

DOI: 10.19650/j.cnki.cjsi.J2108683

# 双阶段帕金森病语音聚类包络卷积稀疏迁移学习算法\*

张小恒<sup>1,2</sup>, 李勇明<sup>2</sup>, 王 品<sup>2</sup>

(1. 重庆开放大学 重庆 400052; 2. 重庆大学微电子与通信工程学院 重庆 400044)

**摘 要:** 帕金森病 (PD) 语音识别算法研究对于其及时诊疗具有重要意义, 但现有 PD 语音识别算法面临小样本数据量问题挑战。针对问题, 本文提出双面双阶段均值聚类包络和卷积稀疏迁移学习算法。在双阶段学习方面, 首先基于源数据集训练多组卷积核, 然后通过中间集得到最优卷积核并对目标集进行编码。在深度样本聚类包络方面, 首先设计迭代均值聚类算法构建深度样本空间; 然后进行样本特征同时选择并训练分类器模型; 最后对不同样本空间的分类结果进行融合。实验选取代表性的 PD 语音数据集进行验证。实验结果表明, 本文算法创新部分有效, 与 10 多个经典和最新相关文献算法相比取得了显著改进, 准确率达 97.8%。此外, 本文算法的时间复杂度不高, 满足临床应用要求。

**关键词:** 帕金森病语音识别; 包络学习; 深度样本学习; 均值聚类; 双阶段卷积稀疏迁移学习

**中图分类号:** TP181 TH77 **文献标识码:** A **国家标准学科分类代码:** 510.4

## Two-stage PD speech clustering envelope and convolution sparse transfer learning algorithm

Zhang Xiaoheng<sup>1,2</sup>, Li Yongming<sup>2</sup>, Wang Pin<sup>2</sup>

(1. Chongqing Open University, Chongqing 400052, China; 2. School of Microelectronics and Communication Engineering, Chongqing University, Chongqing 400044, China)

**Abstract:** The research on the Parkinson's disease (PD) speech recognition algorithm is important for timely diagnosis and treatment. However, the existing public PD speech datasets are characterized by small sample sizes, which is one of the main challenges faced by existing PD speech recognition methods. To address this issue, a novel dual-side two-stage means clustering envelope and convolution sparse transfer learning model is proposed. First, for the dataset side, multiple groups of convolution kernels are trained, which is based on the source domain dataset. Then, the optimal convolution kernels are filtered by the encoded intermediate dataset. Finally, the target domain dataset is encoded by the optimal kernels. In regard to deep instances clustering envelope, an iterative mean clustering algorithm is designed to construct the deep instance space. Secondly, various classifiers are developed after sample/feature parallel selection. Finally, the classification results of different instance layers are fused. In the experiment, the representative PD speech datasets are selected for verification. Experimental results show that the main innovative parts of the proposed algorithm are effective. Compared with more than ten classical algorithms, the obvious improvements in terms of classification accuracy are achieved 97.8%. In addition, the proposed algorithm has potential in clinical applications for acceptable time complexity.

**Keywords:** PD speech recognition; envelope learning; deep instance learning; means clustering; two-stage convolution sparse transfer learning

## 0 引 言

帕金森病 (Parkinson's disease, PD) 是一种长期进行

性神经疾病, 早期诊断和及时干预是提升 PD 诊疗质量的关键因素<sup>[1]</sup>。基于语音识别的帕金森诊断算法具有实时、自动化及非接触非侵入式的优点, 但 PD 语音数据集存在小样本问题, 制约了语音识别性能的进一步提升<sup>[2-6]</sup>。

收稿日期: 2021-10-09 Received Date: 2021-10-09

\* 基金项目: 国家自然科学基金重点项目 (U21A20448)、国家自然科学基金 (61771080)、重庆市自然科学基金 (cstc2020jcyj-msxmX0100)、重庆市社会科学规划项目 (2018YBYY133)、重庆市教委科学技术研究项目 (KJQN202104002) 资助

迁移学习能有效解决此类小样本问题,但极少用于PD语音识别研究中<sup>[7-9]</sup>。此外,现有迁移学习方法主要是单阶段迁移学习,并没有考虑如何有效利用两个及以上源数据集的问题。另外现有研究方法并没有考虑样本间的层次结构信息。

因此有必要研究如下3方面内容。1)有必要研究如何利用两个及以上源数据集进行迁移学习。2)有必要研究如何设计高效编码方法用于迁移学习,从而充分的迁移PD语音样本信息。3)有必要研究如何有效挖掘样本间的层次结构信息。上述3个研究工作对提升小样本PD语音数据集的分类效果以满足实际应用需求具有重要意义。

本文的主要贡献和创新点有:

1)提出了双阶段卷积稀疏迁移学习算法。在迁移学习第1阶段,基于中间集筛选源数据集中的最优结构。在迁移学习第2阶段,将最优结构迁移到目标集,实现目标集的更有效表达,从而利用两个源数据集更好地解决小样本问题。

2)提出了基于迭代均值聚类包络的PD语音深度样本学习算法,实现PD语音样本深度变换,通过多层均值聚类包络重构PD语音样本得到新样本,从而实现对本样本聚类包络的识别,提取了PD语音样本的多层次结构信息,有助于更好的解决小样本问题。

3)基于上述两个创新成果,提出了双面双阶段PD语音均值聚类包络和卷积稀疏迁移学习算法。同时考虑多数据集迁移和样本深度学习,从而深挖PD语音样本信息。该方法为迁移多个PD语音源数据集提供了方法参考。

## 1 双面双阶段均值聚类包络和卷积稀疏迁移学习算法

本文针对上述3个问题展开数据集方面及样本层次结构方面的多阶段迁移学习研究,从而构建一种新颖的PD语音识别算法。为方便描述,两个源数据集中更接近目标集的源数据集记为中间数据集。针对第1个问题本文研究基于中间集得到源数据集最优结构信息并将其迁移到目标集中的方法,从而解决PD语音数据的小样本问题;针对第2个问题,研究如何将卷积稀疏编码(convolutional sparse coding, CSC)<sup>[10]</sup>用于高效提取PD语音源数据集的隐含信息的方法;针对第3个问题研究如何通过构建层次化的样本空间来反映样本间的多层结构信息的方法。目前的深度学习通过对原始数据特征进行多层变换,发现其中的层次结构信息从而获得更高质量特征。受此启发,本文借鉴特征多层变换思路来用于样本多层变换(即“深度样本学

习”)。此外,现有部分研究采用欠采样及样本变换方法一定程度增强了分类性能<sup>[11-15]</sup>,也间接证实了样本的多层结构信息将有益于分类。因此,本文提出了基于迭代均值聚类的深度样本学习机制用于提取样本的层次化结构信息。综上分析,本文提出了一种新的PD语音识别算法-双面双阶段PD语音均值聚类包络和卷积稀疏迁移学习算法。

本文所提出算法的主要流程描述如下:第1阶段迁移学习首先对公共语音数据集进行加噪扩展,提取多类型特征构建源域数据集。然后,对源域数据集进行快速卷积稀疏编码训练得到多组卷积核,并基于不同卷积核组对中间集进行编码并生成对应的特征映射。最后,根据不同卷积核对应特征映射的平均准确率筛选出最优卷积核。第2阶段迁移学习主要从数据样本的角度设计深度样本学习算法,主要流程如下:首先,设计基于迭代均值聚类算法的深度样本学习机制。然后,基于第1阶段的最优卷积核生成对应的特征映射,并对其进行语音段特征同时筛选,从而构建新数据集。在深度空间中,基于训练集训练分类模型并得到受试者在不同空间层的预测标签。最后,基于多层样本预测标签,进行决策层投票融合得到最终预测标签。本文算法分两个版本:具有深度样本机制的双阶段迁移学习版本(two-stage transfer learning with deep instance space construction, TSTL&DISC),没有深度样本机制的版本(two-stage transfer learning, TSTL)。

### 1.1 源域及目标域样本集构建

由于可获取的公共语音集的样本量有限,本文采用对原始语音样本加噪的方式扩展样本量。

扩展后的数据集 $S' = \{S'_j\}_{j=1}^J$ ,其中 $S'_j = \phi(S, N_j, SNR_j)$ , $S$ 为公共语音数据集的原始语音信号。 $N_j$ 表示不同类型的噪声信号, $\phi(\cdot)$ 是可指定不同信噪比的噪声叠加函数。

通过特征提取构建源域特征集 $Y = \{\bar{Y}_{l'}\}_{l'=1}^L = \{\bar{Y}_l\}_{l=1}^L$ ,其中特征集样本 $\bar{Y}_{l'} = [\xi_1(\bar{S}'_{l'}) \xi_2(\bar{S}'_{l'}) \cdots \xi_N(\bar{S}'_{l'})]^T$ , $1 \leq l' \leq L$ , $\bar{S}'_{l'}$ 是扩展集 $S'$ 的样本, $N$ 为特征数量,多类型特征提取方式来源于文献[16], $L$ 为特征集样本总数。

$H_0 \times N$ 块矩阵如下:

$\bar{Y}_l = [\bar{Y}_{(l-1) \times H_0 + 1} \bar{Y}_{(l-1) \times H_0 + 2} \cdots \bar{Y}_{(l-1) \times H_0 + H_0}]^T$ , $1 \leq l \leq L$ 表示卷积稀疏编码学习的受试者(subject)训练样本。

类似的,基于原始的病理语音数据集 $V = \{\bar{V}_{b'}\}_{b'=1}^{B'}$ 构建中间特征数据集 $B = \{\bar{B}_{b'}\}_{b'=1}^{B'} = \{\bar{B}_b\}_{b=1}^B$ ,其中样本 $B_{b'} = [\xi_1(\bar{V}_{b'}) \xi_2(\bar{V}_{b'}) \cdots \xi_N(\bar{V}_{b'})]^T$ , $1 \leq b' \leq B'$ ,

$H_0 \times N$  块矩阵如下:

$$\bar{\mathbf{B}}_b = [\bar{\mathbf{B}}_{(b-1) \times H_0+1} \bar{\mathbf{B}}_{(b-1) \times H_0+2} \cdots \bar{\mathbf{B}}_{(b-1) \times H_0+H_0}]^T, 1 \leq b \leq B$$

为  $H_0 \times N$  块矩阵, 表示中间集中进行稀疏编码运算的受试者样本。

目标集  $\mathbf{A} = \{\bar{\mathbf{A}}_h\}_{h=1}^H = \{\tilde{\mathbf{A}}_m\}_{m=1}^M$ , 其中样本  $\bar{\mathbf{A}} = [a_{i1} \ a_{i2} \ \cdots \ a_{iN}]^T, 1 \leq i \leq H$ ,

$H_0 \times N$  块矩阵如下:

$\tilde{\mathbf{A}}_m = [\bar{\mathbf{A}}_{(m-1) \times H_0+1} \bar{\mathbf{A}}_{(m-1) \times H_0+2} \cdots \bar{\mathbf{A}}_{(m-1) \times H_0+H_0}]^T, 1 \leq m \leq M$  表示目标集中进行稀疏编码运算的受试者样本。

样本段对应标签及受试者样本对应标签分别为  $\mathbf{C} = \{c_h\}_{h=1}^H$  和  $\tilde{\mathbf{C}} = \{\tilde{c}_m\}_{m=1}^M$ 。目标集样本总数为  $H$ , 每个样本的特征数为  $N$ , 总共有  $M$  个受试者, 即每个受试者有  $H_0 = H/M$  个样本。

## 1.2 卷积稀疏编码学习

在卷积稀疏编码学习中, 给定  $L$  个受试者训练样本  $\{\tilde{\mathbf{Y}}_l\}_{l=1}^L$ , 最小化目标式如下:

$$\arg \min_{e,d} \frac{1}{2} \sum_{l=1}^L \left\| \tilde{\mathbf{Y}}_l - \sum_{k=1}^K \mathbf{d}_k * \mathbf{e}_{l,k} \right\|_2^2 + \eta \sum_{l=1}^L \sum_{k=1}^K \|\mathbf{e}_{l,k}\|_1$$

s. t.  $\|\mathbf{d}_k\|_2 \leq 1, \forall k = \{1, \dots, K\}$  (1)

其中,  $\tilde{\mathbf{Y}}$  为  $H_0 \times N$  块矩阵,  $\mathbf{d}_k$  为卷积核,  $\mathbf{e}_{l,k}$  为  $H_0 \times N$  特征映射矩阵,  $\eta$  为大于 0 的正则化因子。

求解上述优化问题可通过经典的交替方向乘子法 ADMM<sup>[17]</sup> 进行求解。为进一步提升求解算法效率, 可采用快速傅里叶域运算<sup>[18]</sup> 及交换迭代次序法<sup>[19]</sup>。

## 1.3 第 1 阶段迁移学习算法

首先重复训练多组卷积核  $\{\hat{\mathbf{d}}_k\}_1, \{\hat{\mathbf{d}}_k\}_2, \dots, \{\hat{\mathbf{d}}_k\}_R$ , 然后基于卷积核将中间集  $\mathbf{B}$  编码为特征映射  $\mathbf{B}'$ , 编码运算如式(2)所示。固定卷积核  $\{\mathbf{d}_k\} = \{\hat{\mathbf{d}}_k\}$ , 特征映射  $\mathbf{e}_{b,k}$  可通过 ADMM 迭代优化生成。 $\mathbf{B}'$  为中间集  $\mathbf{B}$  所有受试者特征映射  $\mathbf{e}_{b,k}$  组成的数据集。

$$\mathbf{B}' = \arg \min_e \frac{1}{2} \sum_{b=1}^B \left\| \bar{\mathbf{B}}_b - \sum_{k=1}^K \hat{\mathbf{d}}_k * \mathbf{e}_{b,k} \right\|_2^2 + \eta \sum_{b=1}^B \sum_{k=1}^K \|\mathbf{e}_{b,k}\|_1 \quad \forall k = \{1, \dots, K\}$$
 (2)

通过 LOSO 交叉验证计算  $\mathbf{B}'$  分类准确率, 分类器 LSVM。最优卷积核  $\hat{\mathbf{d}}_{\text{opt}}$  可通过不同卷积核组对应的平均分类准确率大小进行筛选。

第 1 阶段迁移学习伪代码如算法 1 所示。

## 1.4 基于深度样本集成学习机制的第 2 阶段迁移学习算法

该算法主要包括深度样本学习机制、样本/特征同时选择机制和决策层融合机制。该算法的主要流程如下: 首先构建深度样本空间  $\mathbf{dA} = \{\mathbf{dA}_f\}_{f=0}^F$ , 其中原始样本空间  $\mathbf{dA}_0$  即为目标集  $\mathbf{A}$ , 其余样本空间层  $\mathbf{dA}_f, 1 \leq f \leq F$  可

## 算法 1: 第 1 阶段迁移学习算法

输入: 公共数据集  $\mathbf{S}$ , 中间特征数据集  $\mathbf{B}$

输出: 最优卷积核  $\hat{\mathbf{d}}_{\text{opt}}$

步骤:

**For** kernel\_num = 2, ...,  $K$  **do**

  随机初始化卷积核  $\mathbf{d}_k^0$ ;

  重复训练生成多组卷积核  $\{\hat{\mathbf{d}}_k\}_1, \{\hat{\mathbf{d}}_k\}_2, \dots, \{\hat{\mathbf{d}}_k\}_R$ ;

  对中间集  $\mathbf{B}$  进行编码, 则最优卷积核;

$\hat{\mathbf{d}}_{\text{opt}} = \arg \max_{\hat{\mathbf{d}}_k} (\text{acc}(\mathbf{B}, \{\hat{\mathbf{d}}_k\}_r))$ ;

**End for**

通过迭代均值聚类构建<sup>[20]</sup>; 然后基于最优卷积核  $\hat{\mathbf{d}}_{\text{opt}}$  对  $\mathbf{dA}_f$  进行编码, 生成对应的特征映射; 接下来执行并行优选生成新的特征映射空间  $\mathbf{dG}_f^0$ 。当进行训练测试分类时, 可得到同一受试者不同样本空间层的预测标签  $\mathbf{L}_f$ , 受试者最终预测标签  $\mathbf{L}$  可通过多层标签  $\{\mathbf{L}_f\}_{f=0}^F$  融合得到。

1) 深度样本学习机制

迭代均值聚类的目标函数如下:

$$\arg \min_S \sum_{i=1}^I \sum_{\bar{\mathbf{A}} \in S_i} \|\bar{\mathbf{A}} - \mu_i\|^2$$
 (3)

其中,  $\bar{\mathbf{A}}$  是受试者  $\tilde{\mathbf{A}}_m$  的样本, 目标集  $\mathbf{A} = \{\tilde{\mathbf{A}}_m\}_{m=1}^M$  中的受试者  $\tilde{\mathbf{A}}_m$  被初始化为多个簇  $\{S_i\}_{i=1}^I, \mu_i$  为簇均值, 初始化簇  $s_1, \dots, s_I$  的  $I$  个均值  $m_1^{(1)}, \dots, m_I^{(1)}$ 。则迭代优化后最终的  $I$  个均值  $\hat{m}_1, \dots, \hat{m}_I$  可构造一层深度样本空间。将变量  $I$  设置成  $F$  个不同的值, 则可构建  $F$  层样本空间  $\mathbf{dA}_f, 1 \leq f \leq F$  称为深度样本空间。每个簇可以看作样本包络。

2) 卷积稀疏编码及样本/特征同时选择机制

基于最优卷积核  $\hat{\mathbf{d}}_{\text{opt}}$  对深度样本空间进行编码, 如下所示:

$$\mathbf{dE}_f = \arg \min_e \frac{1}{2} \|\hat{\mathbf{d}}_{\text{opt}} * \mathbf{e} - \mathbf{dA}_f\|_2^2 + \eta \|\mathbf{e}\|_1$$
 (4)

通过 ADMM 迭代优化  $\mathbf{e}$  可生成目标集受试者特征映射集合  $\mathbf{dE}_f$ 。  $\mathbf{dE}_f = \{\mathbf{dE}_{f_m}\}_{m=1}^M$  可通过向量化扩展生成  $\mathbf{dG}_f = \{\mathbf{dG}_{f_m}\}_{m=1}^M$ , 其中:

$$\mathbf{dG}_{f_m} = \text{RESHAPE}(\mathbf{dE}_{f_m} \quad H' \times N \quad 1 \times N'),$$

$$H' = \begin{cases} H_0, & f = 0 \\ I, & 1 \leq f \leq F \end{cases}, \quad N' = H' \times N$$
 (5)

对深度空间每一层而言,  $\mathbf{dG}_f$  被划分为训练集  $\mathbf{dG}_{\text{train}}$

和测试集  $dG_{\text{test}}$ :

$$dG_f = \begin{bmatrix} dG_{\text{train}} \\ dG_{\text{test}} \end{bmatrix} =$$

$$\begin{bmatrix} \vec{\Phi}_{\text{train}(1)} & \vec{\Phi}_{\text{train}(2)} & \dots & \vec{\Phi}_{\text{train}(N')} \\ \vec{\Phi}_{\text{test}(1)} & \vec{\Phi}_{\text{test}(2)} & \dots & \vec{\Phi}_{\text{test}(N')} \end{bmatrix}$$

其中,  $\Phi$  为特征列向量, 可通过 relief 算法计算

$dG_{\text{train}} = [\vec{\Phi}_{\text{train}(1)} \vec{\Phi}_{\text{train}(2)} \dots \vec{\Phi}_{\text{train}(N')}]$  中特征列向量所对应的权重向量  $\vec{W} = [w_1 \ w_2 \ \dots \ w_{N'}]^{[21]}$ 。

经过样本/特征同时选择  $Q$  个特征之后的数据集如下:

$$dG_f^Q = \begin{bmatrix} dG_{\text{train}}^Q \\ dG_{\text{test}}^Q \end{bmatrix} =$$

$$\begin{bmatrix} \vec{\Phi}_{\text{train}(1)} & \vec{\Phi}_{\text{train}(2)} & \dots & \vec{\Phi}_{\text{train}(N')} \\ \vec{\Phi}_{\text{test}(1)} & \vec{\Phi}_{\text{test}(2)} & \dots & \vec{\Phi}_{\text{test}(N')} \end{bmatrix} \times$$

$$\begin{bmatrix} \vec{\xi}_1^T & \vec{\xi}_1^T & \dots & \vec{\xi}_Q^T \end{bmatrix} \quad (6)$$

其中,  $\vec{\xi}_q = [0 \dots 0 \ 1 \ 0 \dots 0]$  且  $index$  是  $Q$  个最大权重

所对应特征向量的列标记。

### 3) 决策层融合机制

深度样本空间的训练部分  $dG_{\text{train}}^Q$  采用线性支持向量机 (linear support vector machines, LSVM) 训练模型, 并得到测试部分  $dG_{\text{test}}^Q$  的标签  $L_f$ , 则多层预测标签融合如下:

$$L = \text{bagging}(L_0, L_1, \dots, L_F) = \text{sign}\left(\frac{1}{F} \sum_{f=0}^F L_f - \frac{1}{2}\right) \quad (7)$$

其中,  $L_f$  为 TRUE=1 或 FALSE=0。

## 1.5 整体算法流程

首先, 公共语音数据集  $S$  作为源域数据集被加噪扩展, 提取多类型特征构造特征数据集  $Y = \{\bar{Y}_l\}_{l=1}^L$ , 然后基于快速卷积稀疏编码算法训练多组卷积核  $\{\hat{d}_k\}_1, \{\hat{d}_k\}_2, \dots, \{\hat{d}_k\}_R$ 。接着将中间集  $B$  编码生成不同卷积核对应的特征映射  $B'$ , 再通过特征映射的分类准确率筛选出最优卷积核  $\hat{d}_{\text{opt}}$ 。

基于目标集  $A$ , 采用迭代均值聚类算法设计深度样本学习机制, 构造深度样本空间  $dA = \{dA_f\}_{f=0}^F$ 。基于最优卷积核  $\hat{d}_{\text{opt}}$  对深度样本空间进行编码生成对应特征映射空间  $dE = \{dE_f\}_{f=0}^F$ , 再向量化扩展为  $dG = \{dG_f\}_{f=0}^F$ 。

对于每一层样本空间,  $dG_f$  可按照 LOSO 交叉验证方式划分为训练集  $dG_{\text{train}}$  和测试集  $dG_{\text{test}}$ ; 基于 relief 算法实现样本/特征同时选择<sup>[16]</sup>, 从而计算训练集特征权重  $\vec{W}$ , 并根据权重大小优选训练集和测试集的前  $Q$  个特征构建新的训练集  $dG_{\text{train}}^Q$  和测试集  $dG_{\text{test}}^Q$ 。之后, 训练

LSVM 分类器模型计算测试集  $dG_{\text{test}}^Q$  的预测标签  $L_f$ , 遍历每个受试者可得到其对应预测标签  $L_{fm}, 1 \leq m \leq M$ 。最后测试集的最终预测标签通过多层空间的预测标签  $\{L_f\}_{f=0}^F$  融合决定。

双阶段迁移学习算法 (TSTL) 伪代码如算法 2 所示。

### 算法 2: TSTL 算法

输入: 公共数据集  $S$ , 中间数据集  $B$ , 目标集  $A$

输出: 准确率, 灵敏度, 特异度

步骤:

第 1 阶段:

基于算法 1 得到最优卷积核  $\hat{d}_{\text{opt}}$ ;

第 2 阶段:

基于最优卷积核  $\hat{d}_{\text{opt}}$  将目标集  $A$  编码成特征映射  $E$  再向量化得到  $G$ ;

**For**  $m = 1, \dots, M$  **do**

将  $G$  划分为训练集  $G_{\text{train}}$  和测试集  $G_{\text{test}}$ , 其中

$$G_{\text{train}} = \{\vec{G}_1, \vec{G}_2, \dots, \vec{G}_M\}, \vec{G}_m \notin G_{\text{train}}, G_{\text{test}} = \vec{G}_m;$$

基于训练集  $G_{\text{train}}$  计算对应权重向量, 并根据权重大小筛选得到新的训练集  $G_{\text{train}}^Q$  及测试集  $G_{\text{test}}^Q$ ;

建立 LSVM 模型预测得到  $G_{\text{test}}^Q$  的预测标签  $L_m$

**End for**

综合考虑目标集所有受试者得到最终测试准确率, 灵敏度, 特异度。

具有深度样本机制的双阶段迁移学习算法 (TSTL&DISC) 伪代码如算法 3 所示。

## 2 实验结果与分析

为进一步验证本文理论方法的有效性本文组织了 3 组实验, 第 1 组消融法实验验证了双阶段迁移学习的有效性以及深度样本学习机制的有效性; 第 2 组实验分析了深度样本层数以及超参数对算法性能的影响; 第 3 组实验将本文算法与经典深度学习以及近年来代表性文献方法进行了比较。

### 2.1 实验条件

#### 1) 数据集

##### (1) 源域数据集

标准的 TIMIT 数据集包含了总共 6 300 个句子, 由 630 名受试者每人提供 10 个句子, 目前网上容易获取的 TIMIT 数据集版本有 240 个样本, 有 40 名男性受试者和 40 名女性受试者, 每人提供 3 个句子。



## 算法 3: TSTL&amp;DISC 算法

输入: 公共数据集  $S$ , 中间数据集  $B$ , 目标集  $A$

输出: 准确率, 灵敏度, 特异度, MCC

步骤:

第 1 阶段:

基于算法 1 得到最优卷积核  $\hat{d}_{opt}$ ;

第 2 阶段:

通过迭代均值聚类构造深度样本空间  $dA$ ;

**For**  $f=1, \dots, F$  **do**

基于最优卷积核  $\hat{d}_{opt}$  将目标集  $dA$  编码成特征映射  $dE$

再向量化得到  $dG$ ;

**For**  $m=1, \dots, M$  **do**

将  $dG$  划分为训练集  $dG_{train}$  和测试集  $dG_{test}$ , 其中

$dG_{train} = \{d\vec{G}_1, d\vec{G}_2, \dots, d\vec{G}_M\}$  and  $d\vec{G}_m \notin dG_{train}$ ,  $dG_{test} = d\vec{G}_m$ ;

基于训练集  $dG_{train}$  计算对应权重向量  $\vec{W}$ , 并根据权重大小筛选得到新的训练集  $dG_{train}^Q$  及测试集  $dG_{test}^Q$ ;

建立 LSVM 模型预测得到  $dG_{test}^Q$  的多层预测标签  $L_{fm}$

**End for**

**End for**

**For**  $m=1, \dots, M$  **do**

多层融合决策形成最终标签  $L_m = \text{sign}\left(\frac{1}{F} \sum_{f=0}^F L_{fm} - \frac{1}{2}\right)$ ;

**End for**

综合考虑目标集所有受试者得到最终测试准确率, 灵敏度, 特异度。

## (2) 目标域数据集

如表 1 所示, 目标域数据集选择描述如下:

①目标数据集 1 为国际流行的 PD 语音公开数据集-Sakar 数据集, 由 Sakar 等<sup>[22]</sup> 提供, 总共包含 40 名受试者, 其中 20 名 PD 患者(6 名女性, 14 男性), 20 名健康人(10 名女性, 10 名男性), 每人提供 26 个语音样本段, 且每个语音样本段包含不同的发音内容, 具体有连续的元音字母发音, 数字发音, 单词发音及短句发音。每个语音段都提取了 26 个特征组成一个特征向量。

②目标数据集 2(研究者自采国内 PD 数据集): 由陆军军医大学第一附属医院的患者提供, 数据总共包含 90 名受试者, 其中 36 名未受治疗 PD 患者(16 名女性, 20 名男性)。54 名已接受治疗 PD 患者(27 名女性, 27 名男性), 未受治疗组与已接受治疗组的年龄分布均衡。每人提供 13 个语音段, 每个语音段提取 26 个特征, 语音信息由录音笔设备(型号: SONY ICD-SX2000)采集, 录音笔放置在距离受试者 15 cm 处, 语音采样率为 44.1 kHz, 量化精

度为 16 bit。受试者朗读了 10 个数字‘1’, ‘2’, ‘3’, ‘4’, ‘5’, ‘6’, ‘7’, ‘8’, ‘9’, ‘10’, 及元音‘a’, ‘o’, ‘u’, 语音特征采用 Praat 软件提取。

表 1 数据集信息

Table 1 Dataset information

数据集名称	样本数	受试者数	语料	数据类型
源域数据集 (TIMIT)	240	80	短句	原始语音信号
目标集 1 (Sakar 数据集)	1 040	40	数字 1~10, 元音‘a’, ‘o’ ‘u’ 及短句	多类型特征数据
目标集 2, 中间集 1(自采集)	1 170	90	数字 1~10, 元音‘a’, ‘o’ ‘u’ 及短句	原始语音信号
中间集 2 (复发性麻痹)	5 226	402	元音‘a’, ‘i’ ‘u’ 及短句	原始语音信号
中间集 3 (声带息肉)	1 144	88	元音‘a’, ‘i’ ‘u’ 及短句	原始语音信号

## (3) 中间集

中间集选择描述如下:

①中间集 1: 选取与目标数据集 2 相同的数据集, 为国内自采数据集, 针对目标数据集 2 进行实验时中间集 1 不再作为中间集被选用。

②中间集 2 和 3 均取自德国萨尔布吕肯大学语音数据库(SVD)<sup>[16]</sup>, 该集包含 2 000 多个文件, 记录了 71 种具体语音障碍数据, 每个语音障碍以正常, 高音调, 低音调及低到高音调记录了 3 个持续元音/a/, /i/和/u/, 短句“早上好, 你好吗?”。文件长度在 1~4 s 之间, 以 50 kHz 采样, 量化位数为 16 位。

其中中间集 2 的患者数据部分为 SVD 的复发性麻痹症数据集, 该集由 213 名患者提供, 丢弃 12 份异常数据(无法提取特征), 留下 201 份, 未患病数据部分为 SVD 数据集中前 201 名正常受试者数据。

中间集 3 的患者数据部分为 SVD 声带息肉数据集, 该集由 45 名患者提供, 丢弃 1 份异常数据(无法提取特征), 留下 44 份, 未患病数据部分为 SVD 数据集中前 44 名正常受试者数据。

## 2) 实验组织方式

双阶段迁移学习的过程是由源域数据集到中间集再到目标域数据集。本文针对目标数据集 1 分为 3 种方式实验: 源域数据集+中间集 1+目标域数据集 1, 源域数据集+中间集 2+目标域数据集 1, 源域数据集+中间集 3+目标域数据集 1。针对目标数据集 2 分为两种方式: 源域数据集+中间集 2+目标域数据集 2, 源域数据集+中间集 3+目标域数据集 2。由于中间集中每个受试者的语音段都是 13 个

样本,为保持迁移的有效性,因此目标集的每个受试者也对对应选取了数字+元音共13个样本,训练基于的样本尺寸为13×26。深度样本空间第1~12层的生成方式是将第0层(原始数据集)的13样本分别聚类成12,11,10,9,8,7,6,5,4,3,2个样本。数据集信息如表1所示。

数据验证方法采用 LOSO 交叉验证法。采用交叉验证法的原因是数据不充分条件下能充分反映算法的实际效果。当前大多数 PD 语音诊断算法采用 k-fold 及 hold-one-out。与这两种交叉验证方法不同,LOSO 能确保训练集和测试集分别属于不同的患者,分类准确率与实际的临床诊断结果更加接近。

### 3) 实验平台及参数设置

设备软硬件环境:CPU (Intel i7-8565U 1.99 GHz), 8 GB 内存, MATLAB R2018b。参数设置:重复训练的卷积核组数为10, ADMM 主迭代数, 特征映射学习迭代数, 核学习迭代数分别为100, 10及10。松弛因子  $\gamma = 1$ , 卷积核数范围取值2~8, 卷积核尺寸为8×8。

4) 评价准则评价准则包含准确率(Acc.), 灵敏度(Sen.), 特异度(Spe.) 和 Matthews 相关系数(MCC)。

$$Acc. = \frac{TP+TN}{TP+FP+TN+FN}$$

$$Sen. = \frac{TP}{TP+FN}$$

$$Spe. = \frac{TN}{FP+TN}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$
(8)

其中, TP 表示真阳, FP 表示假阳, TN 表示真阴, FN 表示假阴。

## 2.2 有效性验证

### 1) 双阶段迁移学习有效性

为了验证本文设计的双阶段迁移学习的有效性,将该算法与单阶段迁移学习进行比较。表2所示,基于 Sakar 数据集在3个不同中间集下的双阶段迁移学习算法(TSTL)的分类性能比较。

表2中未使用中间集的单阶段迁移学习算法记作(one-stage transfer learning, OSTL), 中间集1是PD语音集,与目标集1相似,且受试者的测试发音语料也保持一致,因此基于中间集1的分类准确率优于不加中间集的分类准确率。中间集2和中间集3不是PD语音集,且受试者测试发音也与目标集并不一致,因此作为中间集不能起到明显的正迁移作用。这说明,虽然加中间集的双阶段迁移学习机制是有效的,但是需要构造与目标集正相关的中间集,才能有助于实现正迁移,显著提高目标集的分类准确性。

表2 基于 Sakar 数据集在不同中间集下的双阶段迁移学习性能比较

Table 2 Comparison of two-stage transfer learning performance based on Sakar dataset among different intermediate datasets %

方法	Acc.	Sen.	Spec.
OSTL	86.0	91.0	81.0
TSTL(中间集1)	<b>90.0</b>	<b>90.0</b>	<b>90.0</b>
TSTL(中间集2)	82.5	90.0	75.0
TSTL(中间集3)	80.0	70.0	90.0

### 2) 深度样本学习纠错机制有效性

基于深度样本学习进行纠错包含两种纠错类型。以3层深度样本空间为例,第1种:正确分类标签应为1,原始样本空间预测为0,第1层样本空间和第2层样本空间均预测为1从而实现纠错;反之第2种:正确分类标签应为0,原始样本空间预测为1,第1层样本空间和第2层样本空间均预测为0从而实现纠错。

## 2.3 参数分析

### 1) 深度样本融合层数的影响

表3表示基于 Sakar 目标数据集在中间集1下的准确率,灵敏度,特异度及 Matthews 相关系数,融合层数从原始层至13层。由实验结果可见,采用深度样本集成学习机制后的结果优于原始层的结果,当融合层数为9,准确率达到最大值99.5%,说明取得最优性能需要合理选择融合层数。

表3 基于 Sakar 数据集的深度空间融合层性能(中间集1)  
Table 3 Deep space fusion layer performance (intermediate set 1)

融合层数	Acc./%	Sen./%	Spec./%	MCC
原始层	90.0	90.0	90.0	0.800
2	84.8	84.3	85.3	0.695
3	94.5	94.0	95.0	0.890
4	92.8	91.0	94.5	0.856
5	97.8	97.0	98.5	0.955
6	97.1	96.0	98.3	0.943
7	98.5	98.0	99.0	0.970
8	98.6	98.5	98.8	0.973
<b>9</b>	<b>99.5</b>	<b>99.5</b>	<b>99.5</b>	<b>0.990</b>
10	99.0	98.8	99.3	0.980
11	99.3	99.5	99.0	0.985
12	99.0	98.8	99.3	0.980
13	99.3	98.5	100.0	0.985

因为融合层数为奇数时的性能大都优于相邻偶数层,选取融合层数为奇数时的数据对 3 个中间集进行分析比较。

如图 1 所示基于 Sakar 数据集在不同中间集下的性能比较,随着层数的增加,3 个中间数据集下的分类准确率均近似单调递增。当层数为 9 时,中间集 1 的准确率达到了最大值 99.5%,但中间集 2 为 98.3%,中间集 3 为 98.8%。随着层数的继续增加,中间集 1 的准确率在 99.0% 和 99.3% 的区间摆动,而中间集 2 会继续增加到 99.5%,中间集 3 会继续增加到 99.0%。总之,中间层 1 的性能优于中间集 2 和 3,说明选择中间集 1 相比中间集 2 和 3 更有效。可能的原因是,中间集 1 为自采 PD 语音数据集,与 Sakar 数据集的疾病类型相同,对正迁移有更大作用。图 2 为基于自采数据集在不同中间集下的性能比较,也体现了分类准确率随融合层数迅速提高的趋势,且在融合层数为 9 时达到了最大值 88.0%。值得注意的是,当融合层数较多时,不同中间集的分类准确率接近,即无论中间集的质量如何,当深度样本空间足够深也仍然能取得较高的准确率,这一点充分证实了深度样本学习机制的有效性和必要性。

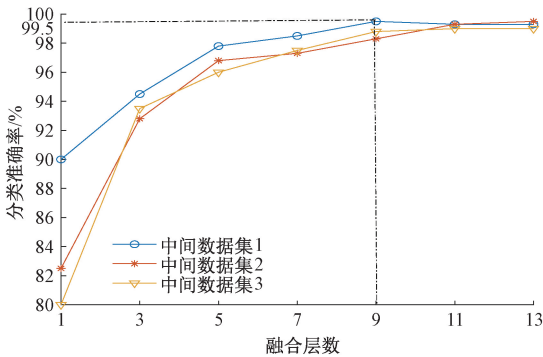


图 1 基于 Sakar 数据集在不同中间集下的性能比较  
Fig. 1 Performance comparison based on Sakar dataset for different intermediate datasets

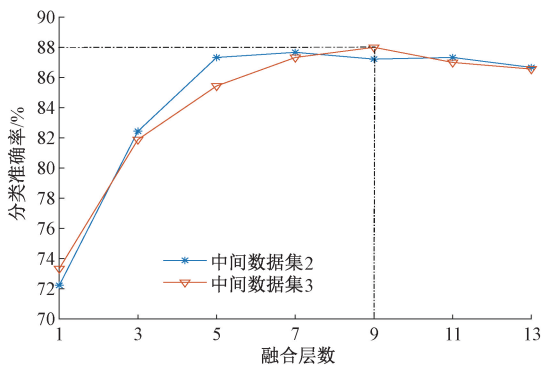


图 2 基于自采数据集在不同中间集下的性能比较  
Fig. 2 Performance comparison based on self-connected dataset for different intermediate datasets

2) 超参数对算法性能的影响

本文算法的可调超参数包括卷积核数,特征映射数及语音段/特征同时选择的特征数。但在 TSTL&DISC 算法中,每一层深度样本空间都有 3 个对应的超参数,随着融合层数的增加,特别是达到 13 层,超参数的规模会变得很大,调参的时间复杂度会变得难以容忍。因此,实验中采用简化的参数调节方式,即将每一层的最优参数作为系统的参数进行统一调节。

如表 4 所示深度样本空间第 0~4 层的超参数对评估性能的影响。具体可见不同深度样本空间层准确率最高时的超参数设置,即原始层当卷积核数为 8,特征映射数为 6,特征选择数为 47 时准确率达到最大值 90%,即算法 TSTL 的分类结果;第 1 层当卷积核数,特征映射数,特征选择数分别为 4,2,4 时准确率达到最大值 77.5%;而深度空间第 2 层当超参数分别为 6,5,18 时准确率达到最大值 75%。由表 4 结果可见,不同层取得最优性能所对应的参数并不相同,因此若要取得融合之后的良好性能需要在每一层选择合适参数。

表 4 基于 Sakar 数据集超参数对分类性能的影响(中间集 1)  
Table 4 Influence of hyper-parameters on classification performance based on Sakar dataset (Intermediate set 1)

层编号	超参数			评价指标			
	卷积核数	特征映射数	特征选择数	Acc. /%	Sen. /%	Spec. /%	MCC
原始层	8	6	47	90.0	90.0	90.0	0.800
第 1 层	4	2	4	77.5	70.0	85.0	0.556
第 2 层	6	5	18	75.0	80.0	70.0	0.503
第 3 层	4	2	71	80.0	75.0	85.0	0.603
第 4 层	4	3	1	87.5	90.0	85.0	0.751

2.4 算法比较

在 2.3 节 1) 中,发现基于 Sakar 目标数据集在中间集 1 下达到,最优性能的融合层数是 9,中间集 2 时则是 13,中间集 3 时则是 11。融合层数太多会导致需要调节过多的超参数,模型复杂且时间代价大,因此在小样本的目标集上容易过拟合且不易于临床应用。因此下面算法比较中,本文算法中深度样本空间的融合层数定为 5,从而较好的平衡系统复杂度和训练准确率,提高系统泛化能力。

如表 5 所示,本文选取了一些主流常用算法进行基于 Sakar 数据集进行性能比较。选择 K 近邻(K-nearest neighbor, KNN)和支持向量机(support vector machines, SVM)算法进行比较是因为是常用的分类器。如表中结果所示,只使用最基本的分类器进行分类准确率并不高。

选择深度置信网络(deep belief networks, DBN)和卷积神经网络(convolutional neural network, CNN)进行比较是因为其是有代表性的深度学习方法。如表5中结果所示,其由于训练样本过小性能也不佳。DBN+SVM(TL)是单阶段深度迁移学习算法,是用TIMIT公共数据集预训练DBN网络,再将目标集输入到训练好的DBN网络,对输出进行分类。类似的,Autoencoder+SVM(TL)方法是用TIMIT公共数据集预训练自动编码器,再用训练好的自动编码器重新编码目标集,再进行分类。选用这两种方法是由于他们都是将TIMIT数据集的信息迁移到目标集,与本文算法的第1阶段迁移学习算法具有相关性。实验表明,TSTL算法和TSTL&DISC算法优于其他比较算法,因此本文提出的算法是有效的。(这里分类器选用SVM)。

表5 基于Sakar数据集的主流算法比较

Table 5 Comparison of main algorithms based on Sakar dataset %

方法	Acc.	Sen.	Spe.
KNN	52.5	55.0	50.0
SVM(RBF)	67.5	80.0	55.0
DBN	54.6	52.4	56.8
CNN	60.0	63.0	57.0
DBN+SVM(TL)	55.5	60.0	51.0
Autoencoder+SVM(TL)	72.5	75.0	70.0
TSTL(中间集1)	90.0	90.0	90.0
TSTL(中间集2)	82.5	90.0	75.0
TSTL(中间集3)	80.0	70.0	90.0
TSTL&DISC(中间集1)	97.8	97.0	98.5
TSTL&DISC(中间集2)	96.8	95.0	98.5
TSTL&DISC(中间集3)	96.0	94.5	97.5

如表6所示,这里选取近年来有代表性的PD语音诊断算法进行比较。首先,表中比较文献均采用LOSO交叉验证方法,与本文交叉验证方法完全相同,且都基于同一目标数据集(Sakar数据集),因此结果比较有意义。通过比较发现,TSTL&DISC在中间集1上取得了97.8%的平均分类准确率,优于其他文献算法。这一优势也体现在敏感度和特异度上。此外,还可以发现,即使采用相关性不大的中间集,本文算法依然能取得较高的准确率,达到96%。这就说明,即使采用质量较差的源数据集作为中间集,本文算法也依然能够获得较单阶段迁移学习算法或者非迁移学习算法更优的准确率,这也间接说明深度样本学习机制是很有效的。

表6 基于Sakar数据集的相关文献算法比较

Table 6 Comparison of relevant literatures based on Sakar dataset %

来源	方法	Acc.	Sen.	Spe.
文献[16]	KNN+SVM	55	60	50
文献[23]	4 Feature Selection Methods+6 Classifiers	57.5	54.28	80
文献[24]	MENN+RF with MENN	81.5	92.50	70.50
文献[25]	HFCC+SVM	87.5	90.00	85.00
文献[26]	Hybrid feature learning+SVM	82.50	85.00	80.00
文献[27]	MFCC+SVM	82.5	80.0	85.0
文献[28]	LDA-NN-GA	95	95	95
文献[29]	$X^2$ Model + Sample Selection + NN	97.5	100	95
文献[30]	CSC+DA	97.5	97.5	97.5
建议算法	TSTL&DISC(中间集1)	<b>97.8</b>	<b>97.0</b>	<b>98.5</b>
建议算法	TSTL&DISC(中间集2)	96.8	95.0	98.5
建议算法	TSTL&DISC(中间集3)	96.0	94.5	97.5

如表7所示,本文提出的算法在国内自采数据集上也达到了较好的效果。与直接使用KNN和SVM分类器相比,本文提出算法的平均分类准确率取得了明显提高,可达87.3%,证明了本文算法在国内PD语音数据集上的有效性。

表7 基于自采数据集的主流算法比较

Table 7 Comparison of mainstream algorithms based on Self-collected dataset %

方法	Acc.	Sen.	Spe.	
KNN	55.6	0	92.6	
SVM(RBF)	60.0	0	100.0	
建议算法	TSTL&DISC(中间集2)	<b>87.3</b>	<b>71.1</b>	<b>98.2</b>
建议算法	TSTL&DISC(中间集3)	85.4	69.2	96.3

## 2.5 时间复杂度分析

卷积核训练筛选完成后可在服务器平台生成固定依赖数据而不需要重复计算,实际工程应用的时间复杂度主要取决于深度样本空间构造及每层样本空间在第2阶段迁移学习中的时间消耗。如表8所示,基于Sakar数据集在中间集1下比较深度样本空间不同层的时间代价。从低层到高层,耗时越来越短,这是由于随着层数越高深度样本聚类之后的样本数越少,即受试者对应的样本尺寸越小,处理时间越短。由于bagging的计算复杂度可忽略不计,各层的耗时累计约等于整个算法的耗时,大约为



21.75 s。由于不同目标集对应样本空间受试者(subject)的尺寸及算法处理均相同,因此时间消耗与受试者数量成正比,即实际应用中,只需等待不到 30 s 能够得到 40 人诊断结果(目标集 1 包含 40 名受试者),不到 1 min 得到 90 人诊断结果(目标集 2 包含 90 名受试者),完全满足临床应用的时间要求。此外,考虑到将算法语言(MATLAB,Python 等)转化为更快速的工程语言(C,Java 等),时间代价还将显著下降。

表 8 基于 Sakar 数据集深度样本空间不同层的算法时间复杂度比较(中间集 1)

Table 8 Comparison of time complexity of algorithms in different layers of deep instance space based on Sakar dataset (Intermediate set 1)

层编号	第 0 层 (原始层)	第 1 层	第 2 层	第 3 层	第 4 层
耗时/s	5.65	4.77	4	3.33	4

### 3 结 论

PD 语音识别的有效性和重要性使得其成为国内外研究热点和难点之一,但始终受制于小样本问题。当前相关的 PD 分类方法仅基于本地 PD 语音数据集,对解决小样本问题效果有限。现有 PD 语音识别算法的第 1 个局限性是很少考虑 1 个以上数据集的迁移学习。第 2 个局限性是很少考虑高性能的样本编码算子。第 3 个局限性是其没有考虑提取样本间的结构信息。

为了解决上述局限性问题,本文提出了一种新颖的 PD 语音识别方法-双面双阶段均值聚类包络和卷积稀疏迁移学习算法。首先,通过双阶段卷积稀疏编码挖掘源域语音结构信息并迁移到目标域使信息更完备,完成两阶段迁移学习。然后,构造深度样本集成学习机制是使训练样本数据信息更加丰富,完成样本深度学习。文末组织了多组实验,基于经典国际公共数据集(Sakar)和自采数据集算法性能均取得较大改善,其中 Sakar 数据集的平均准确率达 97.8%。

需要说明的是,本文方法成功实现了连续迁移两个 PD 语音源数据集,并尝试了 3 种中间源数据集。迁移单个源数据集和迁移两个源数据集在方法上具有本质区别,因此本文方法相较于现有的迁移学习算法有较大突破。但是迁移两个及以上源数据集在方法上没有本质区别,因此简单重复本文算法,可以迁移两个以上的源数据集,所以本文工作为迁移多个 PD 语音源数据集提供了成功案例和方法参考。

### 参考文献

- [1] EMAMZADEH F N, SURGUCHOV A. Parkinson's disease: Biomarkers, treatment, and risk factors [J]. *Frontiers in Neuroscience*, 2018, DOI: 10.3389/fnins.2018.00612.
- [2] KODRASI I, BOURLARD H. Spectro-temporal sparsity characterization for dysarthric speech detection [J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2020, 28:1210-1222.
- [3] NP N, SCHULLER B, ALKU P. The detection of Parkinson's disease from speech using voice source information [J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2021, 29: 1925-1936.
- [4] MAFFIA M, DE MICCO R, PETTORINO M. Speech rhythm variation in early-stage Parkinson's disease: A study on different speaking tasks [J]. *Frontiers in Psychology*, 2021, DOI:10.3389/fpsyg.2021.668291.
- [5] QUAN C, REN K, LUO Z. A deep learning based method for Parkinson's disease detection using dynamic features of speech [J]. *IEEE Access*, 2021, 9: 10239-10252.
- [6] AHMED I, ALJAHDALI S, KHAN M S. Classification of parkinson disease based on Patient's voice signal using machine learning [J]. *Intelligent Automation and Soft Computing*, 2022, 32(2): 705-722.
- [7] NASEER A, RANI M, NAZ S, et al. Refining Parkinson's neurological disorder identification through deep transfer learning [J]. *Neural Computing and Applications*, 2020, 32(3): 839-854.
- [8] 孟浩,田洋,孙宇婷,等. 全局注意力关系网络的小样本船舶识别 [J]. *仪器仪表学报*, 2021, 42(12): 220-227.
- [9] MENG H, TIAN Y, SUN Y T, et al. Few shot ship recognition based on universal attention relationnet [J]. *Chinese Journal of Scientific Instrument*, 2021, 42(12): 220-227.
- [9] 陈保家,陈学力,肖文荣,等. 小样本下滚动轴承故障的多源域迁移诊断方法 [J]. *电子测量与仪器学报*, 2022, 36(2): 219-228.

- CHEN B J, CHEN X L, XIAO W R, et al. Multi-source domain transfer diagnosis method for rolling bearing faults under small samples [J]. *Journal of Electronic Measurement and Instrumentation*, 2022, 36 (2): 219-228.
- [10] REY-OTERO I, SULAM J, ELAD M. Variations on the convolutional sparse coding model [J]. *IEEE Transactions on Signal Processing*, 2020, 68:519-528.
- [11] VUTTIPITTAYAMONGKOL P, ELYAN E. Improved overlap-based undersampling for imbalanced dataset classification with application to epilepsy and Parkinson's disease [J]. *International Journal of Neural Systems*, 2020, DOI:10.1142/S0129065720500434.
- [12] VUTTIPITTAYAMONGKOL P, ELYAN E. Response to discussion on "Improved overlap-based undersampling for imbalanced dataset classification with application to epilepsy and Parkinson's disease" [J]. *International Journal of Neural Systems*, 2020, DOI: 10.1142/S0129065720750027.
- [13] LI X, WAN W, ZHOU Y, et al. Deep multiple instance learning with spatial attention for ROP case classification, instance selection and abnormality localization [C]. 2020 25th International Conference on Pattern Recognition (ICPR), 2020:7293-7298.
- [14] ABAYOMI-ALLI O O, DAMAŠEVIČIUS R, MASKELIŪNAS R, et al. BiLSTM with data augmentation using interpolation methods to improve early detection of Parkinson disease [C]. 2020 15th Conference on Computer Science and Information Systems (FedCSIS), 2020: 371-380.
- [15] XU Z J, WANG R F, WANG J, et al. Parkinson's disease detection based on spectrogram-deep convolutional generative adversarial network sample augmentation [J]. *IEEE Access*, 2020, 8:206888-206900.
- [16] KONONENKO I. Estimating attributes: Analysis and extensions of RELIEF [C]. *European Conference on Machine Learning*, 1994: 171-182.
- [17] SAKARB E, ISENKUL M E, SAKAR C O, et al. Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings [J]. *IEEE Journal of Biomedical and Health Informatics*, 2013, 17 (4): 828-834.
- [18] BOYD S, PARIKH N, CHU E. Distributed optimization and statistical learning via the alternating direction method of multipliers [J]. *Foundations and Trends in Machine Learning*, 2011, 3(1):1-122.
- [19] SOREL M, SROUBEK F. Fast convolutional sparse coding using matrix inversion lemma [J]. *Digital Signal Processing*, 2016, 55(1):44-51.
- [20] CAI X J, GU G Y, HE B, et al. A proximal point algorithms revisit on the alternating direction method of multipliers [J]. *Science China Mathematics*, 2013, 56: 2179-2186.
- [21] 王金甲,洪文学,李昕. 一种 K-均值脸谱图聚类新算法[J]. *仪器仪表学报*, 2007, 28(10): 1916-1920.
- WANG J J, HONG W X, LI X. Novel clustering algorithm using K-means chernoff faces [J]. *Chinese Journal of Scientific Instrument*, 2007, 28 (10): 1916-1920.
- [22] GENG L, SHAN H F, XIAO Z T, et al. Voice pathology detection and classification from speech signals and EGG signals based on a multimodal fusion method [J]. *Biomedizinische Technik*, 2021, 66(6): 613-625.
- [23] CANTURK I, KARABIBER F. A machine learning system for the diagnosis of Parkinson's disease from speech signals and its application to multiple speech signal types [J]. *Arabian Journal for Science and Engineering*, 2016, 41(12): 5049-5059.
- [24] ZHANG HH, YANG L, LIU Y, et al. Classification of Parkinson's disease utilizing multi-edit nearest-neighbor and ensemble learning algorithms with speech samples [J]. *Biomedical Engineering Online*, 2016, 15(1): 122-143.
- [25] BENBA A, JILBAB A, HAMMOUCH A. Using human factor cepstral coefficient on multiple types of voice recordings for detecting patients with Parkinson's disease [J]. *IRBM*, 2017, 38(6): 346-351.
- [26] LI Y, ZHANG C, JIA Y, et al. Simultaneous learning of speech feature and segment for classification of Parkinson disease [C]. 2017 IEEE 19th International Conference on

e-Health Networking, Applications and Services (Healthcom), IEEE, 2017: 1-6.

- [27] BENBA, A, JILBAB A, HAMMOUCH A. Analysis of multiple types of voice recordings in cepstral domain using MFCC for discriminating between patients with Parkinsons disease and healthy people [J]. International Journal of Speech Technology, 2016, 19(3): 449-456.
- [28] ALI L, ZHU C, ZHANG Z, et al. Automated detection of Parkinson's disease based on multiple types of sustained phonations using linear discriminant analysis and genetically optimized neural network [J]. IEEE Journal of Translational Engineering in Health and Medicine, 2019, 7: 1-10.
- [29] ALI L, ZHU C, ZHANG C, et al. Early diagnosis of Parkinson's disease from multiple voice recordings by simultaneous sample and feature selection [J]. Expert Systems with Applications, 2019, 137:22-28.
- [30] LI Y M, ZHANG X Y, WANG P, et al. Insight into an unsupervised two-step sparse transfer learning algorithm for speech diagnosis of Parkinson's disease [J]. Neural Computing and Applications, 2021, 33 ( 15 ): 9733-9750.

## 作者简介



**张小恒**, 2007 年于重庆大学获得硕士学位, 现为重庆开放大学副教授, 主要研究方向为生物医学信息处理和机器学习。

E-mail: 7818320@qq.com

**Zhang Xiaoheng** received his M. Sc. degree from Chongqing University in 2007. He is currently an associate professor at Chongqing Open University. His main research interests include biomedical information processing and machine learning.



**李勇明** (通信作者), 2004 年于重庆大学获得博士学位, 现为重庆大学教授、博士生导师, 主要研究方向为生物医学信息处理和机器学习。

E-mail: liyyc@cqu.edu.cn

**Li Yongming** (Corresponding author) received his Ph. D. degree from Chongqing University in 2004. He is currently a professor and a Ph. D. advisor at Chongqing University. His main research interests include biomedical information processing and machine learning.