型在跨语种的场景中,仍能实现86.69%检测准确率;与传统机器学习方法以及现有的迁移学习算法相比,所提模型在准确率、敏感度和F1分数等方面均有明显提升。

2 本文方法

基于对抗思想的迁移学习技术^[16],其目的就是找到那些既能够完成目标任务,又能够使域鉴别器无法确定其来自源域还是目标域的有效特征。为完成这一目的,往往需要一个特征生成网络和域鉴别网络进行博弈。理想情况下,当域鉴别网络无法识别特征生成网络的输出为源域还是目标域时,即实现了可迁移特征的提取。特征解耦思想^[18]则是指从多个相关特征中提取出彼此独立、不相关的特征表示。特征解耦试图消除特征之间的冗余信息,使得每个特征能够提供独立的、互补的信息。解耦的目

的是为了减少特征之间的相互依赖性,以便更好地捕捉数据的内在结构和特征之间的潜在关系。

基于上述两种技术,本文针对跨语种场景下的帕金森病检测问题,提出一种基于对抗迁移学习的帕金森病跨语种声学分析模型(Cross-Language Speech Analysis Model, CLSAM)。模型包含一个特征提取网络,两个目标任务不一样的对抗迁移学习模块,总体框架如图1所示。

2.1 数据预处理

本文采用公开的MaxLittle帕金森病语音数据集^[19]以及自采的帕金森病语音数据集进行跨语种声学分析研究。文中将前者作为源域数据,将后者作为目标域数据。

对源域数据作如下预处理:将每一条原始语音数据分帧,其中,帧长25 ms,偏移量10 ms;提取40维的Fbank特征序列,将其记为 $\mathbf{x}_{i,s}$ ($1 \leq i \leq N_s$),其中,s代表该数据来源于源域, N_s 为样本数量;将 $\mathbf{x}_{i,s}$ 对应的分类标签记为 $\mathbf{y}_{i,s}$,若受试者患有帕金森病,则分类标签用one-hot向量[1,0]表示,否则用[0,1]表示;将 $\mathbf{x}_{i,s}$ 对应的域标签记为 $\mathbf{d}_{i,s}$,若该样本来自源域,则域标签用one-hot向量[1,0]表示,否则用[0,1]表示;最终,组成源域样本对 $\mathbf{D}_{i,s} = (\mathbf{x}_{i,s}, \mathbf{y}_{i,s}, \mathbf{d}_{i,s})$ 。对目标域数据也作类似的预处理,生成目标域样本对 $\mathbf{D}_{i,t} = (\mathbf{x}_{i,t}, \mathbf{y}_{i,t}, \mathbf{d}_{i,t})$,其中,t代表该数据来源于目标域, $\mathbf{x}_{i,t}$ ($1 \leq i \leq N_t$)为目标域样本, $\mathbf{y}_{i,t}$ 为其对应的分类标签, $\mathbf{d}_{i,t}$ 为其对应的域标签, N_t 为样本数量。各参量具体表示方式同上述源域样本对。

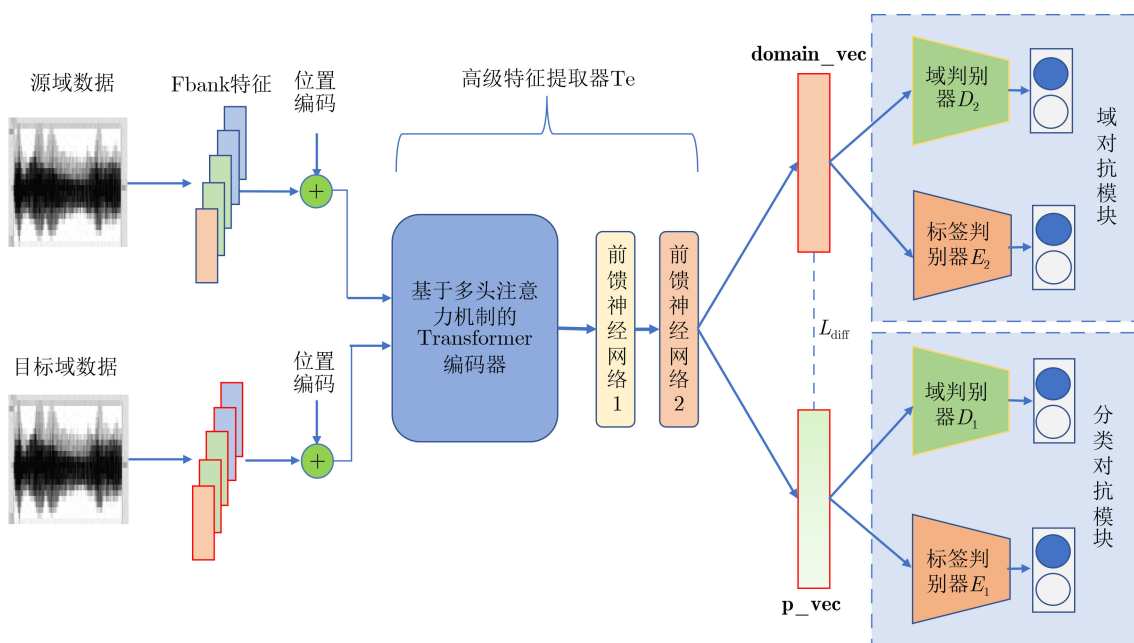


图1 跨语种声学分析模型总体框架图

2.2 特征提取网络

图1中的高级特征提取器Te由基于多头自注意力机制的Transformer编码块^[20]和多层神经网络级联组成。该高级特征提取器的Te作用是通过学习一个变换函数来实现特征变换，将提取的Fbank特征^[21](不论其来自源域还是目标域)映射至一个公共特征空间，使源域和目标域数据的分布差异得以弥合。

2.2.1 基于多头注意力机制的Transformer编码块

基于多头注意力机制的Transformer编码块的主要功能在于从时序序列中提取高级特征。预处理后所得到的Fbank特征矩阵与位置编码相加^[20]，作为多头自注意力机制模块的输入。Fbank特征矩阵表示为 $\mathbf{x}_{i,m} = [\mathbf{x}_{i,m}^1, \mathbf{x}_{i,m}^2, \dots, \mathbf{x}_{i,m}^p, \dots, \mathbf{x}_{i,m}^P]$ ，其中 $\mathbf{x}_{i,m}^p$ 为第*i*个样本第*p*帧的Fbank特征；*P*为特征序列的总帧数；*m*则用于标识该语音特征来自源域还是目标域， $m \in [s, t]$ 。

多头自注意力机制的引入，能够很好地捕获数据中的全局时序信息。基于多头自注意力机制的Transformer编码块的组成如图2所示。该编码块由多头自注意力机制模块(Multihead Self-Attention, MSA)和前馈神经网络模块(Feedforward Neural Network, FNN)组成。为加速网络的训练，还在每个块的输入前引入层归一化(LayerNorm, LN)进行数据的归一化处理，在每个块的输出后进行残差连接操作。

基于多头注意力机制的高级特征提取方式为

$$\mathbf{e}_0 = \mathbf{x}_{i,m} = [\mathbf{x}_{i,m}^1, \mathbf{x}_{i,m}^2, \dots, \mathbf{x}_{i,m}^p, \dots, \mathbf{x}_{i,m}^P] \quad (1)$$

$$\mathbf{e}'_j = \text{MSA}(\text{LN}(\mathbf{e}_{j-1})) + \mathbf{e}_{j-1}, j = 1, 2, \dots, M \quad (2)$$

$$\mathbf{e}_j = \text{FNN}(\text{LN}(\mathbf{e}'_{j-1})) + \mathbf{e}'_{j-1}, j = 1, 2, \dots, M \quad (3)$$

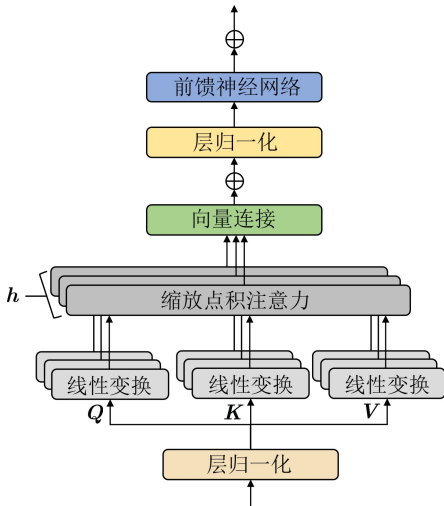


图2 基于多头自注意力机制的Transformer编码块

其中， \mathbf{e}_0 代表多头注意力机制的初始输入， \mathbf{e}'_j 代表经过*j*次多头注意力机制后的输出， \mathbf{e}_j 为 \mathbf{e}'_j 经过层归一化后的输出，*M*代表编码器网络中Transformer编码块的堆叠次数。Transformer编码块的最终输出为 $\mathbf{V}_{\text{Tr}} = [\mathbf{v}_{i,m}^1, \mathbf{v}_{i,m}^2, \dots, \mathbf{v}_{i,m}^p, \dots, \mathbf{v}_{i,m}^P]$ ，其中 $\mathbf{v}_{i,m}^p$ 是经过特征交互后的汇集信息。

2.2.2 两层前馈神经网络

经过基于多头注意力机制的Transformer编码块完成特征的初步提取后，还将再通过两层前馈神经网络进行信息分离，其作用在于对编码块提取到的初步信息进行信息选择，以配合后续的模块进行对抗训练。具体实现为

$$\mathbf{V}_{\text{Te}} = \text{FN}(\mathbf{V}_{\text{Tr}}; \boldsymbol{\theta}_{\text{FN}}) = [\mathbf{p_vec}, \mathbf{domain_vec}] \quad (4)$$

其中， $\boldsymbol{\theta}_{\text{FN}}$ 为两层神经网络层的参数。最终，特征提取网络Te输出两个同样长度的向量，即包含域不变病理信息的16维高级抽象表征向量 $\mathbf{p_vec}$ ，包含域信息的16维高级抽象表征向量 $\mathbf{domain_vec}$ 。

2.3 双重对抗迁移学习网络

由于帕金森病患者的语音样本的特殊性，从高级特征提取器Te中提取的域不变病理信息和域信息往往是纠缠在一起的。如果不能将两种信息进行完整地分离，模型仍会因受到域信息(非病理性因素)的影响而出现误判。为了能够实现对信息的高效解耦，本文设计了一个双重对抗迁移模块，对不同的向量进行不一样的对抗学习。如图1所示，双重对抗迁移学习网络由两部分组成：一个是分类对抗网络，其目的是使得分类器 E_1 基于域不变病理信息特征向量 $\mathbf{p_vec}$ 能够正确进行帕金森病检测，而域鉴别器 D_1 无法识别 $\mathbf{p_vec}$ 来自源域还是目标域；一个是域对抗模块，其目的是使得域鉴别器 D_2 基于 $\mathbf{domain_vec}$ 能够正确进行域鉴别，而分类器 E_2 无法识别 $\mathbf{domain_vec}$ 来自健康人还是帕金森病患者。

2.3.1 分类对抗网络

分类对抗网络由分类器 E_1 和域鉴别器 D_1 组成，它们都由多层全连接神经网络组成。将所述的域不变病理信息特征向量 $\mathbf{p_vec}$ 输入分类器 E_1 ($\mathbf{p_vec}; \boldsymbol{\theta}_{e_1}$) 获取第*i*个样本的症状标签 $\hat{\mathbf{y}}_{i,m_1}$ ($m_1 \in [s, t]$)，其中 $\boldsymbol{\theta}_{e_1}$ 为网络参数。同时，还将 $\mathbf{p_vec}$ 输入到域鉴别器 D_1 ($\mathbf{p_vec}; \boldsymbol{\theta}_{d_1}$) 进行识别，获得其域标签 $\hat{\mathbf{d}}_{i,m_1}$ ($m_1 \in [s, t]$)，其中 $\boldsymbol{\theta}_{d_1}$ 为网络参数。

2.3.2 域对抗网络

域对抗网络由分类器 E_2 和域鉴别器 D_2 组成，也是由多层全连接神经网络组成。将域信息特征向量 $\mathbf{domain_vec}$ 输入到域鉴别器 D_2 ($\mathbf{domain_vec}; \boldsymbol{\theta}_{d_2}$)，

获取第*i*个样本的域标签 \hat{d}_{i,m_2} ($m_2 \in [s, t]$), 其中, θ_{d2} 为网络参数。同时, 还将 **domain_vec** 输入到 $E_2(\text{domain_vec}; \theta_{e2})$ 分类器识别其症状标签, 获得结果 \hat{y}_{i,m_2} ($m_2 \in [s, t]$), 其中, θ_{e2} 为网络参数。

2.4 损失函数

2.4.1 分类对抗损失

通过分类器 E_1 和域鉴别器 D_1 的对抗学习, 提取可实现帕金森病检测的病理信息并排除域信息。本文将最小化分类器 E_1 的损失, 同时最大化域鉴别器 D_1 的损失, 目标函数为

$$L_s(\theta_{Te}, \theta_{e1}, \theta_{d1}) = \frac{1}{n} \sum_{\mathbf{x}_{i,m} \in (D_s, D_t)} (L_{e1}(E_1(\text{Te}(\mathbf{x}_{i,m}; \theta_{Te}); \theta_{e1}), \mathbf{y}_{i,m}) - L_{d1}(D_1(\text{Te}(\mathbf{x}_{i,m}; \theta_{Te}); \theta_{d1}), \mathbf{d}_{i,m})) \quad (5)$$

其中, n 为 N_s 和 N_t 之和。 L_{e1} 为分类器 E_1 的损失函数, 将采用交叉熵损失函数, 形式为

$$L_{e1} = \frac{1}{N} \sum_i \sum_{c=1}^C y_{ic} \ln p_{ic} \quad (6)$$

其中, N 代表样本个数, C 代表类别数量, y_{ic} 为符号函数, 如果样本 i 的真实类别等于 c , 则 y_{ic} 取1, 反之为0; p_{ic} 为分类器将样本 i 预测为类别 c 的概率。域鉴别器 D_1 也将采用上述损失函数。

2.4.2 域对抗损失

domain_vec 通过分类器 E_2 和域鉴别器 D_2 的对抗学习, 融合域信息并排除病理特征信息。本文将最小化域鉴别器 D_2 的损失, 同时最大化分类器 E_2 的损失, 目标函数为

$$L_d(\theta_{Te}, \theta_{e2}, \theta_{d2}) = \frac{1}{n} \sum_{\mathbf{x}_{i,m} \in (D_s, D_t)} (L_{d2}(D_2(\text{Te}(\mathbf{x}_{i,m}; \theta_{Te}); \theta_{d2}), \mathbf{d}_{i,m}) - L_{e2}(E_2(\text{Te}(\mathbf{x}_{i,m}; \theta_{Te}); \theta_{e2}), \mathbf{y}_{i,m})) \quad (7)$$

其中, L_{e2} 为分类器 E_2 的损失函数, L_{d2} 为域鉴别器 D_2 的损失函数, 且皆为交叉熵损失函数。

2.4.3 特征正交约束

为进一步确保病理信息特征向量中排除了域信息, 本文通过对两个特征向量施加正交约束^[22]来定义差分损失, 并将差分损失最小化以促进信息的解纠缠。这里, 将差分损失同时用于源域和目标域, 通过在每个域上的病理信息特征向量和域特征向量之间施加软子空间正交约束来定义损失函数。记矩阵 \mathbf{V}_s^e 为由源域数据提取的病理信息特征向量作为行组成的矩阵, 矩阵 \mathbf{V}_s^d 为由源域数据提取的域信息特征向量作为行组成的矩阵, 矩阵 \mathbf{V}_t^e 为由目标域数据提取的病理特征向量作为行组成的矩阵,

\mathbf{V}_t^d 为由目标域数据提取的域信息特征向量作为行组成的矩阵。差分损失的具体表达为

$$L_{\text{diff}}(\mathbf{V}_s^e; \mathbf{V}_s^d; \mathbf{V}_t^e; \mathbf{V}_t^d) = \left\| [\mathbf{V}_s^e]^T \mathbf{V}_s^d \right\|_F^2 + \left\| [\mathbf{V}_t^e]^T \mathbf{V}_t^d \right\|_F^2 \quad (8)$$

其中, $[\cdot]^T$ 表示矩阵转置, $\|\cdot\|_F^2$ 为平方Frobenius范数。

为避免多个损失共同优化的过程中出现由于尺度不一致导致模型偏向大损失的方向优化, 导致其他模块的性能下降。本文还预设了3个超参数 λ_s , λ_d , λ_{di} , 通过对各个损失进行加权, 减小尺度不一致对模型的影响。最终, 模型的总体优化目标函数为

$$L_{\text{total}}(\theta_{Te}, \theta_{e1}, \theta_{d1}, \theta_{e2}, \theta_{d2}) = \lambda_s L_s + \lambda_d L_d + \lambda_{di} L_{\text{diff}} \quad (9)$$

2.5 模型训练及测试

模型将根据总体优化目标函数采用随机梯度下降法(Stochastic Gradient Descent, SGD)优化方法^[22]进行模型的优化与参数的迭代

$$(\tilde{\theta}_{Te}, \tilde{\theta}_{e1}, \tilde{\theta}_{d1}, \tilde{\theta}_{e2}, \tilde{\theta}_{d2}) = \arg \min_{\theta_{Te}, \theta_{e1}, \theta_{d1}, \theta_{e2}, \theta_{d2}} L_{\text{total}}(\theta_{Te}, \theta_{e1}, \theta_{d1}, \theta_{e2}, \theta_{d2}) \quad (10)$$

固定特征提取器的参数 θ_{Te} 、分类器 E_1 的参数 θ_{e1} 、域鉴别器 D_2 的参数 θ_{d2} , 并最大化如式(11)的损失函数

$$(\tilde{\theta}_{d1}, \tilde{\theta}_{e2}) = \arg \max_{\theta_{d1}, \theta_{e2}} L_{\text{total}}(\theta_{Te}, \theta_{e1}, \theta_{d1}, \theta_{e2}, \theta_{d2}) \quad (11)$$

基于对抗迁移学习的跨语种帕金森病检测算法流程如**算法1**所示。

3 实验

3.1 数据集

本文使用了两个不同语种的帕金森病语音数据集。第1个语音数据集为牛津大学学者MaxLittle与科罗拉多州丹佛的国家语音中心合作创建的帕金森病检测数据集^[19], 受试者以英语为母语。数据采集均在安静的室内环境进行, 采集内容为持续元音/a/, 受试者包括33名帕金森病患者和10名健康人, 每人重复发声6次。经剪辑整理后, 生成96条健康人语音数据和154条帕金森病患者语音数据, 共计260条数据。MaxLittle数据集的统计信息详见**表1**。

第2个语音数据集为本文研究团队与南京医科大学附属老年医院神经内科合作创建的自采帕金森病语音数据集, 受试人群皆为汉语人群。数据集由该医院帕金森病及运动障碍专病门诊筛选出的68名患者和17名健康人的语音数据构成。采集内容为持续元音/a/, 经剪辑处理后共计170个样本。自采帕金森病语音数据集的统计信息详见**表2**。

算法1 基于对抗迁移学习的跨语种帕金森病检测算法

```

输入：源域数据集  $D_s$  和目标域数据集  $D_t$ 
输出：可学习参数  $\tilde{\theta}_{Te}, \tilde{\theta}_{e1}, \tilde{\theta}_{d1}, \tilde{\theta}_{e2}, \tilde{\theta}_{d2}$ 
Repeat
  //特征学习阶段
  For 从源域数据中选取一个批次的样本:
    计算损失  $L_s(\theta_{Te}, \theta_{e1}, \theta_{d1})$ ;
    计算损失  $L_d(\theta_{Te}, \theta_{e2}, \theta_{d2})$ ;
    计算损失  $L_{diff}(V_s^e; V_s^d)$ ;
    根据式(10)计算梯度，并更新  $\tilde{\theta}_{Te}, \tilde{\theta}_{e1}, \tilde{\theta}_{d1}, \tilde{\theta}_{e2}, \tilde{\theta}_{d2}$ 
  End
  For 从目标域数据中选取一个批次的样本:
    计算损失  $L_s(\theta_{Te}, \theta_{e1}, \theta_{d1})$ ;
    计算损失  $L_d(\theta_{Te}, \theta_{e2}, \theta_{d2})$ ;
    计算损失  $L_{diff}(V_t^e; V_t^d)$ ;
    根据式(10)计算梯度，并更新  $\tilde{\theta}_{Te}, \tilde{\theta}_{e1}, \tilde{\theta}_{d1}, \tilde{\theta}_{e2}, \tilde{\theta}_{d2}$ 
  End
  //对抗迁移阶段
  For 对源域或目标域的每一个样本，固定参数  $\theta_{Te}$ 、参数  $\theta_{e1}$ 、参数  $\theta_{d2}$ 
    计算损失  $L_s(\theta_{Te}, \theta_{e1}, \theta_{d1})$ ;
    计算损失  $L_d(\theta_{Te}, \theta_{e2}, \theta_{d2})$ ;
    计算损失  $L_{diff}(V_s^e; V_s^d)$  或  $L_{diff}(V_t^e; V_t^d)$ ;
    根据式(11)计算梯度，并更新  $\tilde{\theta}_{d1}, \tilde{\theta}_{e2}$ 
  End
Until模型收敛

```

3.2 实验条件及设置

本文实验将使用MaxLittle语音数据集和自采的帕金森病语音数据集来验证本文所提方法的有效

性。对所有的语音数据集提取[361,40]的Fbank特征。为了克服由于随机选择测试数据而存在的任何偏差，使用10折交叉验证进行方法可行性验证，90%的语音数据用于模型训练，剩余的10%语音数据用于测试。为了充分验证本文所提方法的有效性，度量标准除常用的准确率外，本文选择准确率(Accuracy, ACC)、敏感度(SENsitivity, SEN)以及F1分数(F1-score)作为度量标准，衡量模型在跨语种场景下评估帕金森病情方面的性能。准确率表示准确区分帕金森病患者和健康人的概率，敏感度代表正确检测出帕金森病患者的概率，F1分数衡量模型的总体预测性能。

本文所提模型为深度神经网络模型，表3是模型各个模块的具体参数，以及训练过程中的实验设置。

3.3 实验结果分析

3.3.1 与传统机器学习模型的性能比较

为验证所提的跨语种声学分析模型的优势，本节将所提模型与传统的机器学习模型(随机森林(Random Forest, RF)和支持向量机(Support Vector Machine, SVM))进行性能比较。实验结果如表4所示，其中，模型名称后括号中的标号含义如下：s代表在MaxLittle数据集上训练及测试；t代表在自采数据集上训练及测试；t-s为在自采数据集上训练，在MaxLittle数据集上测试；s-t为在MaxLittle数据集上训练，自采数据集上测试；st为在联合数据集上训练测试。

由表4可知，传统机器学习模型在跨语种数据集上性能都有下降，即使通过添加跨语种数据到训练数据集中，这种下降的趋势也未有改变。其根本

表1 MaxLittle数据集的统计信息

	男性		女性		合计	
	PD	HC	PD	HC	PD	HC
受试者类别	PD	HC	PD	HC	PD	HC
受试者人数	22	4	11	6	33	10
平均年龄及统计方差	67.2 (9.3)	61(8.6)	67.2(9.3)	61(8.6)	67.2(9.3)	61(8.6)
年龄分布	48~85	46~72	48~85	46~72	48~85	46~72

表2 自采帕金森病语音数据集的统计信息

	男性		女性		合计	
	PD	HC	PD	HC	PD	HC
受试者类别	PD	HC	PD	HC	PD	HC
受试者人数	49	8	19	9	68	17
平均年龄及统计方差	69.3(9.5)	66.5(7.2)	69.8(8.2)	65.3(6.8)	69.4(9.2)	65.9(7.0)
年龄分布	46~88	58~77	56~84	53~74	46~88	53~77
平均病情持续时间及统计方差	5.9 (3.6)	0	5.4 (3.1)	0	5.8 (3.4)	0
HY分期	1~4	0	1~4	0	1~4	0

原因在于,传统模型未能去除语种差异带来的影响。而本文所提的跨语种声学分析模型,则能够很好地在跨语种数据集上实现更加准确稳定的帕金森病检测。相较于传统的机器学习分类模型(RF, SVM),在跨语种数据集上的准确率分别提升了7%, 8%。

3.3.2 与现有的迁移学习算法的性能比较

本节将进行迁移学习域自适应算法之间的性能比较。为保证输入的一致性,对于每个域自适应算法,将其中的特征学习网络改为和本文所提模型一致的特征提取网络,达到去除影响模型性能比较的其他因素。参与比较的模型有:域自适应网络(Domain Adaptive Network, DAN)^[14]、域对抗神经

网络(Domain Adversarial Neural Networks, DANN)^[16]、深度子域自适应网络(Deep Subdomain Adaptation Network, DSAN)^[15]、条件对抗域自适应网络(Conditional Adversarial Domain Adaptation Network, CADAN)^[17]、域分离网络(Domain Separation Network, DSN)^[23]、可转移特征过滤网络(Transferable Feature Filtration Network, TFFN)^[24]等。所提模型与上述各迁移学习模型的性能比较如表5所示。

从实验结果可知,本文所提模型在跨语种场景下的帕金森病检测性能相较于其它模型有更好的表现。根本原因在于,所提模型有效地排除了语种差异的影响,更好地分离了域信息和域不变病理的信息。相较于基于统计准则(采用MMD准则进行差异分布度量)的迁移模型(如DAN, DSAN),对抗学习的隐式距离度量方式基于当前数据进行自适应学习,更加适合当前的数据。而DANN, CADAN两种分类对抗结构由于没有对融合信息进行解耦,因而信息分离能力较差;相较于TFFN采用共享特征提取器和基于对抗学习的域鉴别器来提取和对齐可转移特征,本文采用的双重对抗迁移学习更能有效地捕捉“可转移的”病理信息。与同样具备信息分离的DSN模型比较,本文所提模型不仅同样对两种目标信息进行差异约束,还通过更具迁移学习能力的对抗迁移学习方式实现了信息分离。

3.3.3 消融实验

为进一步探究所提模型的性能,本节通过消融实验来检测所提模型中关键模块的作用,重点探索双重对抗网络的必要性,以及特征正交约束对模型的贡献。实验的详细结果如表6所示。

表3 CLSAM模型参数设置

网络结构参数	参数值
X _s	361×40
X _t	361×40
Transformer编码块Q, K, V向量维度	64
Transformer编码块多头注意力	2
Transformer编码块深度	6
多层前馈神经网络	[32,32]
domain_vec	16
p_vec	16
域鉴别器网络D ₁	[32,16,2]
域鉴别器网络D ₂	[16,2]
帕金森病检测模块E ₁	[16,2]
帕金森病检测模块E ₂	[32,16,2]
周期数	120
学习率	0.001
批大小	36
优化器	SGD
Dropout	0.1

表4 与传统机器学习模型的性能比较(%)

模型	Acc.	Sen.	F1.
CLSAM	86.69	85.98	84.71
RF(s)	79.86	77.41	78.88
RF(t)	78.62	77.32	77.26
RF(s-t)	76.81	75.25	74.75
RF(t-s)	76.38	75.36	74.46
RF(st)	79.15	78.35	78.18
SVM(s)	79.52	77.53	78.35
SVM(t)	77.34	77.61	78.15
SVM(s-t)	75.72	74.46	74.26
SVM(t-s)	75.35	73.45	72.86
SVM(st)	78.95	76.86	75.68

表5 与迁移学习模型的性能比较(%)

模型	Acc.	Sen.	F1.
CLSAM	86.69	85.98	84.71
DAN	80.83	81.86	81.56
DSAN	83.65	83.82	82.61
DANN	82.78	82.98	82.81
CADAN	84.10	83.22	83.56
TFFN	85.64	84.58	83.89
DSN	83.60	82.84	83.15

表6 消融实验(%)

模型	Acc.	Sen.	F1.
CLSAM	86.69	85.98	84.71
CLSAM (不含双重对抗训练)	82.78	82.98	82.31
CLSAM (不含特征正交约束)	85.23	83.74	83.15
CLSAM(带有HSIC约束)	85.96	84.85	84.17

实验结果表明：双重对抗训练能够更好地分离域不变病理信息和域信息，帮助模型实现更好的跨域分析能力。同时，本文中对域信息和域不变病理信息实施的正交约束也能够进一步确保信息的完整分离，再度帮助模型提升性能。在正交约束问题上，将本文方法与常用的希尔伯特-施密特独立性准则(Hilbert-Schmidt Independence Criterion, HSIC)^[25]进行了对比。实验结果表明，后者性能略差，这是由于后者对样本容量要求较高，且无法较好地处理病理特征与非病理信息之间存在的复杂非线性关系。

4 结束语

本文针对当前跨语种场景下帕金森病检测模型易受数据分布差异影响而出现性能下降的现象，提出一种帕金森病跨语种声学分析的方法，用于实现跨语种场景下的帕金森病检测。通过在多语种的语音数据集上训练模型，采用双重对抗网络解耦数据中的域信息和域不变的病理信息。最终实现了从原始语音特征中提取跨语种声学特征，并在此基础上实现跨语种的帕金森病辅助检测，使得研究成果能够适用于不同地域、不同语种患者使用。本文工作主要在于通过迁移学习技术实现跨域分布差异弥合的领域适应问题，未来工作可以考虑将领域泛化技术引入帕金森病跨语种语音分析领域，使得模型能够在未见过某一语种语音的情况下仍然能够实现正确的语音分析，并准确的进行帕金森病检测。

参考文献

- [1] GULLAPALLI A S and MITTAL V K. Early detection of Parkinson's disease through speech features and machine learning: a review[C]. *ICT with Intelligent Applications: Proceedings of ICTIS, Singapore, 2022*: 203–212. doi: [10.1007/978-981-16-4177-0_22](https://doi.org/10.1007/978-981-16-4177-0_22).
- [2] BENBA A, JILBAB A, SANDABAD S, *et al.* Voice signal processing for detecting possible early signs of Parkinson's disease in patients with rapid eye movement sleep behavior disorder[J]. *International Journal of Speech Technology*, 2019, 22(1): 121–129. doi: [10.1007/s10772-018-09588-0](https://doi.org/10.1007/s10772-018-09588-0).
- [3] 季薇, 杨茗淇, 李云, 等. 基于掩蔽自监督语音特征提取的帕金森病检测方法[J]. *电子与信息学报*, 2023, 45(10): 3502–3510. doi: [10.11999/JEIT221041](https://doi.org/10.11999/JEIT221041).
JI Wei, YANG Mingqi, LI Yun, *et al.* Parkinson's disease detection method based on masked self-supervised speech feature extraction[J]. *Journal of Electronics & Information Technology*, 2023, 45(10): 3502–3510. doi: [10.11999/JEIT221041](https://doi.org/10.11999/JEIT221041).
- [4] SUPHINNAPONG P, PHOKAEWVARANGKUL O, THUBTHONG N, *et al.* Objective vowel sound characteristics and their relationship with motor dysfunction in Asian Parkinson's disease patients[J]. *Journal of the Neurological Sciences*, 2021, 426: 117487. doi: [10.1016/j.jns.2021.117487](https://doi.org/10.1016/j.jns.2021.117487).
- [5] HSU S C, JIAO Yishan, MCAULIFFE M J, *et al.* Acoustic and perceptual speech characteristics of native Mandarin speakers with Parkinson's disease[J]. *The Journal of the Acoustical Society of America*, 2017, 141(3): EL293–EL299. doi: [10.1121/1.4978342](https://doi.org/10.1121/1.4978342).
- [6] KOVAC D, MEKYSKA J, GALAZ Z, *et al.* Multilingual analysis of speech and voice disorders in patients with Parkinson's Disease[C]. *The 44th International Conference on Telecommunications and Signal Processing, Brno, Czech Republic, 2021*: 273–277. doi: [10.1109/TSP52935.2021.9522597](https://doi.org/10.1109/TSP52935.2021.9522597).
- [7] VÁSQUEZ-CORREA J C, ARIAS-VERGARA T, RIOS-URREGO C D, *et al.* Convolutional neural networks and a transfer learning strategy to classify Parkinson's Disease from speech in three different languages[C]. *24th Iberoamerican Congress on Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, Havana, Cuba, 2019*: 697–706. doi: [10.1007/978-3-030-33904-3_66](https://doi.org/10.1007/978-3-030-33904-3_66).
- [8] KIM Y and CHOI Y. A cross-language study of acoustic predictors of speech intelligibility in individuals with Parkinson's Disease[J]. *Journal of Speech, Language, and Hearing Research*, 2017, 60(9): 2506–2518. doi: [10.1044/2017_JSLHR-S-16-0121](https://doi.org/10.1044/2017_JSLHR-S-16-0121).
- [9] NISHIO M and NIIMI S. Comparison of speaking rate, articulation rate and alternating motion rate in dysarthric speakers[J]. *Folia Phoniatrica et Logopaedica*, 2006, 58(2): 114–131. doi: [10.1159/000089612](https://doi.org/10.1159/000089612).
- [10] OROZCO-ARROYAVE J R, HÖNIG F, ARIAS-LONDOÑO J D, *et al.* Automatic detection of Parkinson's disease in running speech spoken in three different languages[J]. *The Journal of the Acoustical Society of America*, 2016, 139(1): 481–500. doi: [10.1121/1.4939739](https://doi.org/10.1121/1.4939739).
- [11] YEO E J, CHOI K, KIM S, *et al.* Cross-lingual dysarthria severity classification for English, Korean, and Tamil[C]. *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Chiang Mai, Thailand, 2022*: 566–574. doi: [10.23919/APSIPAASC55919.2022.9980124](https://doi.org/10.23919/APSIPAASC55919.2022.9980124).
- [12] VÁSQUEZ-CORREA J C, RIOS-URREGO C D, ARIAS-VERGARA T, *et al.* Transfer learning helps to improve the accuracy to classify patients with different speech disorders in different languages[J]. *Pattern Recognition Letters*, 2021, 150: 272–279. doi: [10.1016/j.patrec.2021.04.011](https://doi.org/10.1016/j.patrec.2021.04.011).
- [13] JIANG Junguang, SHU Yang, WANG Jianmin, *et al.*

- Transferability in deep learning: A survey[J]. arXiv: 2201.05867, 2022. doi: [10.48550/arXiv.2201.05867](https://doi.org/10.48550/arXiv.2201.05867).
- [14] GHIFARY M, KLEIJN W B, and ZHANG Mengjie. Domain adaptive neural networks for object recognition[C]. 13th Pacific Rim International Conference on Artificial Intelligence, Gold Coast, QLD, Australia, 2014: 898–904. doi: [10.1007/978-3-319-13560-1_76](https://doi.org/10.1007/978-3-319-13560-1_76).
- [15] ZHU Yongchun, ZHUANG Fuzhen, WANG Jindong, *et al.* Deep subdomain adaptation network for image classification[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 32(4): 1713–1722. doi: [10.1109/tnnls.2020.2988928](https://doi.org/10.1109/tnnls.2020.2988928).
- [16] GANIN Y, USTINOVA E, AJAKAN H, *et al.* Domain-adversarial training of neural networks[J]. *The Journal of Machine Learning Research*, 2016, 17(1): 2096–2030. doi: [10.1007/978-3-319-58347-1_10](https://doi.org/10.1007/978-3-319-58347-1_10).
- [17] LONG Mingsheng, CAO Zhangjie, WANG Jianmin, *et al.* Conditional adversarial domain adaptation[C]. The 32nd International Conference on Neural Information Processing Systems, Montréal, Canada, 2018: 1647–1657. doi: [10.5555/3326943.3327094](https://doi.org/10.5555/3326943.3327094).
- [18] CAI Ruichu, LI Zijian, WEI Pengfei, *et al.* Learning disentangled semantic representation for domain adaptation[C]. International Joint Conferences on Artificial Intelligence (IJCAI), Macao, China, 2019: 2060–2066. doi: [10.24963/ijcai.2019/285](https://doi.org/10.24963/ijcai.2019/285).
- [19] TSANAS A, LITTLE M A, MCSHARRY P E, *et al.* Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease[J]. *IEEE Transactions on Biomedical Engineering*, 2012, 59(5): 1264–1271. doi: [10.1109/TBME.2012.2183367](https://doi.org/10.1109/TBME.2012.2183367).
- [20] VASWANI A, SHAZEER N, PARMAR N, *et al.* Attention is all you need[C]. Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, USA, 2017: 6000–6010. doi: [10.5555/3295222](https://doi.org/10.5555/3295222).
- [21] OROZCO-ARROYAVE J R, VÁSQUEZ-CORREA J C, VARGAS-BONILLA J F, *et al.* NeuroSpeech: An open-source software for Parkinson's speech analysis[J]. *Digital Signal Processing*, 2018, 77: 207–221. doi: [10.1016/j.dsp.2017.07.004](https://doi.org/10.1016/j.dsp.2017.07.004).
- [22] CAI D, HE X, HAN J, *et al.* Orthogonal Laplacianfaces for face recognition[J]. *IEEE Transactions on Image Processing*, 2006, 15(11): 3608–3614. doi: [10.1109/TIP.2006.881945](https://doi.org/10.1109/TIP.2006.881945).
- [23] BOUSMALIS K, TRIGEORGIS G, SILBERMAN N, *et al.* Domain separation networks[C]. The 30th International Conference on Neural Information Processing Systems, Barcelona, Spain, 2016: 343–351. doi: [10.5555/3157096.3157135](https://doi.org/10.5555/3157096.3157135).
- [24] LI Yiyang, WANG Shengsheng, WANG Bilin, *et al.* Transferable feature filtration network for multi-source domain adaptation[J]. *Knowledge-Based Systems*, 2023, 260: 110113. doi: [10.1016/J.KNOSYS.2022.110113](https://doi.org/10.1016/J.KNOSYS.2022.110113).
- [25] SONG L, SMOLA A, GRETTON A, *et al.* Supervised feature selection via dependence estimation[C]. The 24th International Conference on Machine Learning, 2007: 823–830. doi: [10.1145/1273496.1273600](https://doi.org/10.1145/1273496.1273600).
- 季 薇: 女, 博士, 教授, 硕士生导师, 研究方向为机器学习与信号处理的交叉研究、无线通信与通信信号处理等。
- 王传瑜: 男, 硕士生, 研究方向为机器学习与信号处理的交叉研究。
- 吴 迪: 男, 硕士生, 研究方向为机器学习与信号处理的交叉研究。
- 李 云: 男, 博士, 教授, 博士生导师, 研究方向为机器学习、特征选择、信息安全等。
- 郑慧芬: 女, 博士, 主任医师, 研究方向帕金森病及相关运动障碍性疾病。

责任编辑: 余 蓉