

# 面向帕金森病语音诊断的非监督两步式卷积 稀疏迁移学习算法

张小恒<sup>1,2</sup>, 张馨月<sup>1</sup>, 李勇明<sup>1</sup>, 王 品<sup>1</sup>, 刘玉川<sup>1</sup>

(1. 重庆大学微电子与通信工程学院, 重庆 400030; 2. 重庆广播电视大学, 重庆 400052)

**摘 要:** 帕金森病(Parkinson's Disease, PD)语音诊断存在小样本问题, 如果借助相关语音数据集进行迁移学习, 容易加重训练集和测试集之间的分布差异, 影响分类准确率. 为了解决上述矛盾问题, 本文提出了两步式稀疏迁移学习算法. 该算法分为两大步: 第一步算法为语音段特征同时优选的快速卷积稀疏编码算法, 构造卷积稀疏编码算子用于快速学习公共语音数据集的结构信息, 然后将其迁移到PD语音目标集以弥补后者样本信息的不足, 接着再同时对语音段和特征进行同时优选以获得更有价值的信息; 第二步算法为联合局部结构信息分布对齐算法, 对训练集和测试集进行域适应, 在保持各自样本结构信息的同时, 最小化分布误差. 实验结果表明: 本文算法中每一步迁移学习算法均有效; 与相关算法相比, 本文算法准确率显著较高, 达97.5%.

**关键词:** 语音诊断; 帕金森症(PD); 两步式稀疏迁移学习; 卷积稀疏迁移学习; 域适应

**中图分类号:** TP319.41

**文献标识码:** A

**文章编号:** 0372-2112(2022)01-0177-08

**电子学报 URL:** <http://www.ejournal.org.cn>

**DOI:** 10.12263/DZXB.20201003

## An Unsupervised Two-Step Convolution Sparse Transfer Learning Algorithm for Parkinson's Disease Speech Diagnosis

ZHANG Xiao-heng<sup>1,2</sup>, ZHANG Xin-yue<sup>1</sup>, LI Yong-ming<sup>1</sup>, WANG Pin<sup>1</sup>, LIU Yu-chuan<sup>1</sup>

(1. College of Communication Engineering, Chongqing University, Chongqing 400030, China;

2. Chongqing Radio & TV University, Chongqing 400052, China)

**Abstract:** Parkinson's disease(PD) speech diagnosis has a small sample problem. Although it is possible to transfer learning with the help of relevant speech datasets. The introduction of other samples will lead to the distribution difference between samples of different subjects, so the classification accuracy is greatly affected. Therefore, in this paper, to solve the problems above, we propose a novel unsupervised two-step convolutional sparse transfer learning algorithm. The algorithm is divided into two steps: fast convolutional sparse coding with coordinate selection of samples and features(FCSC&SF), joint local structure distribution alignment(JLSDA). In the FCSC&SF, speech structure among public speech dataset is quickly learned by fast convolution sparse coding(FCSC), and transferred into the target dataset, after that, the more valuable information is obtained by coordinate selection of samples and features. JLSDA is designed to maintain the local structure information in the two domains, and reduce the distribution difference between the two domains at the same time. The experimental results showed that each step of the proposed algorithm has a positive effect on the classification results; compared with the representative relevant algorithms, the accuracy of the proposed method is significantly higher at 97.5%.

**Key words:** speech diagnosis; parkinson's disease(PD); two-step sparse transfer learning; convolutional sparse coding transfer learning; domain adaptation

## 1 引言

帕金森病(Parkinson's Disease, PD)是全球第二大

退行性神经疾病, 危害严重<sup>[1]</sup>. 语音障碍作为PD的显著症状又叫作PD构音障碍<sup>[2]</sup>. 基于语音数据的PD诊

收稿日期: 2020-09-10; 修回日期: 2021-02-24; 责任编辑: 孙瑶

基金项目: 国家自然科学基金(No.61771080); 重庆市自然科学基金(No.cstc2020jcyj-msxmX0100, No.cstc2020jcyj-msxmX0196); 重庆市社会科学规划项目(No.2018YBY133)

断方法具有简易、高效、非接触等特点,因此深入研究PD语音诊断具有重要的科学意义及实用价值.近年来已有大量PD语音诊断算法出现,但其预测精度仍有较大提升空间.

常用的PD语音特征分为基音相关类、能量相关类、语速相关类及内容相关类<sup>[3,4]</sup>.特征选择/变换方法主要可分为基于神经网络方法<sup>[5-7]</sup>、基于主成分分析方法<sup>[8]</sup>、基于串行搜索方法<sup>[3,8]</sup>、基于进化计算方法<sup>[7]</sup>,还有P值<sup>[4]</sup>、线性判别分析(Linear Discriminant Analysis, LDA)<sup>[9]</sup>等方法.常用分类器主要有支持向量机(Support Vector Machines, SVM)<sup>[3-5]</sup>和K近邻(K-Nearest-Neighbor, KNN)<sup>[3,4,10]</sup>,还有随机森林(Random Forest, RF)<sup>[3]</sup>、贝叶斯网络<sup>[11]</sup>、概率神经网络<sup>[12,13]</sup>和决策树<sup>[13]</sup>等.

值得注意的是,上述算法均仅基于本地PD语音数据集.如前所述,本地PD语音数据集存在小样本问题,限制了准确性的提高.研究表明,迁移学习方法可以有效解决这类难题<sup>[5]</sup>.相关研究<sup>[3,5]</sup>已证实了迁移学习在PD诊断中的有效性,但这些诊断方法都不是基于语音数据的.此外,这些文献中的迁移学习方法都只涉及从源域到目标域的一步迁移,且未考虑本地PD语音集中训练集和测试集的不同受试者数据之间的分布差异.虽然已有学者将训练集和测试集看作不同域进行了迁移学习研究<sup>[14]</sup>,但都未涉及PD语音分类领域.总之,现有的PD语音诊断算法研究,既未考虑利用迁移学习来解决小样本问题,也没有考虑通过域适应方法减少训练集与测试集的数据分布差异.上述问题亟待解决.

近年来,图像处理研究领域出现一种新颖的稀疏学习算法——卷积稀疏编码(Convolutional Sparse Coding, CSC)算法.该算法具有很强的稀疏学习能力,能够有效地获取数据中隐含的结构与模式<sup>[15]</sup>,适合于提取PD语音数据中的隐藏信息.此外,基于relief算法<sup>[16]</sup>,本文构造了语音段特征同时优选算法,挖掘了更有效的信息,拟与CSC结合,设计第一步迁移学习算法(First Step Transfer Learning Algorithm, FT).域适应的目的是在不同但相关的任务或领域之间转移共享知识<sup>[17]</sup>.非监督域适应(Unsupervised Domain Adaptation, UDA)的常见做法是尽量减少域之间的差异,以保持域不变特征<sup>[18]</sup>或在执行域对齐时学习更多有差异的特征.因此,本文考虑将其用于PD语音数据域适应,保持训练集(源域)和测试集(目标域)各自局部结构特性不变的同时,减小领域之间的分布差异.基于此,本文提出了一种新的域适应方法——联合局部结构信息分布对齐(Joint Local Structure Distribution Alignment, JLSDA)算法,从而在分类之前降低了训练集与测试集之间的分布差异,有助于提高分类准确率.结合上述两步迁移学

习算法,本文提出了一种新的PD语音诊断方法——非监督两步式卷积稀疏迁移学习(Unsupervised Two-Step Convolutional Sparse Transfer Learning, TSTL)算法.

本文的主要贡献和创新点如下.

(1)为了解决PD语音诊断中小样本和训练集测试集间分布差异这两大问题,本文提出了非监督两步式卷积稀疏迁移学习算法(TSTL).该算法同时将不同数据集之间的迁移学习与同一数据集中训练集与测试集之间的迁移学习相结合.

(2)为了提高迁移源数据集效率,将CSC算法与语音段特征同时优选算法相结合,基于傅里叶域并交换迭代次序的交替方向乘法提出了第一步迁移学习算法——语音段特征同时优选的快速卷积稀疏编码(Fast Convolutional Sparse Coding with Coordinate Selection of Samples and Seatures, FCSC&SF)算法.

(3)本文针对PD语音集中训练集和测试集的分布差异问题,首次提出了联合局部结构信息分布对齐(JLSDA)算法.

(4)本文将受试者多个语音段样本合并转化为二维数据样本,用于迁移学习算法的建模和验证,并充分考虑了语音段和特征之间的协同关系.

## 2 方法

### 2.1 相关符号描述

PD语音集即目标域 $\mathbf{F} = [\vec{\mathbf{F}}_1, \vec{\mathbf{F}}_2, \dots, \vec{\mathbf{F}}_G]^\top$ ,其中 $\vec{\mathbf{F}}_i = [f_{i1}, f_{i2}, \dots, f_{iN}]$ ,  $1 \leq i \leq G$ ,基于受试者的分块矩阵为 $\mathbf{F} =$

$$[\vec{\mathbf{F}}_1, \vec{\mathbf{F}}_2, \dots, \vec{\mathbf{F}}_M]^\top, \text{其中}, \vec{\mathbf{F}}_i = \begin{bmatrix} f_{i1} & f_{i2} & \dots & f_{iN} \\ f_{21} & f_{22} & \dots & f_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ f_{G_01} & f_{G_02} & \dots & f_{G_0N} \end{bmatrix}, 1 \leq$$

$i \leq M$ .

$M$ 个受试者合在一起的样本总数为 $G$ ,每个样本的特征数为 $N$ ,即每个受试者的样本数为 $G_0 = G/M$ .在第一步迁移之前,作为源域的公共语音集可通过加入不同信噪比及类型的噪声扩展为更大尺度的数据集.扩展后的公共语音集为 $\mathbf{S}' = [\vec{\mathbf{S}}'_1, \vec{\mathbf{S}}'_2, \dots, \vec{\mathbf{S}}'_J]^\top$ ,其中, $\vec{\mathbf{S}}'_j = \varphi(\vec{\mathbf{S}}_j, \vec{\mathbf{N}}_j, \text{SNR}_j)$ , $\vec{\mathbf{S}}_j$ 是公共语音集中的原始语音样本, $\vec{\mathbf{N}}_j$ 是不同类型的噪声, $\varphi(\cdot)$ 是加噪函数.然后提取扩展数据集的特征并形成特征集 $\mathbf{Y} = [\vec{\mathbf{Y}}_1, \vec{\mathbf{Y}}_2, \dots, \vec{\mathbf{Y}}_L]^\top$ ,其中,特征向量总数为 $L$ , $N$ 个不同类型特征构建特征向量 $\vec{\mathbf{Y}}_i = [\xi_1(\vec{\mathbf{S}}'_i), \xi_2(\vec{\mathbf{S}}'_i), \dots, \xi_N(\vec{\mathbf{S}}'_i)]$ ,  $1 \leq i \leq L$ ,特征提取方式与文献[7]相同.特征集 $\mathbf{Y}$ 还可表示为分块矩阵形式 $\mathbf{Y} =$

$[\tilde{Y}_1, \tilde{Y}_2, \dots, \tilde{Y}_L]^T$ , 其中  $\tilde{Y}_i$  是二维  $G_0 \times N$  分块矩阵, 即卷积稀疏学习的训练样本。

在第二步迁移学习(ST)算法中, 领域  $\mathcal{Q}$  包含了  $d$  维特征空间  $\mathbf{X}'$  及边缘概率分布  $P(x')$ , 源域数据记作  $\mathbf{X}'_S = [x'_{S_1}, x'_{S_2}, \dots, x'_{S_{N_S}}]^T \in \mathbf{R}^{N_S \times L}$ , 目标域数据记作  $\mathbf{X}'_T = [x'_{T_1}, x'_{T_2}, \dots, x'_{T_{N_T}}]^T \in \mathbf{R}^{N_T \times L}$ ,  $N_S$  和  $N_T$  分别为源域和目标域的样本数量, 其中下标  $S$  表示源域相关样本或源域变换后数据, 下标  $T$  表示目标域相关样本或目标域变换后数据, 标签向量记作  $\mathbf{Y}'_S = [y'_{S_1}, y'_{S_2}, \dots, y'_{S_{N_S}}]^T \in \mathbf{R}^{N_S}$ , 类别数为  $C$ . 符号  $\|\cdot\|_{\mathbb{H}}$  是重构希尔伯特核空间(RKHS)范数,  $\phi(\cdot)$  为核映射函数,  $\text{tr}(\cdot)$  表示矩阵迹算子。

## 2.2 算法研究

本文提出的面向帕金森语音诊断的非监督两步式卷积稀疏迁移学习(TSTL)算法主要包括两部分: 第一步迁移学习算法为语音段特征同时优选的快速卷积稀疏编码(FCSC&SF)算法, 第二步迁移学习算法为联合局部结构信息分布对齐(JLSDA)算法. 第一步迁移学习算法的目的是学习公共语音集(源域)中的有用信息并将其迁移到目标域. 第二步迁移(JLSDA)的目的是对齐数据分布同时保持原始结构不变, 降低训练集和测试集之间的分布差异。

### 2.3 第一步迁移学习算法(FT)——语音段特征同时优选的快速卷积稀疏编码(FCSC&SF)算法

#### 2.3.1 快速卷积稀疏编码

在 CSC 中, 给定样本集  $\{\mathbf{x}_g\}_{g=1}^G$ , 卷积核组  $\{\mathbf{d}_k\}_{k=1}^K$  可以通过最小化如下目标函数得到:

$$\arg \min_{\mathbf{e}, \mathbf{d}} \frac{1}{2} \sum_{g=1}^G \left\| \mathbf{x}_g - \sum_{k=1}^K \mathbf{d}_k * \mathbf{e}_{g,k} \right\|_2^2 + \eta \sum_{g=1}^G \sum_{k=1}^K \|\mathbf{e}_{g,k}\|_1 \quad (1)$$

s.t.  $\|\mathbf{d}_k\|_2 \leq 1, \forall k = \{1, 2, \dots, K\}$

其中,  $\mathbf{x}_g = \tilde{\mathbf{Y}}_g$  是  $G_0 \times N$  分块矩阵,  $\mathbf{e}_{g,k}$  是  $G_0 \times N$  特征映射矩阵, 与对应的卷积核  $\mathbf{d}_k$  进行卷积运算逼近  $\mathbf{x}_g$ , 符号  $*$  表示二维卷积运算,  $\eta$  是大于 0 的正则化因子. 式(1)可进一步简洁表达如下:

$$\arg \min_{\mathbf{e}} \frac{1}{2} \|\mathbf{D}\mathbf{e} - \mathbf{x}\|_2^2 + \eta \|\mathbf{e}\|_1, \quad (2)$$

s.t.  $\|\mathbf{d}_k\|_2 \leq 1$

其中,  $\sum_{k=1}^K \mathbf{d}_k * \mathbf{e}_{g,k} = \mathbf{D}\mathbf{e}$ ,  $\mathbf{D} = [\mathbf{D}_1 \ \mathbf{D}_2 \ \dots \ \mathbf{D}_K]$  是  $[\mathbf{d}_1 \ \mathbf{d}_2 \ \dots \ \mathbf{d}_K]$  对应的向量化卷积算子,  $\mathbf{e} =$

$[\mathbf{e}_1^T \ \mathbf{e}_2^T \ \dots \ \mathbf{e}_K^T]^T$  是特征映射向量组。

上述优化问题可基于经典的交替方向乘子法(Alternating Direction Multiplier Method, ADMM)实现<sup>[19]</sup>. 由于源数据集规模较大, 训练时间较长不利于实际应用, 为实现高效运算, 本文基于文献[20]的傅里叶域快速运算并结合文献[21]的交换迭代次序法加以解决. 其特征映射部分训练部分如下:

$$\begin{aligned} & \begin{cases} \bar{\mathbf{e}}^j = \left( I - \lambda \mathbf{D}^T \left( I + \lambda \sum \mathbf{D}_k \mathbf{D}_k^T \right)^{-1} \mathbf{D} \right) (\lambda \mathbf{D}^T \mathbf{x} + \mathbf{b}^j - \mathbf{u}^j) \\ \bar{\mathbf{u}}^j = \mathbf{u}^j + \bar{\mathbf{e}}^j - \mathbf{b}^j \\ \bar{\mathbf{b}}^j = \begin{cases} \bar{\mathbf{e}}^j + \bar{\mathbf{u}}^{(j)} - \alpha, & \bar{\mathbf{e}}^j + \bar{\mathbf{u}}^{(j)} > \alpha \\ 0, & |\bar{\mathbf{e}}^j + \bar{\mathbf{u}}^{(j)}| \leq \alpha \\ \bar{\mathbf{e}}^j + \bar{\mathbf{u}}^{(j)} + \alpha, & \bar{\mathbf{e}}^j + \bar{\mathbf{u}}^{(j)} < -\alpha \end{cases} \end{cases} \\ \text{step1.} & \\ \text{step2.} & \begin{pmatrix} \mathbf{b}^{(j+1)} \\ \mathbf{u}^{(j+1)} \end{pmatrix} = \begin{pmatrix} \mathbf{b}^j \\ \mathbf{u}^j \end{pmatrix} - \gamma \begin{pmatrix} \mathbf{b}^j - \bar{\mathbf{b}}^j \\ \mathbf{u}^j - \bar{\mathbf{u}}^j \end{pmatrix} \end{aligned} \quad (3)$$

其中,  $\lambda$  是大于 0 的 ADMM 缩放因子,  $\alpha = \lambda \eta$ ,  $\left( I + \lambda \sum \mathbf{D}_k \mathbf{D}_k^T \right)^{-1}$  可简化为傅里叶域的确系数形式  $\frac{1}{1 + \sum_{k=1}^K |\hat{\mathbf{d}}_k|^2}$ ,  $\hat{\mathbf{d}}_k$  是  $\mathbf{d}_k$  傅里叶变换后的幅度值,  $S_\alpha(\cdot)$  是元素运算的软阈值函数。

卷积核训练部分生成如下:

$$\begin{aligned} & \begin{cases} \bar{\mathbf{d}}^j = \left( I - \lambda \mathbf{E}^T \left( I + \lambda \sum \mathbf{E}_k \mathbf{E}_k^T \right)^{-1} \mathbf{E} \right) (\lambda \mathbf{E}^T \mathbf{x} + \mathbf{c}^j - \mathbf{v}^j) \\ \bar{\mathbf{v}}^j = \mathbf{v}^j + \bar{\mathbf{d}}^j - \mathbf{c}^j \\ \bar{\mathbf{c}}^j = \begin{cases} \frac{(\bar{\mathbf{d}}^j + \bar{\mathbf{v}}^j) F_{\text{supp}}(\bar{\mathbf{d}}^j)}{\|\bar{\mathbf{d}}^j + \bar{\mathbf{v}}^j\|}, & \text{if } \|\bar{\mathbf{d}}^j + \bar{\mathbf{v}}^j\| F_{\text{supp}}(\bar{\mathbf{d}}^j) > 1 \\ (\bar{\mathbf{d}}^j + \bar{\mathbf{v}}^j) F_{\text{supp}}(\bar{\mathbf{d}}^j), & \text{otherwise} \end{cases} \end{cases} \\ \text{step1.} & \\ \text{step2.} & \begin{pmatrix} \mathbf{c}^{(j+1)} \\ \mathbf{v}^{(j+1)} \end{pmatrix} = \begin{pmatrix} \mathbf{c}^j \\ \mathbf{v}^j \end{pmatrix} - \gamma \begin{pmatrix} \mathbf{c}^j - \bar{\mathbf{c}}^j \\ \mathbf{v}^j - \bar{\mathbf{v}}^j \end{pmatrix} \end{aligned} \quad (4)$$

其中,  $F_{\text{supp}}(\mathbf{d})$  是一个蒙板函数, 支撑集  $\text{SUPP}(\mathbf{d})$  上为 1 否则为 0,  $\gamma$  是松弛因子。

式(3)、式(4)中 step1 的时间复杂度等同于文献[20], 但由于 step2 的时间复杂度极低, 因此本文提出的快速卷积稀疏编码(FCSC)算法整体迭代效率远高于文献[20].

目标域  $\mathbf{F}$  的特征映射变换如下:  $\tilde{\mathbf{F}}_i$  替换  $\mathbf{x}$ , 用训练好的卷积核  $\hat{\mathbf{D}}$  替换  $\mathbf{D}$ , 基于式(3)通过有限次迭代可将目标域特征矩阵  $\tilde{\mathbf{F}}_i$  可转换为特征映射矩阵  $\mathbf{e} =$

$[e_1 \ e_2 \ \dots \ e_k]^T$ , 可以选择固定的  $e_k$  作为映射  $\tilde{E}_i$ , 从

$$\text{而构建转换后的目标域特征矩阵 } \mathbf{E} = \begin{bmatrix} \tilde{E}_1 \\ \tilde{E}_2 \\ \vdots \\ \tilde{E}_M \end{bmatrix}.$$

### 2.3.2 语音段特征同时优选

为进行语音段特征同时优选, 将矩阵  $\tilde{E}_i$  进行向量化扩展可生成  $\vec{G}_i = \text{RESHAPE}(\tilde{E}_i, H_0 \times N - 1 \times N')$ ,  $M$  个 subject 对应的行向量  $\vec{G}_i$  组合可看作列特征向量形式, 即  $\mathbf{G} = \begin{bmatrix} \vec{G}_1 \\ \vec{G}_2 \\ \vdots \\ \vec{G}_M \end{bmatrix}$  并归一化为  $\mathbf{G}'$ , 其中训练集  $\mathbf{G}'_m =$

$$\begin{bmatrix} \vec{G}'_1 \\ \vec{G}'_2 \\ \vdots \\ \vec{G}'_M \end{bmatrix} \text{ 可看作列向量形式, 其中 } \vec{G}'_n \text{ 为列特征向量, 使用 relief 算法计算特征 } \vec{G}' \text{ 所对应的权重向量 } \vec{W}, \text{ 可根据权重值大小优选 } Q \text{ 个最大权重所对应的列特征向量得到新的训练集 } \mathbf{G}^Q = \begin{bmatrix} \vec{G}'_1 \\ \vec{G}'_2 \\ \vdots \\ \vec{G}'_Q \end{bmatrix} = \begin{bmatrix} \vec{G}_1 & \vec{G}_2 & \dots & \vec{G}_N \end{bmatrix}$$

$\begin{bmatrix} \vec{\xi}_1^T \\ \vec{\xi}_2^T \\ \vdots \\ \vec{\xi}_Q^T \end{bmatrix}$ , 其中  $\vec{\xi}_i = \begin{bmatrix} 0 \dots 0 1 \\ \text{index} \end{bmatrix}$  且 index 是权重  $w_i$  所对应特征向量的列标记, 同理测试集也根据  $\vec{W}$  进行特征优选.

FCSC&SF 算法伪代码描述如算法 1 所示.

$$\min \left[ \left\| \frac{1}{n_S} \sum_{i=1}^{n_S} \phi(x_{S_i}) - \frac{1}{n_T} \sum_{j=1}^{n_T} \phi(x_{T_j}) \right\|_H^2 + \frac{1}{2} \sum_{m,n} \left\| \phi(x_{S_m}) - \phi(x_{S_n}) \right\|_H^2 S_{mn} + \frac{1}{2} \sum_{p,q} \left\| \phi(x_{T_p}) - \phi(x_{T_q}) \right\|_H^2 S_{pq} \right] \quad (5)$$

根据多数非监督域适应方法的关键假设, 不同域的边缘分布不同, 即  $P \neq Q$ , 但预测分布却相同, 即  $P(Y_S | X_S) = Q(Y_T | X_T)$ <sup>[17]</sup>. 式(5)中第一个范数即源域和目标域之间的最大均值差异(MMD), 余下的范数累加

$$\begin{aligned} & \min \text{tr} \left( \frac{1}{n_S^2} \phi(X_S) \mathbf{H}^T \phi^T(X_S) + \frac{1}{n_T^2} \phi(X_T) \mathbf{H}^T \phi^T(X_T) - \frac{1}{n_S n_T} \phi(X_S) \mathbf{H}^T \phi^T(X_T) - \frac{1}{n_S n_T} \phi^T(X_S) \mathbf{H} \phi(X_T) \right) \\ & + \frac{1}{2} \sum_{m,n} \text{tr} \left( \phi(x_{S_m}) \phi^T(x_{S_m}) + \phi(x_{S_n}) \phi^T(x_{S_n}) - \phi(x_{S_m}) \phi^T(x_{S_n}) - \phi(x_{S_n}) \phi^T(x_{S_m}) \right) S_{mn} \\ & + \frac{1}{2} \sum_{p,q} \text{tr} \left( \phi(x_{T_p}) \phi^T(x_{T_p}) + \phi(x_{T_q}) \phi^T(x_{T_q}) - \phi(x_{T_p}) \phi^T(x_{T_q}) - \phi(x_{T_q}) \phi^T(x_{T_p}) \right) S_{pq} \end{aligned} \quad (6)$$

根据矩阵迹运算相关特性及之前的相关定义, 式(6)可简化为

$$\begin{aligned} & \min \text{tr}(\hat{\mathbf{K}} \hat{\mathbf{M}}) + \text{tr}(\phi(X'_S) \phi^T(X'_S) \hat{\mathbf{L}}_S) \\ & + \text{tr}(\phi(X'_T) \phi^T(X'_T) \hat{\mathbf{L}}_T) \end{aligned} \quad (7)$$

式(6)中第一部分与 TCA 域适应目标式类似<sup>[17]</sup>. 式

### 算法 1 第一步迁移学习算法-语音段特征同时优选的卷积稀疏编码(FCSC&SF)算法

**输入:** 公共数据集  $S$ , 目标域数据集  $F$ , 总样本数  $H$ , 每个样本的特征数  $N$ , 患者数  $M$

**输出:** 新的目标域数据集  $X'$

**步骤:**

- (1) 加不同类型及信噪比噪声到公共数据集中并扩展为数据集  $S$
- (2) 从  $S$  中提取多类型特征并构建特征数据集, 即源域数据集  $Y$
- (3) **For** iteration = 1...iter\_num **do**
- (4) **For** iteration1 = 1...iter\_num1 **do**
- (5) 根据式(3), 固定卷积核求解特征映射
- (6) **End for**
- (7) **For** iteration2 = 1...iter\_num2 **do**
- (8) 根据式(4), 固定特征映射求解卷积核
- (9) **End for**
- (10) **End for**
- (11) 通过训练出的卷积核提取 PD 数据集的特征映射  $E$ , 将  $E$  扩展为  $G$  并归一化得到  $G'$
- (12) 根据 relief 算法计算权重向量  $\vec{W}$
- (13) 选择权重向量  $\vec{W}$  中  $Q$  个最大权重所对应的特征向量并构建新的训练集及测试集

### 2.4 第二步迁移学习算法——联合局部结构信息分布对齐(JLSDA)算法

本文提出一种新的通过映射源域(训练集)和目标域(测试集)公共流行空间以适配分布及保持样本间结构的方法. 公式如下:

求和表示同一域中样本之间的结构关系<sup>[22]</sup>.

#### 2.4.1 联合优化

根据式(5), JLSDA 算法可重新表示如下:

(7) 中  $\hat{\mathbf{K}}$  是核矩阵,  $\hat{\mathbf{M}}$  是 MMD 矩阵, 都可以通过第一步迁移学习得到.  $\phi(X'_S) \phi^T(X'_S)$  记作  $\hat{\mathbf{K}}_S$  且  $\phi(X'_T) \phi^T(X'_T)$  记作  $\hat{\mathbf{K}}_T$ .  $\hat{\mathbf{L}}_S = \hat{\mathbf{D}}_S - \hat{\mathbf{S}}_S$  和  $\hat{\mathbf{L}}_T = \hat{\mathbf{D}}_T - \hat{\mathbf{S}}_T$  分别是源域和目标域的拉普拉斯矩阵, 其中,  $\hat{\mathbf{D}}_S$  和  $\hat{\mathbf{D}}_T$  为源域和目标域的度矩阵,  $\hat{\mathbf{S}}_S$  和  $\hat{\mathbf{S}}_T$  为源域和目标域的邻接矩阵. 源域及目

标域度矩阵对角线上元素分别为  $\hat{D}_{mm} = \sum_n \hat{S}_{mn}$  和  $\hat{D}_{qq} = \sum_p \hat{S}_{pq}$ ,  $\hat{S}_{mn}$  和  $\hat{S}_{pq}$  分别为源域及目标域邻接矩阵的行列元素. 为方便推导,式(7)可简化为

$$\min \text{tr}(\hat{K}\hat{M}) + \text{tr}(\hat{K}\hat{L}) \quad (8)$$

其中,  $(\cdot)$  为  $\hat{K}$  与  $\hat{L}$  的点乘运算符,  $\hat{L} = \begin{bmatrix} \hat{L}_S & \mathbf{0} \\ \mathbf{0} & \hat{L}_T \end{bmatrix} \in \mathbf{R}^{N \times N}$ .

由于利用最大平均偏差嵌入法(Maximum Mean Discrepancy Embedding, MMDE)求解需要解决半定规划问题(Semidefinite Programming Problem, SDP), 计算复杂度极高, 因此本文采纳一种利用显式低秩表示的统一核学习方法<sup>[17]</sup>, 可引入变换矩阵  $W \in \mathbf{R}^{d \times N}$ . 设计低秩表示的核矩阵  $\tilde{K}$  替换核矩阵  $\hat{K}$ , 由  $\text{tr}(\hat{K}\hat{M})$  得  $\text{tr}(\tilde{K}\hat{M}) = \text{tr}((\hat{K}WW^T\hat{K})\hat{M}) = \text{tr}(W^T\hat{K}\hat{M}\hat{K}W)$ . 令  $\hat{L}^*$  为  $\hat{K}$  和  $\hat{L}$  的点积, 且设计低秩表示的  $\tilde{L}^*$  替换  $\hat{L}^*$ , 由  $\text{tr}(\hat{K}\hat{L}) = \text{tr}(\hat{L}^*)$  得  $\text{tr}(\tilde{L}^*) = \text{tr}(W^T\hat{L}^*W)$ .

引入尺度归一化  $W^T\hat{K}\hat{D}\hat{K}W = I$  及正则项, 其中度矩阵  $\hat{D} = \begin{bmatrix} \hat{D}_S & \mathbf{0} \\ \mathbf{0} & \hat{D}_T \end{bmatrix} \in \mathbf{R}^{N \times N}$ ,  $\mu$  为正则因子, 可得

$$\begin{aligned} \min_W \text{tr}(W^T\hat{K}\hat{M}\hat{K}W + W^T\hat{L}^*W + \mu W^TW) \\ \text{s.t. } W^T\hat{K}\hat{D}\hat{K}W = I \end{aligned} \quad (9)$$

基于拉格朗日乘子法, 问题(9)可重新表达为

$$\text{tr}(W^T(\hat{K}\hat{M}\hat{K} + \hat{L}^* + \mu I)W) - \text{tr}(W^T\hat{K}\hat{D}\hat{K}W - I)Z \quad (10)$$

其中,  $Z$  是包含拉格朗日乘子的对角阵. 式(10)对  $W$  取导后并置 0, 可得

$$(\hat{K}\hat{M}\hat{K} + \hat{L}^* + \mu I)W = \hat{K}\hat{D}\hat{K}WZ \quad (11)$$

其中,  $W$  为  $(\hat{K}\hat{M}\hat{K} + \hat{L}^* + \mu I)^{-1}\hat{K}\hat{D}\hat{K}$  前  $d$  个主特征向量构成,  $d \leq N = n_1 + n_2$ ,  $n_1$  为源域样本数,  $n_2$  为目标域样本数.

JLSDA 算法的伪代码如算法 2 所示.

基于 JLSDA 算法得到变换矩阵  $W$ , 进一步得到变换

**算法 2 第二步迁移学习算法-联合局部结构信息分布对齐(JLSDA)算法**

**输入:** 经过第一步迁移之后的 Sakar 训练集  $X'_S$ , 经过第一步迁移之后的 Sakar 测试集  $X'_T$ , 初始化参数: 正则化参数, 核函数, 邻居数, 关联矩阵模式

**输出:** 变换矩阵  $W$

**步骤:**

- (1) 计算 MMD 矩阵  $\hat{M}$
- (2) 选择合适的核函数, 获取核矩阵  $\hat{K}$
- (3) 构建源域和目标域的关联矩阵  $\hat{S}$
- (4) 分别构建源域和目标域的对角矩阵  $\hat{D}$ ,  $\hat{D}_i = \sum_j \hat{S}_{ij}$
- (5) 分别构建源域和目标域的拉普拉斯矩阵  $\hat{L}$ ,  $\hat{L} = \hat{D} - \hat{S}$
- (6) 按照式(11)求解  $W$

后的源域和目标域数据集, 即  $[\phi(X'_S) \ \phi(X'_T)] = W^T\hat{K}$ , 变换后的源域  $\phi(X'_S)$  为前  $n_1$  个列向量构成, 变换后的目标域  $\phi(X'_T)$  为后  $n_2$  个列向量构成.

## 3 结果

### 3.1 实验条件

#### 3.1.1 数据集

本文使用了 3 个语音数据集, 有一定代表性. 它们分别是 TIMIT 公共语音数据集、Sakar 数据集<sup>[4]</sup>和 DNSH 数据集.

第一个语音数据集作为第一步迁移学习的源域数据集. TIMIT 集总共包含 6300 个句子, 由 630 个朗读者每人提供 10 个句子, 但目前只有 240 个语音样本可用, 40 名男性和 40 名女性每人提供 3 段语音. 数据集扩展使用的噪声来源于 NOISEX-9 标准噪声集.

Sakar 数据集是 PD 语音公开数据集, 由 Sakar 等人<sup>[7]</sup>提供, 作为第一步迁移学习的目标域数据集. 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, 包括频率类、振幅类、谐波类、基音频率等.

第 3 个数据集由本文作者自采, 且受试者来自陆军军医大学第一附属医院. 数据包含了未接受治疗的 36 名 PD 患者, 其中 16 名女性(年龄的均值  $\pm$  标准差(mean  $\pm$  std): 57.9  $\pm$  9.0), 20 名男性(mean  $\pm$  std: 60.8  $\pm$  10.6)(患病时间的均值和标准差分别是 7.38 年和 3.58 年); 54 名已接受治疗的 PD 患者, 其中 27 名女性(mean  $\pm$  std: 59.7  $\pm$  8.1), 27 名男性(mean  $\pm$  std: 63.2  $\pm$  10.8)(患病时间的均值和标准差分别是 6.82 年和 3.50 年). 每人提供 13 个语音段, 每个语音段提取 26 个特征.

#### 3.1.2 评价准则

为验证算法的有效性, 本文使用分类准确率、灵敏度和特异度作为实验结果的评估准则.

根据数据集中一个受试者对应多个语音样本的特性, 本文所提出算法使用 LOSO(Leave-One-Subject-Out) 交叉验证方法. 分类器采用线性 SVM, 预测分类准确率累加再算数平均即得最终准确率.

#### 3.1.3 实验平台

本文实验使用的计算机硬件配置为 CPU (intel i3-4170M), 6 GB memory, 操作系统为 64-bit Windows 7, 实验运行在 Matlab R2018b 版本上. 论文参数设置如下. 重复实验次数设置为 10 次, 在第一步迁移学习中, 主训练迭代次数、特征映射训练迭代次数及卷积核训练迭代次数分别为 100, 10 和 10, ADDM 缩放因子  $\lambda = 1$ , 松弛

因子 $\gamma=1$ ,卷积核数量从2增加到8,尺寸为 $8*8$ .在第二步迁移学习中,正则化因子 $\mu=0.01$ ,核类型为‘rbf’,‘rbf’核宽度 $\gamma$ 为100,关联矩阵模式为“simple”模式且最小邻居样本数为1.

### 3.2 两步式迁移学习算法的有效性验证

#### 3.2.1 第一步迁移学习的有效性验证

以Sakar数据集为例.在第一步迁移学习(FT)中,本文主要使用CSC将TIMIT的知识迁移到Sakar数据集.本文尝试处理第一步迁移前后的数据,比较准确率差异.这里分类器选择KNN和SVM,分类结果见表1.

表1 基于Sakar集的第一步迁移学习分类准确率(LOSO)

方法	ACC / %	TPR / %	TNR / %
KNN	52.5	55.0	50.0
SVM(linear)	50.0	50.0	50.0
FT&KNN	90.0	85.0	95.0
FT&SVM(linear)	92.5	95.0	90.0

如表1所示,KNN分类器的准确率为52.5%,优于SVM分类器的50.0%.但是,经过第一步迁移学习之后分类准确率得到显著改善,FT&KNN准确率为90%,FT&SVM准确率为92.5%,灵敏度和特异度也明显提升.实验结果表明,FT从公共语音集迁移到目标域的新增信息有助于目标域分类,其是有效的.

#### 3.2.2 第二步迁移学习的有效性验证

以Sakar数据集为例.在第二步迁移学习(ST)中,JLSDA算法使训练集和测试集的分布差异减小,并保持原始结构信息.参照3.2.1节,本文可通过比较迁移前后数据的分类准确率,来验证ST的有效性.请见表2分类结果.

如表2所示,根据第二步迁移学习(ST)的实验结果,尽管性能提升程度不如第一步迁移学习(FT),但仍然有效.经过第二步迁移学习,基于KNN准确率增加了15%,基于SVM准确率增加了12.5%.

表2 基于Sakar数据集的第二步迁移学习分类准确率(LOSO)

方法	ACC / %	TPR / %	TNR / %
KNN	52.5	55.0	50.0
SVM(linear)	50.0	50.0	50.0
ST&KNN	67.5	65.0	70.0
ST&SVM(linear)	62.5	80.0	45.0

### 3.3 算法比较

基于Sakar数据集,本文选取了基于不同智能算法的代表性PD语音诊断算法进行比较.文献[23]手动选择特征;文献[24]采用样本选择法,文献[25]是样本特征混合选择的代表;文献[26,27]均使用了倒谱系数;

而文献[28]结合利用语音段选择,特征选择及神经网络等多种手段,且取得了相对更好的性能.此外,深度学习近年十分流行,因此本文选取了深度置信网络(Deep Belief Networks, DBN)、卷积神经网络(Convolutional Neural Network, CNN)、自动编码器(Autoencoder)算法,及SVM和KNN相结合的算法进行了比较.为验证迁移学习的效果,本文还选取了卷积稀疏编码迁移学习(Convolutional Sparse Coding Transfer Learning, CSCTL)算法、经典域适应方法——迁移成分分析(Transfer Component Analysis, TCA)<sup>[27]</sup>结合SVM分类器等一步式迁移学习算法,这些方法只单一考虑基于源域结构对目标域进行更有效表达或源域和目标域分布一致问题,通过比较证明,本文的两步式迁移学习在同时考虑上述两个问题后更有效.

表3显示了准确率比较结果.

表3 基于Sakar数据集的相关算法性能比较(LOSO)

研究文献	方法	ACC / %	TPR / %	TNR / %
[23]	4 Feature Selection Methods & 6 Classifiers	57.5	54.30	80.0
[24]	MENN&RF	81.5	92.5	70.5
[25]	Hybrid feature learning & SVM	82.5	85.0	80.0
[26]	HFCC+SVM	87.5	90.0	85.0
[27]	MFCC&SVM	82.5	80.0	85.0
[28]	LDA-NN-GA	95.0	95.0	95.0
-	DBN	54.6	52.4	56.8
-	CNN	60.0	63.0	57.0
-	DBN&SVM	50.5	53.0	48.0
-	Autoencoder&SVM	67.5	65.0	70.0
-	CSC&SVM	70.0	75.0	65.0
-	TCA&SVM	62.5	80.0	45.0
Proposed algorithm	TSTL&SVM	97.5	97.5	97.5

如表3所示,仅有小部分算法能达到90%的准确率.从方法论的角度看,深度学习并不比传统机器学习的效果更好,这也印证了深度学习依赖大样本量训练特性,并不适合PD语音数据集这类小样本数据集.以上实验结果基于LOSO交叉验证.

如表4所示,本文提出的算法在DNSH数据集上也达到了较好的效果.与直接使用SVM和KNN分类器相比,本文提出的算法达到了90.63%的平均分类准确率,证明本文算法在中国人的PD语音数据集上也十分有效.

表 4 基于 DNSH 数据集的相关算法比较 (LOSO)

来源	方法	ACC /%	TPR /%	TNR /%
-	KNN	52.5	55.0	50.0
-	SVM(linear kernel)	42.5	35.0	50.0
Proposed algorithm	TSTL&SVM	90.6	90.6	90.6

#### 4 结论

虽然迁移学习可以解决 PD 语音小样本问题,但面临不同数据集之间和训练集测试集之间的差异. 为了解决这个问题,本文提出了一种新的解决方法——非监督两步式卷积稀疏迁移学习 (TSTL) 算法. 实验结果表明,本文所提出算法的主要创新部分是有效的,在准确率、灵敏度及特异度上都较现有算法显著更好.

尽管本文算法被验证是有效的,但仍存在较大改进空间. 下一步工作拟考虑结合不同大小和类型的公共语音数据集和各种代表性域适应方法,探索进一步提升本文算法性能的可能途径.

#### 参考文献

- [1] MIRARCHI D, VIZZA P, TRADIGO G, et al. Signal analysis for voice evaluation in Parkinson's disease[C]//2017 IEEE International Conference on Healthcare Informatics (ICHI). Park City: IEEE, 2017: 530-535.
- [2] GILLIVAN-MURPHY P, MILLER N, CARDING P. Voice tremor in Parkinson's disease: An acoustic study[J]. Journal of Voice, 2019, 33(4): 526-535.
- [3] ZOU N, HUANG X. Empirical bayes transfer learning for uncertainty characterization in predicting Parkinson's disease severity[J]. IISE Transactions on Healthcare Systems Engineering, 2018, 8(3): 209-219.
- [4] SAKAR B 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.
- [5] 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.
- [6] AL-FATLAWI A H, JABARDI M H, LING S H. Efficient diagnosis system for Parkinson's disease using deep belief network[C]//2016 IEEE Congress on Evolutionary Computation (CEC). Vancouver: IEEE, 2016: 1324-1330.
- [7] AVCI D, DOGANTEKIN A. An expert diagnosis system for Parkinson disease based on genetic algorithm-wavelet kernel-extreme learning machine[J]. Parkinson's Disease, 2016, 2016: 5264743.
- [8] CAI Z N, GU J H, CHEN H L. A new hybrid intelligent framework for predicting Parkinson's disease[J]. IEEE Access, 2017, 5: 17188-17200.
- [9] CAESARENDRA W, ARIYANTO M, SETIAWAN J D, et al. A pattern recognition method for stage classification of Parkinson's disease utilizing voice features[C]//2014 IEEE Conference on Biomedical Engineering and Sciences (IECBES). Kuala Lumpur: IEEE, 2014: 87-92.
- [10] OZKAN H. A comparison of classification methods for tediagnosis of Parkinson's disease[J]. Entropy, 2016, 18(4): 115.
- [11] BENBA A, JILBAB A, HAMMOUCH A. Hybridization of best acoustic cues for detecting persons with Parkinson's disease[C]//2014 Second World Conference on Complex Systems (WCCS). Agadir: IEEE, 2014: 622-625.
- [12] NARANJO L, PÉREZ C J, CAMPOS-ROCA Y, et al. Addressing voice recording replications for Parkinson's disease detection[J]. Expert Systems with Applications, 2016, 46: 286-292.
- [13] HIRSCHAUER T J, ADELI H, BUFORD J A. Computer-aided diagnosis of Parkinson's disease using enhanced probabilistic neural network[J]. Journal of Medical Systems, 2015, 39(11): 179.
- [14] DAS D, LEE C S G. Sample-to-sample correspondence for unsupervised domain adaptation[J]. Engineering Applications of Artificial Intelligence, 2018, 73: 80-91.
- [15] ZHANG H, PATEL V M. Convolutional sparse and low-rank coding-based image decomposition[J]. IEEE Transactions on Image Processing, 2018, 27(5): 2121-2133.
- [16] KONONENKO I. Estimating attributes: Analysis and extensions of RELIEF[C]//European conference on machine learning. Catania: Springer, 1994: 171-182.
- [17] PAN S J, TSANG I W, KWOK J T, et al. Domain adaptation via transfer component analysis[J]. IEEE Transactions on Neural Networks, 2011, 22(2): 199-210.
- [18] 赵鹏, 王美玉, 纪霞, 等. 基于张量表示的域适配的迁移学习中特征表示方法[J]. 电子学报, 2020, 48(2): 359-368.  
ZHAO P, WANG M Y, JI X, et al. A novel feature representation based on tensor and domain adaption for transfer learning[J]. Acta Electronica Sinica, 2020, 48(2): 359-368. (in Chinese)
- [19] BOYD S, PARIKH N. Distributed optimization and statistical learning via the alternating direction method of multipliers[J]. Foundations and Trends in Machine Learning,

2010, 3(1): 1-122.

- [20] SOREL M, SROUBEK F. Fast convolutional sparse coding using matrix inversion lemma[J]. Digital Signal Processing, 2016, 55(1): 44-51.
- [21] CAI X J, GU G Y, HE B S, et al. A proximal point algorithm revisit on the alternating direction method of multipliers[J]. Science China Mathematics, 2013, 56(10): 2179-2186.
- [22] HE X, NIYOGI P. Locality preserving projections[C]// Proceedings of Conference on Advances in Neural Information Processing Systems(NIPS). Vancouver and Whistler: NIPS foundation, 2004: 153-160.
- [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 H H, 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] LI Y M, ZHANG C, JIA Y J, 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). Dalian: IEEE, 2017: 1-6.
- [26] 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.
- [27] BENBA A, JILBAB A, HAMMOUCH A. Analysis of multiple types of voice recordings in cepstral domain using MFCC for discriminating between patients with Parkinson's disease and healthy people[J]. International Journal of Speech Technology, 2016, 19(3): 449-456.
- [28] ALI L, ZHU C, ZHANG Z H, 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.

#### 作者简介



张小恒 男,1980年生,四川达州人. 博士研究生,副教授. 主要研究领域为医学信号处理、机器学习.

E-mail:7818320@qq.com



张馨月 女,1996年生,四川泸县人. 硕士研究生. 主要研究领域为医学信号处理.

E-mail:1029323666@qq.com



李勇明(通信作者) 男,1976年生,四川绵阳人. 博士,教授、博士生导师. 主要研究领域为医学信号处理、机器学习.

E-mail:yongmingli@cqu.edu.cn