

# 转置投影包络线性判别分析

李勇明<sup>1\*</sup>, 赵文强<sup>1</sup>, 李帆<sup>2</sup>, 张小恒<sup>1</sup>, 王品<sup>1</sup>

(1. 重庆大学微电子与通信工程学院, 重庆 400044; 2. 重庆交通大学信息科学与工程学院, 重庆 400074)

**摘要:** 线性判别分析(Linear Discriminant Analysis, LDA)是一种运用广泛的特征提取方法,其以Fisher判别准则为指导,增强子空间中异类样本区分性和同类样本紧凑性,提高了降维结果质量,具有成熟、易解释、简单高效等优点,迄今为止仍是学术界和产业界的研究热点之一。诸多学者对LDA进行了改进以进一步提高其性能,然而这些LDA变体方法直接建模在原样本粒度上,只利用了样本自身存在的信息。由DIK(Data-Information-Knowledge)模型表明,人类获取知识有三个层次,即数据层、信息层,以及知识层,数据首先应当被转换为信息,然后再从信息中学习知识。由人类认知机制表明,信息层不仅包含了原始输入自身信息,还包含了与其相似输入间的关联信息。将其类比于LDA降维过程,提取到的特征即为信息,该信息也应当包含相似样本间的关联信息,以提高下游任务性能。且已有相关研究表明,相似样本间存在的关联信息对机器学习模型构建、知识获取至关重要。即现有LDA存在缺陷,其对样本信息的利用不够完备。针对上述问题,本文提出了转置投影包络线性判别分析(Transposed Projection Envelope Linear Discriminant Analysis, TPELDA)。首先,通过转置投影将原始样本转换为包含相似样本间关联信息的包络样本,转置投影的核心思想为在样本维度上对一批最近邻样本进行降维,使得降维所得包络样本尽量包含该批样本所含信息;随后基于包络样本利用Fisher判别准则学习降维子空间;同时引入分布差异惩罚项确保降维子空间对原始样本的适配性;最后通过联合优化,该方法在考虑相似样本间关联信息的基础上使得投影到子空间中的样本具有更好的判别特征,即该特征同时代表了样本自身信息以及相似样本间存在的关联信息。实验结果表明,TPELDA在给定的多个数据集上相比相关对比方法性能更优,提升范围在2.25%至13.19%之间。此外,结合其他实验结果,表明了本文方法的有效性。

**关键词:** 线性判别分析;降维;关联信息;分布差异;包络学习;特征提取

**基金项目:** 国家自然科学基金(No.62201106)

**中图分类号:** TP391.4

**文献标识码:** A

**文章编号:** 0372-2112(2026)01-0032-18

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

**DOI:** 10.12263/DZXB.20250938

## Transposed Projection Envelope Linear Discriminant Analysis

LI Yongming<sup>1\*</sup>, ZHAO Wenqiang<sup>1</sup>, LI Fan<sup>2</sup>, ZHANG Xiaoheng<sup>1</sup>, WANG Pin<sup>1</sup>

(1. School of Microelectronics and Communication Engineering, Chongqing University, Chongqing 400044, China;

2. School of Information Science and Engineering, Chongqing Jiaotong University, Chongqing 400074, China)

**Abstract:** Linear discriminant analysis (LDA) is a widely used feature extraction method guided by Fisher's discriminant criterion. It enhances the separability of dissimilar samples and the compactness of similar samples within the subspace, thereby improving the quality of dimensionality reduction results. With its mature, interpretable, simple, and efficient advantages, it remains one of the research hotspots in both academia and industry to date. Numerous scholars have refined LDA to further enhance its performance. However, these LDA variants model directly at the original sample granularity, utilizing only the information inherent within the samples themselves. The data-information-knowledge (DIK) model indicates that human knowledge acquisition occurs across three levels: data, information, and knowledge. Data must first be transformed into information, from which knowledge is then learned. Human cognitive mechanisms reveal that the information layer encompasses not only the inherent properties of raw inputs but also correlation information among similar inputs. Analogously, in LDA's dimensionality reduction process, extracted features represent information that should also incorporate correlation information among similar samples to enhance downstream task performance. Furthermore, existing research demonstrates that the correlation information between similar samples is crucial for machine learning model construction and knowledge acquisition. This indicates that existing LDA has limitations, as it does not fully utilize sample information. To address these issues, this paper proposes transposed projection envelope linear discriminant analysis (TPELDA). First, transposed projection transforms original samples into envelope samples that encapsulate correlation informa-

tion among similar samples. The core idea of transposed projection is to reduce the dimensionality of a batch of nearest neighbor samples along the sample dimension, ensuring the resulting envelope sample retain as much information as possible from the original batch. Subsequently, Fisher's discriminant criterion is employed to learn a reduced-dimension subspace based on these envelope samples. A distribution-difference penalty term is introduced to ensure the reduced subspace's adaptability to the original samples. Finally, through joint optimization, this method enhances the discriminative features of samples projected into the subspace by incorporating the correlation information among similar samples. Thus, the resulting features simultaneously represent both the intrinsic information of individual samples and the correlation information among similar samples. Experimental results demonstrate that TPELDA outperforms relevant comparison methods across multiple datasets, achieving performance improvements ranging from 2.25% to 13.19%. Furthermore, combined with other experimental findings, the effectiveness of the proposed method is confirmed.

**Keywords:** linear discriminant analysis; dimensionality reduction; correlation information; distribution discrepancy; envelope learning; feature extraction

**Foundation Item(s):** National Natural Science Foundation of China (No.62201106)

## 0 引言

目前,高维数据的处理在机器学习领域日益普遍,但面临数据稀疏性增大、计算开销增大等“维度诅咒”问题<sup>[1-2]</sup>。特征间的相关性、冗余特征以及存在的噪声往往也会导致机器学习模型性能的下降<sup>[3]</sup>。降维通过将原始数据从高维空间映射到低维子空间,有助于解决上述问题,其至今仍是研究热点之一<sup>[4-5]</sup>。

降维可分为无监督降维和有监督降维。无监督降维在建模过程中未使用样本类别信息。常见的无监督降维方法有:保留主要信息的PCA(Principal Components Analysis)<sup>[6]</sup>、考虑数据流形结构的ISOMAP(Isometric MAPping)<sup>[7]</sup>等。有监督降维在建模过程中使用样本的类别信息。线性判别分析(Linear Discriminant Analysis, LDA)<sup>[8]</sup>作为一种有监督降维方法,其以Fisher判别准则为指导,增强子空间中异类样本区分性和同类样本紧凑性,提高了降维结果的质量,具有成熟、易解释、简单高效等优点。因此,LDA在图像识别、语音识别等方面运用广泛,仍是学术界和产业界的研究热点之一<sup>[9]</sup>。

近年来有许多学者从各方面对LDA进行了改进研究。虽然这些研究取得了一定成效,但大多数着眼于原样本进行建模,未考虑相似样本间存在的关联信息,导致构建的降维矩阵性能有限。有研究表明,挖掘相似样本间的关联信息用于机器学习模型构建,能有效提高模型性能<sup>[10-11]</sup>。因此,本文从考虑相似样本间关联信息出发,对LDA进行改进,提出了转置投影包络线性判别分析(Transposed Projection Envelope LDA, TPELDA)。

本文主要贡献和创新如下:(1)提出转置投影,用于挖掘相似样本间的关联信息,将原样本自身信息以及所挖掘的关联信息富集于生成的包络样本。(2)与现有LDA改进方法基于原样本建模不同,本文方法基于包络样本建模,以实现考虑相似样本间关联信息的投

影降维,为相关研究提供了新思路。(3)为确保降维子空间对原始样本的适配性,引入分布差异损失项,提高模型在原始样本上的泛化性能。(4)本文方法将上述三部分进行联合优化,实现了降维子空间中包络样本的最优构建。所提方法根据相似样本间的关联信息寻找降维子空间,在给定的多个数据集上性能超过了相关对比算法,表明本文方法有效。

## 1 相关工作

近年来,有诸多学者针对原始LDA进行改进以克服其存在的各种问题,一些代表性的改进工作如下。

原始LDA为迹比值问题,不易求解。有学者对其分母添加约束转为另一优化问题,然而两优化问题并不等价,且存在小样本量问题。Mahadi等人<sup>[12]</sup>利用非线性协方差矩阵估计器解决此问题。还有一些方法如OLDA(Orthogonal LDA)<sup>[13]</sup>使用逆代替求解,然而该方式缺乏理论依据。Wang等人<sup>[14]</sup>成功将原始LDA优化问题转化为另一个等价优化问题,并设计一种迭代求解算法,解决了LDA的求解以及小样本量问题。Ye等人<sup>[15]</sup>进一步对该方法的收敛性进行了严格证明。

此外,LDA隐含各类数据由多元高斯分布生成的假设,所以其在具有复杂分布的多模态数据上表现不佳<sup>[16]</sup>。一些学者通过引入邻接矩阵来解决此问题<sup>[17-19]</sup>。Zhu等人<sup>[20]</sup>通过构建内在图和惩罚图考虑数据的结构信息。Zhou等人<sup>[21]</sup>则将LDA建模为一个考虑观察数据和隐藏数据区分性的潜在表示框架,并提出等距结构学习以捕获内在局部结构信息。

另一些学者从减小噪声角度出发对LDA进行改进。Li等人<sup>[22]</sup>将迹比值LDA与 $l_{2,p}$ 范数相结合,同时考虑特征提取以及特征选择。Wen等人<sup>[23]</sup>提出了RobSLDA(Robust Sparse LDA),其对映射矩阵添加 $l_{2,1}$ 范数惩罚,同时还受PCA启发添加重构误差最小约束,并对重构误差进行 $l_1$ 范数约束。Li等人<sup>[24]</sup>提出了类内和类

间核约束,同时考虑欧式距离以及余弦距离,进一步提高了相近数据点间的可分离性和幅度差异。Liu 等人<sup>[25]</sup>在 RobSLDA 的基础上进行改进提出了 RobSLDA+ (RobSLDA Plus),将 RobSLDA 优化目标中的  $l_{2,1}$ 、 $l_1$  范数分别替换为  $l_{2,0}$ 、 $l_0$  范数,以获得更优的抗噪声性能。

还有一些学者也从其他方面对 LDA 进行改进。Yang 等人<sup>[26]</sup>基于调和平均的性质,提出了一种比值和最小化准则,避免了原始迹比值最大化准则中存在的最大目标值支配问题。Li 等人<sup>[27]</sup>通过发掘 LDA 与 K-means<sup>[28]</sup>之间的联系,提出了一种权重自学习的无监督 LDA 方法。Lam 等人<sup>[29]</sup>则基于分类理论误差的新表达式开发了一种 LDA 方法。

总的来说,现有针对 LDA 算法的改进研究已在一定程度上提升了分类性能。然而,这些方法普遍基于原始样本进行建模,忽略了相似样本间潜在的关联信息,尚未从该视角对 LDA 进行系统性优化。

## 2 动机

在知识建构理论中, DIK (Data-Information-Knowledge) 模型<sup>[30]</sup>为理解人类知识获取与转化过程提供了一个层次框架。DIK 模型的示意图如图 1 所示,其由数据层、信息层以及知识层组成。数据代表人类通过感官接收到的原始输入。这些数据本身并无结构或意义,仅仅是对外部世界的原始反映,经过初步加工后,数据被转化为具备上下文和语义的信息。进一步地,人类在不断的经验积累、信息整合中,将信息内化为知识,即能够用于推理、判断的认知结构。类比于机器学习模型,数据对应于模型接收到的原始输入样本。其次,通过特征提取等操作,原始样本被转化为具备结构和上下文的信息。最后,通过学习大量信息,提取其中的潜在模式和规律,形成知识,完成模型的训练。

人类认知机制表明,人类大脑在处理输入时,并非孤立地对该输入进行独立解读,而是倾向于与已有经验和记忆中相似的输入建立关联,以构建更丰富的认知表征<sup>[31]</sup>。这意味着 DIK 模型当中的信息层除了包含对单一输入自身加工所得信息,还包含了该输入与其相似输入间的关联信息。类比于机器学习模型,对样本的处理不应当单独进行,而应当与其相似样本建立关联后再进行处理,以此模式提取到的特征不仅包含了样本自身的信息,还包含了该样本与其相似样本间的关联信息,能进一步提高模型的性能,即机器学习模型的构建应当考虑相似样本间的关联信息。近年来也有研究表明考虑相似样本间的关联信息能有效提高机器学习模型性能。Xia 等人<sup>[32-33]</sup>通过构建粒球包含相似样本间的关联信息,李帆等人<sup>[34]</sup>通过聚类挖掘相似样本间的关联信息,分类性能都有显著提升。Li 等

人<sup>[35]</sup>通过对同属于一个对象的多个帕金森语音样本进行变换以提取它们间的关联信息,显著提高了诊断准确率。

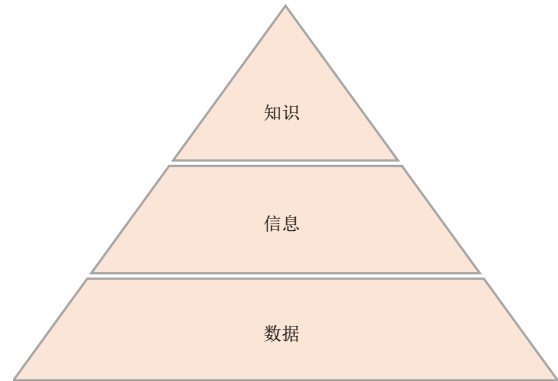


图 1 DIK 模型

Figure 1 The DIK model

基于以上分析,本文通过考虑相似样本间的关联信息,对 LDA 进行改进。首先定义样本的相似样本:与该样本同类别的  $K$  近邻样本视为与其相似的  $K$  个样本。两样本相似,其类别相同为一大前提。其次,利用一种非常朴素的思想来衡量两个样本的相似程度,即欧氏距离,某样本离该样本越近,则认为其与该样本的相似程度越高。将该样本与其  $K$  近邻样本构成的集合称之为该样本的样本包络,样本包络生成过程示意图如图 2①所示。随后利用待学习向量  $\mathbf{p} \in \mathbb{R}^{K+1}$  将该样本与其相似样本进行加权融合以使得该样本与其相似样本建立关联,加权融合生成的新样本包含了该样本自身信息以及与其相似样本间的关联信息,将生成的新样本称之为该样本的包络样本。 $\mathbf{p}$  的学习主要通过转置投影进行。随后基于包络样本进行 Fisher 准则建模学习降维矩阵  $\mathbf{Z}$ ,此种方式获得的降维矩阵  $\mathbf{Z}$  考虑了相似样本间的关联信息,质量得到提升。

然而, $\mathbf{Z}$  基于包络样本建模获得,不适用于原样本。测试样本由于类别未知,按前述定义无法找到其相似样本以构建样本包络,为规避该问题,引入分布差异损失项,要求包络样本在子空间中的分布和原始样本在子空间中的分布差异尽可能小。此外,将转置投影也更改为在子空间进行,能确保映射后包络样本的质量。将上述三部分进行联合优化即构成了本文所提方法,其示意图如图 2②所示。

此外,基于流形学习的降维方法通过构建图也引入了样本间的关联信息,然而本文方法与此类方法具有明显差异:(1)基于流形学习的降维方法对于关联信息的利用为“隐式”,原始目的为结构保持;而本文方法的目的则是要充分利用样本间存在的关联信息以改善性能,即直接利用转置投影进行关联信息的“显

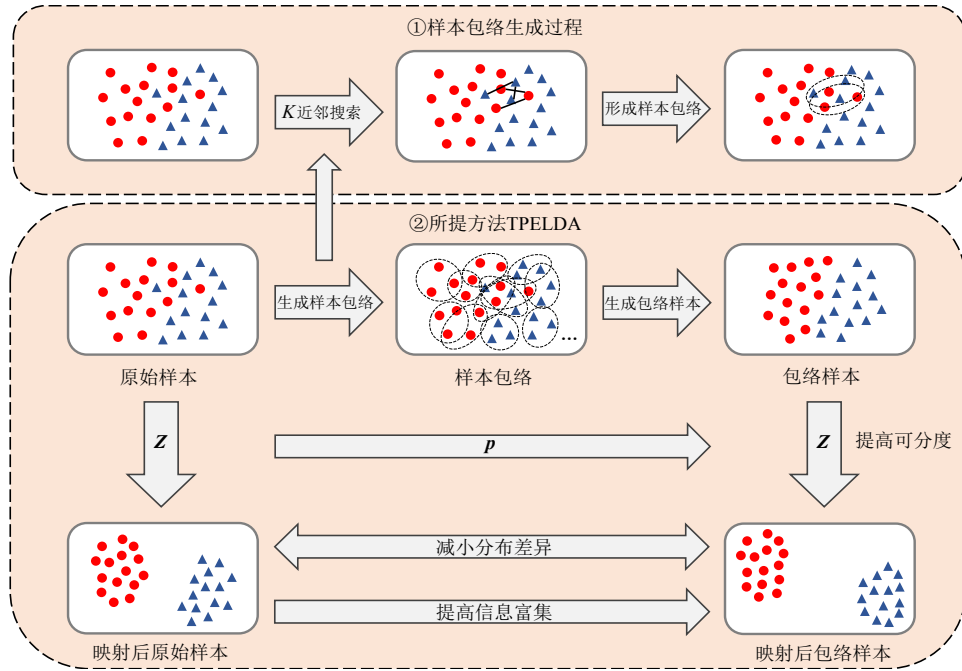


图2 TPELDA思路示意图

Figure 2 Schematic diagram of TPELDA

式”挖掘。(2)基于流形学习的降维方法在建模过程中未生成新样本,而本文方法则生成了包络样本,以利于后续使用相似样本间存在的关联信息,更符合由DIK模型指出的知识获取过程。(3)基于流形学习的降维方法通常为无监督学习,而本文方法为有监督学习,能够提高子空间的判别能力。

### 3 方法

#### 3.1 符号说明

设  $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$  包含所有样本,其中  $d$  为特征数,  $n$  为总样本数。设每类样本构成的集合为  $C_1, C_2, \dots, C_{q_0}$ 。  $|C_i|$  代表第  $i$  类样本的个数。令  $\phi(\cdot): \mathbb{R}^d \rightarrow \mathbb{R}^m$  为一非线性映射函数,用于将输入数据从原始空间  $\mathbb{R}^d$  映射至高维特征空间  $\mathbb{R}^m$ ,以便在该空间中应用线性方法<sup>[36]</sup>,那么有  $\phi(X) = [\phi(\mathbf{x}_1), \phi(\mathbf{x}_2), \dots, \phi(\mathbf{x}_n)] \in \mathbb{R}^{m \times n}$ ,同时定义核函数  $\mathcal{K}(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^\top \phi(\mathbf{x}_j)$ 。令  $Z \in \mathbb{R}^{d' \times m}$  为一线性变换矩阵,用于将样本从高维特征空间  $\mathbb{R}^m$  映射至潜在子空间  $\mathbb{R}^{d'}$ ,  $d'$  即为降维目标维度。令  $\text{tr}(\cdot)$  代表矩阵的迹。 $\|\cdot\|_F$  和  $\|\cdot\|_2$  代表矩阵的  $l_F$  范数以及向量的  $l_2$  范数。上标 T 代表转置。 $I$  代表单位矩阵,  $\mathbf{0}$  代表零矩阵。

#### 3.2 转置投影包络线性判别分析

##### 3.2.1 转置投影项

为挖掘相似样本间的关联信息,首先使用转置投影将原样本变换为包络样本。在特征空间进行建模,对  $\phi(\mathbf{x}_i)$  在原始空间中按照欧式距离寻找与

其标签相同的  $K$  近邻样本,对应特征空间中的  $\phi(\mathbf{x}_{i,1}), \phi(\mathbf{x}_{i,2}), \dots, \phi(\mathbf{x}_{i,K})$ , 这些样本视作其相似样本。记  $E_i = [\phi(\mathbf{x}_i), \phi(\mathbf{x}_{i,1}), \phi(\mathbf{x}_{i,2}), \dots, \phi(\mathbf{x}_{i,K})] \in \mathbb{R}^{m \times (K+1)}$  为  $\phi(\mathbf{x}_i)$  的样本包络。令  $p \in \mathbb{R}^{K+1}$ , 用其对  $E_i$  中的样本进行加权求和以将  $\phi(\mathbf{x}_i)$  变换为包络样本  $\overline{\phi(\mathbf{x}_i)} = E_i p$ , 包络样本的标签和原样本保持一致。

对于包络样本  $\overline{\phi(\mathbf{x}_i)}$  而言,其第  $j$  个特征由  $\phi(\mathbf{x}_i), \phi(\mathbf{x}_{i,1}), \dots, \phi(\mathbf{x}_{i,K})$  的第  $j$  个特征与  $p$  线性加权获得,此举所得包络样本  $\overline{\phi(\mathbf{x}_i)}$  的第  $j$  个特征融合了  $\phi(\mathbf{x}_i)$  在第  $j$  个维度上所含的自身信息以及与其相似样本间在第  $j$  个维度上所含的关联信息。包络样本质量越高,要求每一维度信息融合越多。由于PCA具有保留主要信息这一特点,将其用于指导  $p$  的学习。由PCA可知,当使用  $p$  以及  $\overline{\phi(\mathbf{x}_i)}$  的第  $j$  个特征对  $\phi(\mathbf{x}_i), \phi(\mathbf{x}_{i,1}), \dots, \phi(\mathbf{x}_{i,K})$  的第  $j$  个特征进行重构,重构误差最小时,第  $j$  维度上信息保留达到最大。当所有维度重构误差  $\|E_i - E_i p p^\top\|_F^2$  最小时,此时  $\overline{\phi(\mathbf{x}_i)}$  质量最优。将所有样本都按照这种方式变换为包络样本,那么转置投影需最小化式(1):

$$\text{loss}_{\text{转置投影}} = \sum_{i=1}^n \|E_i - E_i p p^\top\|_F^2 \quad (1)$$

传统PCA投影在特征维度方向进行。包络样本  $\overline{\phi(\mathbf{x}_i)}$  则通过对  $E_i$  在列方向上使用PCA进行投影获得,而由  $E_i$  的构成方式可知,其列方向代表的是样本维度方向,相当于进行了一个转置后再投影,所以该方法称为转置投影。

### 3.2.2 可分度项

基于转置投影获得的包络样本进行 Fisher 准则建模,在考虑相似样本间关联信息的同时提高子空间中包络样本的可分度,以优化子空间的寻找。令  $\mathbf{Z} = \mathbf{Q}\phi(\mathbf{X})^T$ , 其中  $\mathbf{Q} \in \mathbb{R}^{d' \times n}$ 。  $\mathbf{Z}$  用于将包络样本从特征空间  $\mathbb{R}^m$  投影至子空间  $\mathbb{R}^{d'}$ , 在这个子空间中类别相同的包络样本尽量靠近, 类别不同的包络样本尽量远离, 利于分类。在该子空间中, 同属第  $i$  类别的包络样本类中心为  $\frac{1}{|C_i|} \sum_{x_j \in C_i} \mathbf{Z} \overline{\phi(x_j)} = \frac{1}{|C_i|} \sum_{x_j \in C_i} \mathbf{Z} \mathbf{E}_j \mathbf{p}$ , 所有包络样本的中心为  $\frac{1}{n} \sum_{q=1}^n \mathbf{Z} \overline{\phi(x_q)} = \frac{1}{n} \sum_{q=1}^n \mathbf{Z} \mathbf{E}_q \mathbf{p}$ 。根据 Fisher 准则, 类内距离定义为所有样本到其类中心的欧式距离平方之和, 那么投影到该子空间的包络样本类内距离如式(2)所示:

$$\begin{aligned} & \sum_{i=1}^q \sum_{x_i \in C_i} \left( \left( \mathbf{Z} \mathbf{E}_k \mathbf{p} - \frac{1}{|C_i|} \sum_{x_j \in C_i} \mathbf{Z} \mathbf{E}_j \mathbf{p} \right)^T \right. \\ & \quad \left. \cdot \left( \mathbf{Z} \mathbf{E}_k \mathbf{p} - \frac{1}{|C_i|} \sum_{x_j \in C_i} \mathbf{Z} \mathbf{E}_j \mathbf{p} \right) \right) \\ & = \text{tr}(\mathbf{Q} \mathbf{S}_{\text{wQ}} \mathbf{Q}^T) = \text{tr}(\mathbf{p}^T \mathbf{S}_{\text{wp}} \mathbf{p}) \end{aligned} \quad (2)$$

其中,

$$\begin{aligned} \mathbf{S}_{\text{wQ}} &= \sum_{i=1}^q \sum_{x_i \in C_i} \left( \left( \phi(\mathbf{X})^T \mathbf{E}_k - \frac{1}{|C_i|} \sum_{x_j \in C_i} \phi(\mathbf{X})^T \mathbf{E}_j \right) \mathbf{p} \right. \\ & \quad \left. \cdot \mathbf{p}^T \left( \phi(\mathbf{X})^T \mathbf{E}_k - \frac{1}{|C_i|} \sum_{x_j \in C_i} \phi(\mathbf{X})^T \mathbf{E}_j \right) \right) \end{aligned} \quad (3)$$

$$\begin{aligned} \mathbf{S}_{\text{wp}} &= \sum_{i=1}^q \sum_{x_i \in C_i} \left( \left( \phi(\mathbf{X})^T \mathbf{E}_k - \frac{1}{|C_i|} \sum_{x_j \in C_i} \phi(\mathbf{X})^T \mathbf{E}_j \right)^T \right. \\ & \quad \left. \cdot \mathbf{Q} \left( \phi(\mathbf{X})^T \mathbf{E}_k - \frac{1}{|C_i|} \sum_{x_j \in C_i} \phi(\mathbf{X})^T \mathbf{E}_j \right) \right) \end{aligned} \quad (4)$$

根据 Fisher 准则, 类间距离定义为每个类别的类中心到所有样本中心欧式距离的平方乘以该类样本数之和, 那么投影到该子空间的包络样本类间距离如式(5)所示:

$$\begin{aligned} & \sum_{i=1}^q \left( |C_i| \cdot \left( \frac{1}{|C_i|} \sum_{x_j \in C_i} \mathbf{Z} \mathbf{E}_j \mathbf{p} - \frac{1}{n} \sum_{q=1}^n \mathbf{Z} \mathbf{E}_q \mathbf{p} \right)^T \right. \\ & \quad \left. \cdot \left( \frac{1}{|C_i|} \sum_{x_j \in C_i} \mathbf{Z} \mathbf{E}_j \mathbf{p} - \frac{1}{n} \sum_{q=1}^n \mathbf{Z} \mathbf{E}_q \mathbf{p} \right) \right) \\ & = \text{tr}(\mathbf{Q} \mathbf{S}_{\text{bQ}} \mathbf{Q}^T) = \text{tr}(\mathbf{p}^T \mathbf{S}_{\text{bp}} \mathbf{p}) \end{aligned} \quad (5)$$

其中,

$$\begin{aligned} \mathbf{S}_{\text{bQ}} &= \sum_{i=1}^q \left( |C_i| \cdot \left( \frac{1}{|C_i|} \sum_{x_j \in C_i} \phi(\mathbf{X})^T \mathbf{E}_j - \frac{1}{n} \sum_{q=1}^n \phi(\mathbf{X})^T \mathbf{E}_q \right) \mathbf{p} \right. \\ & \quad \left. \cdot \mathbf{p}^T \left( \frac{1}{|C_i|} \sum_{x_j \in C_i} \phi(\mathbf{X})^T \mathbf{E}_j - \frac{1}{n} \sum_{q=1}^n \phi(\mathbf{X})^T \mathbf{E}_q \right)^T \right) \end{aligned} \quad (6)$$

$$\begin{aligned} \mathbf{S}_{\text{bp}} &= \sum_{i=1}^q \left( |C_i| \cdot \left( \frac{1}{|C_i|} \sum_{x_j \in C_i} \phi(\mathbf{X})^T \mathbf{E}_j - \frac{1}{n} \sum_{q=1}^n \phi(\mathbf{X})^T \mathbf{E}_q \right)^T \right. \\ & \quad \left. \cdot \mathbf{Q} \left( \frac{1}{|C_i|} \sum_{x_j \in C_i} \phi(\mathbf{X})^T \mathbf{E}_j - \frac{1}{n} \sum_{q=1}^n \phi(\mathbf{X})^T \mathbf{E}_q \right) \right) \end{aligned} \quad (7)$$

根据 Fisher 准则, 要求类间距离与类内距离的比值尽可能大以使得样本可分度变大, 即最大化式(8):

$$\text{loss}_{\text{可分度}} = \frac{\text{tr}(\mathbf{Q} \mathbf{S}_{\text{bQ}} \mathbf{Q}^T)}{\text{tr}(\mathbf{Q} \mathbf{S}_{\text{wQ}} \mathbf{Q}^T)} = \frac{\text{tr}(\mathbf{p}^T \mathbf{S}_{\text{bp}} \mathbf{p})}{\text{tr}(\mathbf{p}^T \mathbf{S}_{\text{wp}} \mathbf{p})} \quad (8)$$

在式(2)~(7)中出现的  $\phi(\mathbf{X})^T$  与样本包络的乘积可以通过核函数计算, 例如  $\phi(\mathbf{X})^T \mathbf{E}_i = \begin{bmatrix} \mathcal{K}(\mathbf{x}_1, \mathbf{x}_i) & \cdots & \mathcal{K}(\mathbf{x}_1, \mathbf{x}_{i,K}) \\ \cdots & \cdots & \cdots \\ \mathcal{K}(\mathbf{x}_n, \mathbf{x}_i) & \cdots & \mathcal{K}(\mathbf{x}_n, \mathbf{x}_{i,K}) \end{bmatrix}$ 。

### 3.2.3 分布差异项

可分度的建模基于包络样本进行, 这意味着需要将训练样本和测试样本都通过转置投影变换为包络样本。构建样本包络需寻找目标样本的相似样本, 两样本相似, 前提为标签相同, 对于测试样本, 标签未知, 只能在所有样本中寻找  $K$  近邻样本以构成样本包络, 这可能会使得异类样本(即非相似样本)也参与到样本包络的构建, 即产生了错误的样本包络, 进而导致不理想的性能。

为规避上述问题, 引入分布差异损失项, 要求包络样本在子空间中的分布和原始样本在子空间中的分布差异尽可能小。已知包络样本在该子空间中具有较好的分布, 现直接对原始样本使用  $\mathbf{Z}$  进行映射, 由于分布差异的惩罚, 映射之后的原始样本也会获得不错的分布。通过这种方式将原始训练样本使用  $\mathbf{Z}$  映射到子空间作为新训练样本, 将测试样本使用  $\mathbf{Z}$  映射到子空间后再进行测试, 避免了上述问题。

使用线性核的最大均值差异<sup>[37]</sup>平方来衡量包络样本和原始样本在子空间中的分布差异, 即两种样本中心的欧式距离平方, 表示为式(9):

$$\begin{aligned} \text{loss}_{\text{分布差异}} &= \left\| \frac{1}{n} \sum_{k=1}^n \mathbf{Z} \phi(\mathbf{x}_k) - \frac{1}{n} \sum_{j=1}^n \mathbf{Z} \mathbf{E}_j \mathbf{p} \right\|_2^2 \\ &= \left\| \frac{1}{n} \sum_{k=1}^n \mathbf{Q} \mathbf{I}_k - \frac{1}{n} \sum_{j=1}^n \mathbf{Q} \mathbf{K}_j \mathbf{p} \right\|_2^2 \end{aligned} \quad (9)$$

其中,  $\mathbf{I}_k = \phi(\mathbf{X})^T \phi(\mathbf{x}_k)$ ;  $\mathbf{K}_j = \phi(\mathbf{X})^T \mathbf{E}_j$ 。

### 3.2.4 联合优化目标函数

结合式(1)(8)(9),所提方法优化目标可以写为式(10):

$$\min_{\mathbf{p}, \mathbf{Q}} \sum_{i=1}^n \|\mathbf{E}_i - \mathbf{E}_i \mathbf{p} \mathbf{p}^T\|_F^2 - \mu \frac{\text{tr}(\mathbf{Q} \mathbf{S}_{\text{bQ}} \mathbf{Q}^T)}{\text{tr}(\mathbf{Q} \mathbf{S}_{\text{wQ}} \mathbf{Q}^T)} + \gamma \left\| \frac{1}{n} \sum_{k=1}^n \mathbf{Q} \mathbf{I}_k - \frac{1}{n} \sum_{j=1}^n \mathbf{Q} \mathbf{K}_j \mathbf{p} \right\|_2^2 \quad \text{s.t. } \mathbf{p}^T \mathbf{p} = 1, \mathbf{Q} \mathbf{Q}^T = \mathbf{I} \quad (10)$$

其中,  $\mu$  以及  $\gamma$  为两个权衡参数,用于控制不同损失项的重要性。为避免平凡解添加约束  $\mathbf{p}^T \mathbf{p} = 1$  以及  $\mathbf{Q} \mathbf{Q}^T = \mathbf{I}$ 。在式(10)中,如果进一步约束子空间的生成,使映射到该子空间的包络样本仍然能从映射到该子空间的原始样本融合尽可能多的关联信息,那么映射到该子空间的包络样本质量有了进一步保障;从另一方面来看,经映射的原始样本进行了特征提取,优于未映射的原始样本,直接对映射后的原始样本进行转置投影变换为映射后的包络样本,所得映射后包络样本质量也得到提升。所以将式(10)改为式(11),作为所提方法最终优化目标:

$$\min_{\mathbf{p}, \mathbf{Q}} \sum_{i=1}^n \|\mathbf{Q} \mathbf{K}_i - \mathbf{Q} \mathbf{K}_i \mathbf{p} \mathbf{p}^T\|_F^2 - \mu \frac{\text{tr}(\mathbf{Q} \mathbf{S}_{\text{bQ}} \mathbf{Q}^T)}{\text{tr}(\mathbf{Q} \mathbf{S}_{\text{wQ}} \mathbf{Q}^T)} + \gamma \left\| \frac{1}{n} \sum_{k=1}^n \mathbf{Q} \mathbf{I}_k - \frac{1}{n} \sum_{j=1}^n \mathbf{Q} \mathbf{K}_j \mathbf{p} \right\|_2^2 \quad \text{s.t. } \mathbf{p}^T \mathbf{p} = 1, \mathbf{Q} \mathbf{Q}^T = \mathbf{I} \quad (11)$$

### 3.3 优化

所提方法优化目标式(11)中含有迹比值项,直接该式进行优化并不容易,受文献[14]启发,通过引入一个新的待优化变量  $a$ ,可以将式(11)的优化等价于优化式(12):

$$\min_{\mathbf{p}, \mathbf{Q}, a} a^2 \text{tr}(\mathbf{Q} \mathbf{S}_{\text{wQ}} \mathbf{Q}^T) - 2a \sqrt{\mu \text{tr}(\mathbf{Q} \mathbf{S}_{\text{bQ}} \mathbf{Q}^T)} + \sum_{i=1}^n \|\mathbf{Q} \mathbf{K}_i - \mathbf{Q} \mathbf{K}_i \mathbf{p} \mathbf{p}^T\|_F^2 + \gamma \left\| \frac{1}{n} \sum_{k=1}^n \mathbf{Q} \mathbf{I}_k - \frac{1}{n} \sum_{j=1}^n \mathbf{Q} \mathbf{K}_j \mathbf{p} \right\|_2^2 \quad \text{s.t. } \mathbf{p}^T \mathbf{p} = 1, \mathbf{Q} \mathbf{Q}^T = \mathbf{I} \quad (12)$$

在式(12)中,原有的迹比值项不再出现,优化目标形式变得简单。通过定理1可以证明优化式(12)等价于优化式(11)。在式(12)中需要优化的变量有  $\mathbf{p}, \mathbf{Q}, a$ , 可以通过分块坐标下降法<sup>[38]</sup>进行优化。

#### 定理 1

优化式(12)等价于优化式(11)。

#### 证明

为证明简洁,将式(11)记为  $\min_{\mathbf{p}^T \mathbf{p} = 1, \mathbf{Q} \mathbf{Q}^T = \mathbf{I}} J(\mathbf{p}, \mathbf{Q})$ , 将式(12)记为  $\min_{\mathbf{p}^T \mathbf{p} = 1, \mathbf{Q} \mathbf{Q}^T = \mathbf{I}, a} J_1(\mathbf{p}, \mathbf{Q}, a)$ 。式(12)对  $a$  的梯度为

$$\frac{\partial J_1(\mathbf{p}, \mathbf{Q}, a)}{\partial a} = 2a \text{tr}(\mathbf{Q} \mathbf{S}_{\text{wQ}} \mathbf{Q}^T) - 2 \sqrt{\mu \text{tr}(\mathbf{Q} \mathbf{S}_{\text{bQ}} \mathbf{Q}^T)} \quad (13)$$

令  $\frac{\partial J_1(\mathbf{p}, \mathbf{Q}, a^*)}{\partial a^*} = 0$ , 则有:

$$a^* = \frac{\sqrt{\mu \text{tr}(\mathbf{Q} \mathbf{S}_{\text{bQ}} \mathbf{Q}^T)}}{\text{tr}(\mathbf{Q} \mathbf{S}_{\text{wQ}} \mathbf{Q}^T)} \quad (14)$$

进而有:

$$J_1(\mathbf{p}, \mathbf{Q}, a^*) \leq J_1(\mathbf{p}, \mathbf{Q}, a) \quad \forall \mathbf{p}, \mathbf{Q}, a, \mathbf{p}^T \mathbf{p} = 1, \mathbf{Q} \mathbf{Q}^T = \mathbf{I} \quad (15)$$

其中,当  $a = a^*$  时,等号成立。根据上式,对于  $\forall \mathbf{p}, \mathbf{Q}, a, \mathbf{p}^T \mathbf{p} = 1, \mathbf{Q} \mathbf{Q}^T = \mathbf{I}$ , 要使得  $J_1(\mathbf{p}, \mathbf{Q}, a)$  最小,必然满足式(14),所以最小化了  $J_1(\mathbf{p}, \mathbf{Q}, a) \mathbf{p}^T \mathbf{p} = 1, \mathbf{Q} \mathbf{Q}^T = \mathbf{I}$ ,  $J_1(\mathbf{p}, \mathbf{Q}, a^*) \mathbf{p}^T \mathbf{p} = 1, \mathbf{Q} \mathbf{Q}^T = \mathbf{I}$  的值也达到了最小。而  $J_1(\mathbf{p}, \mathbf{Q}, a^*) = J(\mathbf{p}, \mathbf{Q})$ 。那么有:

$$\min_{\mathbf{p}^T \mathbf{p} = 1, \mathbf{Q} \mathbf{Q}^T = \mathbf{I}, a} J_1(\mathbf{p}, \mathbf{Q}, a) \Rightarrow \min_{\mathbf{p}^T \mathbf{p} = 1, \mathbf{Q} \mathbf{Q}^T = \mathbf{I}} J_1(\mathbf{p}, \mathbf{Q}, a^*) \Rightarrow \min_{\mathbf{p}^T \mathbf{p} = 1, \mathbf{Q} \mathbf{Q}^T = \mathbf{I}} J(\mathbf{p}, \mathbf{Q}) \quad (16)$$

即优化式(12)等价于优化式(11)。证毕。

#### 3.3.1 更新 $a$

在  $t+1$  次迭代中,固定变量  $\mathbf{p}, \mathbf{Q}$ , 丢弃与  $a$  的无关项,并通过式(4)(7)更新  $\mathbf{S}_{\text{wp}}, \mathbf{S}_{\text{bp}}$  为  $\mathbf{S}_{\text{wp}}^{t+1}, \mathbf{S}_{\text{bp}}^{t+1}$ ,  $a$  通过式(17)进行更新:

$$a_{t+1} = \min_a a^2 \text{tr}(\mathbf{p}_t^T \mathbf{S}_{\text{wp}}^{t+1} \mathbf{p}_t) - 2a \sqrt{\mu \text{tr}(\mathbf{p}_t^T \mathbf{S}_{\text{bp}}^{t+1} \mathbf{p}_t)} \quad (17)$$

直接对  $a$  求梯度并令梯度为0即可,即  $a$  的更新如式(18)所示:

$$2a_{t+1} \text{tr}(\mathbf{p}_t^T \mathbf{S}_{\text{wp}}^{t+1} \mathbf{p}_t) - 2 \sqrt{\mu \text{tr}(\mathbf{p}_t^T \mathbf{S}_{\text{bp}}^{t+1} \mathbf{p}_t)} = 0 \Rightarrow a_{t+1} = \frac{\sqrt{\mu \text{tr}(\mathbf{p}_t^T \mathbf{S}_{\text{bp}}^{t+1} \mathbf{p}_t)}}{\text{tr}(\mathbf{p}_t^T \mathbf{S}_{\text{wp}}^{t+1} \mathbf{p}_t)} \quad (18)$$

#### 3.3.2 更新 $\mathbf{p}$

在  $t+1$  次迭代中,固定变量  $\mathbf{Q}, a$ , 丢弃与  $\mathbf{p}$  的无关项,  $\mathbf{p}$  通过式(19)进行更新:

$$\begin{aligned} \mathbf{p}_{t+1} &= \min_{\mathbf{p}^T \mathbf{p} = 1} a_{t+1}^2 \text{tr}(\mathbf{p}^T \mathbf{S}_{\text{wp}}^{t+1} \mathbf{p}) - 2a_{t+1} \sqrt{\mu \text{tr}(\mathbf{p}^T \mathbf{S}_{\text{bp}}^{t+1} \mathbf{p})} \\ &+ \sum_{i=1}^n \|\mathbf{Q}_i \mathbf{K}_i - \mathbf{Q}_i \mathbf{K}_i \mathbf{p} \mathbf{p}^T\|_F^2 + \gamma \left\| \frac{1}{n} \sum_{k=1}^n \mathbf{Q}_i \mathbf{I}_k - \frac{1}{n} \sum_{j=1}^n \mathbf{Q}_i \mathbf{K}_j \mathbf{p} \right\|_2^2 \\ &= \max_{\mathbf{p}^T \mathbf{p} = 1} \text{tr}(\mathbf{p}^T (\mathbf{sI} - \mathbf{A}_{t+1}) \mathbf{p}) + \text{tr}(\mathbf{b}_{t+1}^T \mathbf{p}) \\ &+ 2a_{t+1} \sqrt{\mu \text{tr}(\mathbf{p}^T \mathbf{S}_{\text{bp}}^{t+1} \mathbf{p})} \end{aligned} \quad (19)$$

其中,

$$\begin{aligned} \mathbf{A}_{t+1} &= a_{t+1}^2 \mathbf{S}_{\text{wp}}^t - \sum_{i=1}^n \mathbf{K}_i^T \mathbf{Q}_i^T \mathbf{Q}_i \mathbf{K}_i \\ &+ \gamma \frac{1}{n^2} \left( \sum_{j=1}^n \mathbf{Q}_i \mathbf{K}_j \right)^T \left( \sum_{j=1}^n \mathbf{Q}_i \mathbf{K}_j \right) \end{aligned} \quad (20)$$

$$\mathbf{b}_{t+1}^T = \gamma \frac{2}{n^2} \left( \sum_{k=1}^n \mathbf{Q}_i \mathbf{I}_k \right)^T \left( \sum_{j=1}^n \mathbf{Q}_i \mathbf{K}_j \right) \quad (21)$$

$s$  为一常数且足够大,使得  $s\mathbf{I}-\mathbf{A}_{t+1}$  为半正定矩阵。优化问题 (19) 可以使用文献 [39] 中的 GRA (General Reweighted Algorithm) 进行优化。GRA 适用于如下形式的最优化问题:

$$\max_{\mathbf{x} \in \Omega} f(\mathbf{x}) + h(\mathbf{x}) \quad (22)$$

其中,  $h(\mathbf{x})$  为凸函数,其通过多次迭代以使得式 (22) 收敛到最大值。迭代方式如式 (23) 所示:

$$\mathbf{x}_{t+1} = \max_{\mathbf{x} \in \Omega} f(\mathbf{x}) + \text{tr}(h'(\mathbf{x}_t)^T \mathbf{x}) \quad (23)$$

其中,  $h'(\mathbf{x}_t)$  为  $h(\mathbf{x})$  在  $\mathbf{x}_t$  处的梯度。在式 (19) 中,由于  $s\mathbf{I}-\mathbf{A}_{t+1}$  为半正定矩阵,那么  $\text{tr}(\mathbf{p}^T(s\mathbf{I}-\mathbf{A}_{t+1})\mathbf{p}) + \text{tr}(\mathbf{b}_{t+1}^T \mathbf{p})$  为凸函数,为满足 GRA 的使用条件,下面通过定理 2 证明式 (19) 中的  $2a_{t+1}\sqrt{\mu \text{tr}(\mathbf{p}^T \mathbf{S}_{\text{bp}}^{t+1} \mathbf{p})}$  也为一凸函数。

### 定理 2

$2a_{t+1}\sqrt{\mu \text{tr}(\mathbf{p}^T \mathbf{S}_{\text{bp}}^{t+1} \mathbf{p})}$  为一凸函数。

### 证明

$\mathbf{S}_{\text{bp}}^{t+1}$  为半正定矩阵,分解为  $\mathbf{S}_{\text{bp}}^{t+1} = \mathbf{O}\mathbf{O}^T$ 。则有:

$2a_{t+1}\sqrt{\mu \text{tr}(\mathbf{p}^T \mathbf{S}_{\text{bp}}^{t+1} \mathbf{p})} = 2a_{t+1}\sqrt{\mu} \|\mathbf{O}^T \mathbf{p}\|_2$ , 由于范数为凸函数,对自变量进行仿射变换保凸,所以  $\|\mathbf{O}^T \mathbf{p}\|_2$  为凸函数,  $a, \mu$  非负,所以  $2a_{t+1}\sqrt{\mu} \|\mathbf{O}^T \mathbf{p}\|_2$  为凸函数,继而原函数为凸函数。证毕。

凸函数与凸函数之和仍然为凸函数,所以式 (19) 优化目标为凸函数,满足 GRA 的使用条件,对  $\mathbf{p}$  求梯度有

$$\begin{aligned} & \frac{\partial \left( \text{tr}(\mathbf{p}^T(s\mathbf{I}-\mathbf{A}_{t+1})\mathbf{p}) + \text{tr}(\mathbf{b}_{t+1}^T \mathbf{p}) + 2a_{t+1}\sqrt{\mu \text{tr}(\mathbf{p}^T \mathbf{S}_{\text{bp}}^{t+1} \mathbf{p})} \right)}{\partial \mathbf{p}} \\ & = 2(s\mathbf{I}-\mathbf{A}_{t+1})\mathbf{p} + \mathbf{b}_{t+1} + 2a_{t+1}\mu \mathbf{S}_{\text{bp}}^{t+1} \mathbf{p} / \sqrt{\mu \text{tr}(\mathbf{p}^T \mathbf{S}_{\text{bp}}^{t+1} \mathbf{p})} \end{aligned} \quad (24)$$

GRA 第  $r+1$  次迭代时,  $\mathbf{p}$  的更新方式如式 (25):

$$\begin{aligned} \mathbf{p}_{r+1} &= \max_{\mathbf{p}^T \mathbf{p} = 1} \text{tr} \left( \left( 2(s\mathbf{I}-\mathbf{A}_{t+1})\mathbf{p}_r + \mathbf{b}_{t+1} \right. \right. \\ & \quad \left. \left. + 2a_{t+1}\mu \mathbf{S}_{\text{bp}}^{t+1} \mathbf{p}_r / \sqrt{\mu \text{tr}(\mathbf{p}_r^T \mathbf{S}_{\text{bp}}^{t+1} \mathbf{p}_r)} \right)^T \mathbf{p} \right) \\ & \Rightarrow \mathbf{p}_{r+1} = \left\| \frac{2(s\mathbf{I}-\mathbf{A}_{t+1})\mathbf{p}_r + \mathbf{b}_{t+1} + \frac{2a_{t+1}\mu \mathbf{S}_{\text{bp}}^{t+1} \mathbf{p}_r}{\sqrt{\mu \text{tr}(\mathbf{p}_r^T \mathbf{S}_{\text{bp}}^{t+1} \mathbf{p}_r)}}}{\left\| 2(s\mathbf{I}-\mathbf{A}_{t+1})\mathbf{p}_r + \mathbf{b}_{t+1} + \frac{2a_{t+1}\mu \mathbf{S}_{\text{bp}}^{t+1} \mathbf{p}_r}{\sqrt{\mu \text{tr}(\mathbf{p}_r^T \mathbf{S}_{\text{bp}}^{t+1} \mathbf{p}_r)} \right\|_2} \right\|_2 \end{aligned} \quad (25)$$

所以,优化问题式 (19) 的求解可以通过式 (24) (25) 得到,其完整过程如算法 1 所示。

### 算法 1 基于 GRA 求解优化问题式 (19)

输入:  $s, \mathbf{A}_{t+1}, \mathbf{b}_{t+1}, \mathbf{S}_{\text{bp}}^{t+1}, \mu, a_{t+1}$

输出:  $\mathbf{p}_{t+1}$

1. 初始化  $\mathbf{p}_{t+1}$  满足  $\mathbf{p}_{t+1}^T \mathbf{p}_{t+1} = 1$
2. 重复
3. 根据式 (24) 计算目标函数在  $\mathbf{p}_{t+1}$  处的梯度
4. 根据式 (25) 将上一步所得梯度进行归一化并赋值给  $\mathbf{p}_{t+1}$
5. 直到收敛

### 3.3.3 更新 $\mathbf{Q}$

在  $t+1$  次迭代中,固定变量  $\mathbf{p}, a$ , 通过式 (3) (6) 更新  $\mathbf{S}_{\text{wQ}}, \mathbf{S}_{\text{bQ}}$  为  $\mathbf{S}_{\text{wQ}}^{t+1}, \mathbf{S}_{\text{bQ}}^{t+1}$ , 丢弃与  $\mathbf{Q}$  的无关项,  $\mathbf{Q}$  通过式 (26) 进行更新:

$$\begin{aligned} \mathbf{Q}_{t+1} &= \min_{\mathbf{Q}^T = \mathbf{I}} a_{t+1}^2 \text{tr}(\mathbf{Q} \mathbf{S}_{\text{wQ}}^{t+1} \mathbf{Q}^T) - 2a_{t+1} \sqrt{\mu \text{tr}(\mathbf{Q} \mathbf{S}_{\text{bQ}}^{t+1} \mathbf{Q}^T)} \\ & \quad + \sum_{i=1}^n \|\mathbf{Q} \mathbf{K}_i - \mathbf{Q} \mathbf{K}_i \mathbf{p}_{t+1} \mathbf{p}_{t+1}^T\|_F^2 + \gamma \left\| \frac{1}{n} \sum_{k=1}^n \mathbf{Q} \mathbf{I}_k - \frac{1}{n} \sum_{j=1}^n \mathbf{Q} \mathbf{K}_j \mathbf{p}_{t+1} \right\|_2^2 \\ & = \max_{\mathbf{Q}^T = \mathbf{I}} \text{tr}(\mathbf{Q}(s_1 \mathbf{I} - \mathbf{C}_{t+1}) \mathbf{Q}^T) + 2a_{t+1} \sqrt{\mu \text{tr}(\mathbf{Q} \mathbf{S}_{\text{bQ}}^{t+1} \mathbf{Q}^T)} \end{aligned} \quad (26)$$

其中,

$$\begin{aligned} \mathbf{C}_{t+1} &= a_{t+1}^2 \mathbf{S}_{\text{wQ}}^{t+1} + \sum_{i=1}^n \mathbf{K}_i \mathbf{K}_i^T - \mathbf{K}_i \mathbf{p}_{t+1} \mathbf{p}_{t+1}^T \mathbf{K}_i^T \\ & \quad + \gamma \frac{1}{n^2} \left( \sum_{k=1}^n \mathbf{I}_k - \sum_{j=1}^n \mathbf{K}_j \mathbf{p}_{t+1} \right) \left( \sum_{k=1}^n \mathbf{I}_k - \sum_{j=1}^n \mathbf{K}_j \mathbf{p}_{t+1} \right)^T \end{aligned} \quad (27)$$

$s_1$  为常数且足够大,使得  $s_1 \mathbf{I} - \mathbf{C}_{t+1}$  为半正定矩阵。参照定理 2 的证明过程同样可以得出  $2a_{t+1}\sqrt{\mu \text{tr}(\mathbf{Q} \mathbf{S}_{\text{bQ}}^{t+1} \mathbf{Q}^T)}$  为凸函数,而由于  $s_1 \mathbf{I} - \mathbf{C}_{t+1}$  为半正定矩阵,所以  $\text{tr}(\mathbf{Q}(s_1 \mathbf{I} - \mathbf{C}_{t+1}) \mathbf{Q}^T)$  为凸函数,继而  $\text{tr}(\mathbf{Q}(s_1 \mathbf{I} - \mathbf{C}_{t+1}) \mathbf{Q}^T) + 2a_{t+1}\sqrt{\mu \text{tr}(\mathbf{Q} \mathbf{S}_{\text{bQ}}^{t+1} \mathbf{Q}^T)}$  为凸函数,满足 GRA 的使用条件,在该式中对  $\mathbf{Q}$  求梯度有

$$\begin{aligned} & \frac{\partial \left( \text{tr}(\mathbf{Q}(s_1 \mathbf{I} - \mathbf{C}_{t+1}) \mathbf{Q}^T) + 2a_{t+1}\sqrt{\mu \text{tr}(\mathbf{Q} \mathbf{S}_{\text{bQ}}^{t+1} \mathbf{Q}^T)} \right)}{\partial \mathbf{Q}} \\ & = 2\mathbf{Q}(s_1 \mathbf{I} - \mathbf{C}_{t+1}) + 2a_{t+1} \frac{\mu \mathbf{Q} \mathbf{S}_{\text{bQ}}^{t+1}}{\sqrt{\mu \text{tr}(\mathbf{Q} \mathbf{S}_{\text{bQ}}^{t+1} \mathbf{Q}^T)}} \end{aligned} \quad (28)$$

GRA 第  $r+1$  次迭代时,  $\mathbf{Q}$  的更新方式如式 (29):

$$\begin{aligned} \mathbf{Q}_{r+1} &= \max_{\mathbf{Q}^T = \mathbf{I}} \text{tr} \left( \left( 2\mathbf{Q}_r(s_1 \mathbf{I} - \mathbf{C}_{t+1}) \right. \right. \\ & \quad \left. \left. + 2a_{t+1}\mu \mathbf{Q}_r \mathbf{S}_{\text{bQ}}^{t+1} / \sqrt{\mu \text{tr}(\mathbf{Q}_r \mathbf{S}_{\text{bQ}}^{t+1} \mathbf{Q}_r^T)} \right)^T \mathbf{Q} \right) \end{aligned} \quad (29)$$

进行奇异值分解获得  $(2\mathbf{Q}_r(s_1 \mathbf{I} - \mathbf{C}_{t+1}) +$

$2a_{t+1} \frac{\mu \mathbf{Q}_r \mathbf{S}_{bQ}^{t+1}}{\sqrt{\mu \text{tr}(\mathbf{Q}_r \mathbf{S}_{bQ}^{t+1} \mathbf{Q}_r^T)}} \Big)^T = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T$ , 则优化问题式(29)的

最优解为  $\mathbf{Q}_{r+1} = \mathbf{V}[\mathbf{I}, \mathbf{0}] \mathbf{U}^T$ 。证明见定理3。

### 定理3

对  $\left( 2\mathbf{Q}_r (s_1 \mathbf{I} - \mathbf{C}_{t+1}) + 2a_{t+1} \frac{\mu \mathbf{Q}_r \mathbf{S}_{bQ}^{t+1}}{\sqrt{\mu \text{tr}(\mathbf{Q}_r \mathbf{S}_{bQ}^{t+1} \mathbf{Q}_r^T)}} \right)^T$  进行

奇异值分解获得  $\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}^T$ , 则优化问题式(29)的最优解为  $\mathbf{Q}_{r+1} = \mathbf{V}[\mathbf{I}, \mathbf{0}] \mathbf{U}^T$ 。

### 证明

式(29)目标函数可以重新写为  $\text{tr}(\mathbf{U} \mathbf{\Sigma} \mathbf{V}^T \mathbf{Q}) = \text{tr}(\mathbf{\Sigma} \mathbf{V}^T \mathbf{Q} \mathbf{U})$ , 令  $\mathbf{Y} = \mathbf{V}^T \mathbf{Q} \mathbf{U}$ , 则  $\mathbf{Y} \mathbf{Y}^T = (\mathbf{V}^T \mathbf{Q} \mathbf{U})(\mathbf{V}^T \mathbf{Q} \mathbf{U})^T = \mathbf{I}$ 。

$\text{tr}(\mathbf{\Sigma} \mathbf{V}^T \mathbf{Q} \mathbf{U}) = \sum_{i=1}^{d'} \Sigma_{(i,i)} \mathbf{Y}_{(i,i)}$ ,  $\Sigma_{(i,i)}, \mathbf{Y}_{(i,i)}$  分别代表  $\mathbf{\Sigma}, \mathbf{Y}$  的第  $i$  个对角线元素, 由于  $\mathbf{Y} \mathbf{Y}^T = \mathbf{I}$ , 则  $|\mathbf{Y}_{(i,i)}| \leq 1$ 。考虑到奇异值  $\Sigma_{(i,i)}$  非负, 要使得  $\sum_{i=1}^{d'} \Sigma_{(i,i)} \mathbf{Y}_{(i,i)}$  最大, 则  $\mathbf{Y}_{(i,i)} = 1$ , 即  $\mathbf{Y} = [\mathbf{I}, \mathbf{0}]$ ,

根据  $\mathbf{Y} = \mathbf{V}^T \mathbf{Q} \mathbf{U}$  得出最优解  $\mathbf{Q}_{r+1} = \mathbf{V}[\mathbf{I}, \mathbf{0}] \mathbf{U}^T$ 。证毕。

因此, 优化问题式(26)的求解过程如算法2所示。

#### 算法2 基于GRA求解优化问题式(26)

输入:  $s_1, \mathbf{C}_{t+1}, \mathbf{S}_{bQ}^{t+1}, \mu, a_{t+1}$

输出:  $\mathbf{Q}_{t+1}$

1. 初始化  $\mathbf{Q}_{t+1}$  满足  $\mathbf{Q}_{t+1} \mathbf{Q}_{t+1}^T = \mathbf{I}$
2. 重复
3. 根据式(28)计算目标函数在  $\mathbf{Q}_{t+1}$  处的梯度转置
4. 对所得梯度转置进行奇异值分解获得  $\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}^T$
5. 将  $\mathbf{Q}_{t+1}$  更新为  $\mathbf{V}[\mathbf{I}, \mathbf{0}] \mathbf{U}^T$
6. 直到收敛

通过交替更新相关变量, 所提方法可按照算法3进行迭代求解。

#### 算法3 基于分块坐标下降法求解所提方法

输入:  $\mathbf{X} \in \mathbb{R}^{d \times n}, d', K, \gamma, \mu$

输出:  $\mathbf{Q}, \mathbf{p}$

1. 构建样本包络并根据核函数计算  $\mathbf{L}_i, \mathbf{K}_i, i = 1, 2, \dots, n$
2. 初始化  $\mathbf{Q}, \mathbf{p}$  满足  $\mathbf{Q} \mathbf{Q}^T = \mathbf{I}, \mathbf{p}^T \mathbf{p} = 1$
3. 重复
4. 根据式(4)(7)更新  $\mathbf{S}_{wp}, \mathbf{S}_{bp}$
5. 根据式(18)更新变量  $a$
6. 根据式(20)(21)更新  $\mathbf{A}, \mathbf{b}$
7. 执行算法1更新变量  $\mathbf{p}$
8. 根据式(3)(6)更新  $\mathbf{S}_{wQ}, \mathbf{S}_{bQ}$
9. 根据式(27)更新  $\mathbf{C}$
10. 执行算法2更新变量  $\mathbf{Q}$
11. 直到收敛

### 3.4 复杂度分析

所提方法计算时间主要由距离矩阵计算耗时、核矩阵计算耗时、执行算法3计算耗时组成。

构建样本包络时需计算距离矩阵, 时间复杂度为  $\mathcal{O}(n^2 d)$ 。为了构建  $\mathbf{L}_i, \mathbf{K}_i, i = 1, 2, \dots, n$ , 需要计算核矩阵, 本文核函数选用高斯核函数, 即  $\mathcal{K}(\mathbf{x}_i, \mathbf{x}_j) = e^{-\lambda \|\mathbf{x}_i - \mathbf{x}_j\|^2}$ , 时间复杂度为  $\mathcal{O}(n^2 d)$ 。在算法3中: 算法1单次迭代的时间复杂度为  $\mathcal{O}(K^2)$ ; 算法2单次迭代的时间复杂度为  $\mathcal{O}(nd'^2 + n^2 d')$ ; 算法3迭代一次除开执行算法1、2的时间复杂度为  $\mathcal{O}(nd'K^2 + n^2 d'K + n^3)$ , 实际运用中有  $K \ll n$ , 所以进一步约等于  $\mathcal{O}(n^2 d'K + n^3)$ 。此外, 更新  $\mathbf{C}$  涉及到计算  $\sum \mathbf{K}_i \mathbf{K}_i^T$ , 其时间复杂度为  $\mathcal{O}(n^3 K)$ , 这一项可以提前计算而不参与迭代。

设算法1、2、3分别固定迭代  $w_1, w_2, w_3$  次结束, 那么所提方法的时间复杂度为  $\mathcal{O}(n^2 d + n^3 K + w_1 w_3 K^2 + w_2 w_3 n d'^2 + w_2 w_3 n^2 d' + w_3 n^2 d' K + w_3 n^3)$ 。

## 4 实验结果与分析

### 4.1 实验环境

11个数据集被用来验证所提方法的有效性, 其中包含8个结构化数据集和3个图像数据集。所选数据集涵盖多个领域, 被广泛运用于测试机器学习方法性能。表1具体给出了各数据集信息。在表1中, 前8个数据集为结构化数据集, 均来源于UCI Machine Learning Repository (<https://archive.ics.uci.edu/>); 后3个数据集为图像数据集。对于Jaffe数据集 (<http://www.kasrl.org/JAFFE.html>), 其包含了来自10个人的7种面部表情图像。对于Coil20数据集 (<http://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php>), 其包含了20个不同物体在360度旋转中不同角度成像的灰度图像, 每个物体具有72种不同的姿势。对于Umist数据集 (<https://www.visioneng.org.uk/datasets/>), 其包含了20个人的多张面部图像, 这些图像在不同的姿态下拍摄。

对于结构化数据集, 训练集和测试集按照7:3的比例划分; 对于图像数据集, 从每类中分别选取相同数量的样本作为训练集, 剩余样本作为测试集, 具体为: 对于Jaffe数据集, 测试从每类中挑选8、10以及12个样本作为训练样本时的性能; 对于Coil20数据集, 测试从每类中挑选10、15以及20个样本作为训练样本时的性能; 对于Umist数据集, 测试从每类中挑选12、15以及18个样本作为训练样本时的性能。

在实验中, 所提方法与近年来发表的其他LDA改进算法进行性能对比, 以验证其有效性。对比方法包括: TR LDA (Trace Ratio LDA)<sup>[14]</sup>、RSLDA (Ratio Sum LDA)<sup>[40]</sup>、RSMLDA (Ratio Sum Minimization-based

表 1 数据集信息

Table 1 Dataset information

数据集名称	样本数量	特征数量	类别数量	类型
Ionosphere	351	33	2	雷达数据
Control	600	60	6	控制数据
Urban	675	147	9	环境数据
Sonar	208	60	2	声纳数据
Dermatology	366	34	6	医学数据
German	1 000	20	2	信用数据
Isolet	1 650	617	26	语音数据
Madelon	2 600	500	2	人工数据
Jaffe	213	1 024	10	图像数据
Coil20	1 440	1 024	20	图像数据
Umist	575	10 304	20	图像数据

LDA)<sup>[26]</sup>、GRLDA(Generalized Robust LDA)<sup>[20]</sup>、ROBSLDA+(ROBust and Sparse LDA plus)<sup>[25]</sup>以及 SWULDA (Self-Weighted Unsupervised LDA)<sup>[27]</sup>。此外,考虑到基于流形学习的降维方法也会涉及到样本间的关联关系,所以额外添加两种相应算法进行对比。它们分别是 SPFDR (Structure Preserved Fast Dimensionality Reduction)<sup>[41]</sup>以及 EGCFS(Embedded Graph learning and Constrained Feature Selection)<sup>[42]</sup>。

为使实验结果具有说服力,实验重复十次,以十次重复随机实验的平均分类准确率作为评价指标。分类器选用最近邻分类器。对于维度较大的 Isolet、Madelon、Jaffe、Coil20 以及 Umist 数据集,提前使用 PCA 进行降维预处理以提高计算效率,降维时要求保留 98% 的能量。实验所使用的电脑配置为 Intel(R) Core(TM) i7-12700F CPU、64 GB RAM,操作系统为 Windows 10 20H2,所使用的软件为 MATLAB R2019a。

## 4.2 参数设置

五折交叉验证被用来为所有方法选择合适的参数。对于对比算法,其参数设置与参数搜索范围与原文保持一致。对于本文算法,参数有  $K$  以及  $\gamma, \mu, \lambda$ 。针对不同的数据集,通过网络搜索法寻找最优参数组合。表 2 给出了所提方法在 Ionosphere 数据集上针对不同  $K$  设置下的性能表现情况,  $K$  的取值范围设置为 [1, 2, 3, 4, 5, 6, 7]。图 3 给出了所提方法在 Ionosphere 数据集上针对不同  $\gamma, \mu, \lambda$  设置下的性能表现情况,  $\gamma, \mu$  的取值范围设置为  $[10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10^1, 10^2, 10^3]$ ,  $\lambda$  的取值范围设置为 [0.001, 0.005, 0.01, 0.05, 0.1, 0.5]。此外,所提方法在更多数据集上的参数实验结果见 <https://github.com/SEMC7/Supplementary-Materials>。

表 2 所提方法在 Ionosphere 数据集上针对不同  $K$  的平均准确率 $\pm$ 标准差

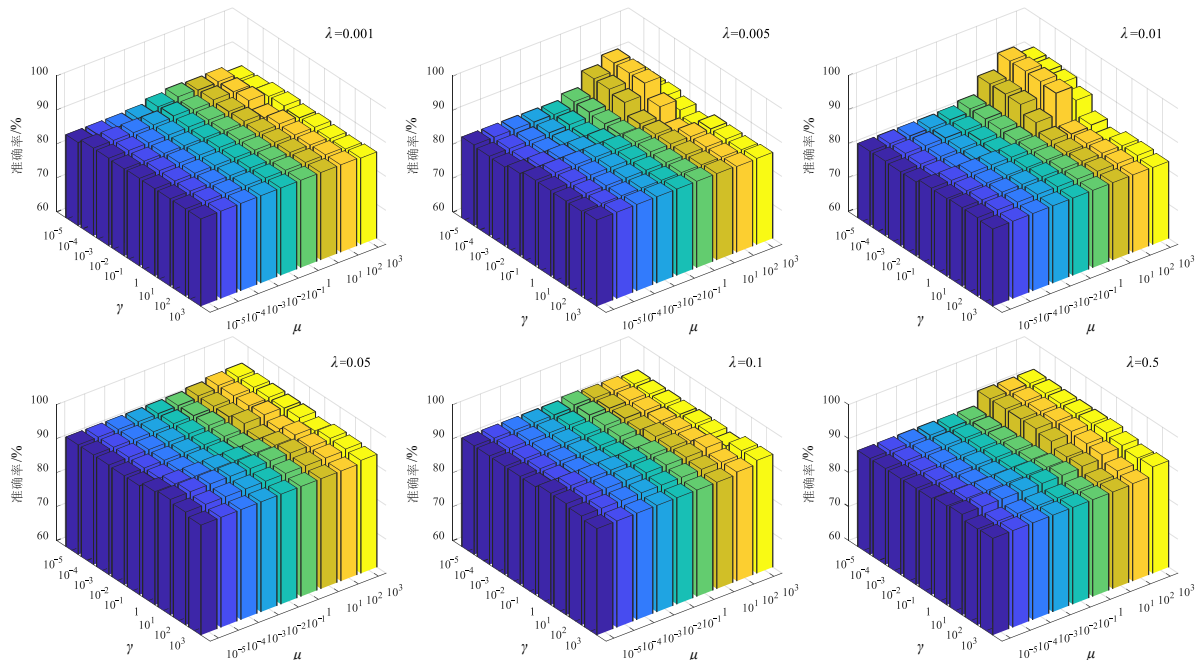
单位: %

Table 2 Average accuracy  $\pm$  standard deviation of proposed method on Ionosphere dataset for different  $K$  values

unit: %

数据集名称	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$	$K=6$	$K=7$
Ionosphere	90.67 $\pm$ 2.34	90.48 $\pm$ 2.86	91.14 $\pm$ 2.00	92.00 $\pm$ 1.42	<b>92.95<math>\pm</math>2.46</b>	90.29 $\pm$ 3.06	89.81 $\pm$ 2.02

注:最优结果以粗体显示。

图 3 所提方法在 Ionosphere 数据集上针对不同  $\gamma, \mu, \lambda$  的性能表现情况Figure 3 Performance of the proposed method on the Ionosphere dataset under different  $\gamma, \mu, \lambda$  settings

由表 2 可知,  $K$  的取值对方法性能有一定影响, 具体表现为随着  $K$  的增大, 方法性能呈现上升趋势,  $K=5$  时性能达到最大, 随后开始下降。具体原因如下: 当  $K$  较小时, 随着  $K$  的增大, 样本包络中的相似样本越多, 生成的包络样本富集了更多的关联信息, 方法性能不断上升; 当  $K$  到达一定程度时, 随着  $K$  的增大, 引入样本包络中的样本相似程度不断下降, 甚至可能是受噪声影响的假相似样本, 此时生成的包络样本质量受到影响, 方法性能不断下降。由图 3 可知,  $\lambda$  的取值对方法性能影响明显, 从高斯核函数形式  $\mathcal{K}(\mathbf{x}_i, \mathbf{x}_j) = e^{-\lambda \|\mathbf{x}_i - \mathbf{x}_j\|^2}$  来看, 若  $\lambda$  取得过大, 核函数的值将会趋近于 0, 而取得过小, 不同样本在特征空间中的点积又将变得区别不大。 $\gamma, \mu$  对方法性能的影响与  $\lambda$  的取值有关, 总体而言, 其值不能取得过大, 否则会与其他损失项产生竞争, 也不能取得过小, 否则对应损失项会丧失其作用。从图 3 结果来看,  $\lambda$  取 0.05 以及 0.1 时, Ionosphere 数据集上的效果比较好, 且此时所提方法对参数  $\gamma, \mu$  较为鲁棒。具体而

言, 对于 Ionosphere 数据集,  $K$  设置为 5,  $\lambda, \gamma, \mu$  分别设置为 0.05、0.01 以及 10。

### 4.3 消融实验

为验证所提方法损失函数各组成部分的有效性, 本节在多个数据集上进行消融实验。记录对照组和所提方法在这些数据集上所有降维维度下的最高分类准确率。分别有以下对照组: (1) 只考虑转置投影损失以及可分度损失的对比方法  $\text{loss}_{\text{转置投影}} + \text{loss}_{\text{可分度}}$ , 以验证分布差异损失的作用。由于未添加分布差异损失, 所以将训练样本和测试样本在子空间中变换为包络样本后进行分类, 此时会出现前文所提的测试样本变换问题。(2) 只考虑转置投影损失和分布差异损失的对比方法  $\text{loss}_{\text{转置投影}} + \text{loss}_{\text{分布差异}}$ , 以验证可分度损失的作用。(3) 只考虑可分度损失的对比方法  $\text{loss}_{\text{可分度}}$ , 此时  $\mathbf{p}$  固定为第一个元素为 1, 其余元素为 0 的向量。该对照组用于验证基于包络样本进行 Fisher 准则建模的作用。实验结果如表 3 所示。

表 3 所提方法消融实验结果

Table 3 Ablation experiment results for proposed method

数据集	$\text{loss}_{\text{转置投影}} + \text{loss}_{\text{可分度}}$	$\text{loss}_{\text{转置投影}} + \text{loss}_{\text{分布差异}}$	$\text{loss}_{\text{可分度}}$	$\text{loss}_{\text{转置投影}} + \text{loss}_{\text{可分度}} + \text{loss}_{\text{分布差异}}$ (TPELDA)
Ionosphere	86.19±2.26	92.86±2.26	92.95±2.63	<b>94.19±1.45</b>
Urban	75.99±2.65	70.40±3.48	73.12±1.24	<b>78.81±2.15</b>
Sonar	76.77±4.88	85.97±2.85	80.81±4.33	<b>86.13±3.66</b>
German	<b>71.30±1.77</b>	65.87±2.94	69.63±1.89	71.17±2.69
Isolet	86.82±1.88	91.62±1.08	86.60±1.88	<b>91.97±1.10</b>
Jaffe(8)	92.11±1.88	98.27±1.42	93.53±2.75	<b>99.02±1.23</b>
Jaffe(10)	93.19±1.77	98.85±1.45	93.45±1.83	<b>99.20±0.88</b>
Jaffe(12)	92.69±1.50	<b>99.14±1.32</b>	91.51±2.93	<b>99.14±1.32</b>
Umist(12)	52.33±2.77	86.69±2.54	78.48±2.38	<b>88.75±3.42</b>
Umist(15)	50.95±2.95	87.24±1.80	83.67±0.95	<b>89.42±1.59</b>
Umist(18)	44.88±2.01	84.14±3.29	80.56±0.90	<b>85.77±2.96</b>

注: 平均准确率%±标准差%, 最优结果以粗体显示。

从表 3 可以看出: (1) 除了 German 数据集,  $\text{loss}_{\text{转置投影}} + \text{loss}_{\text{可分度}} + \text{loss}_{\text{分布差异}}$  相比  $\text{loss}_{\text{转置投影}} + \text{loss}_{\text{可分度}}$  的性能都有明显提升, 体现了分布差异损失项的作用, 在一定程度上解决了测试样本变换为包络样本的问题。在 German 数据集中, 添加  $\text{loss}_{\text{分布差异}}$  后性能有所下降, 考虑到下降幅度微小, 仅为 0.13%, 且在更加细致地调节  $\gamma, \mu$  后, 性能上界仍然保持 71.17% 不变, 那么造成该问题的原因为分布差异项的假设在该数据集上不完全适用。(2) 除了 Jaffe 数据集在每类中选取 12 个样本作为训练样本的情况下,  $\text{loss}_{\text{转置投影}} + \text{loss}_{\text{可分度}} + \text{loss}_{\text{分布差异}}$  相比  $\text{loss}_{\text{转置投影}} + \text{loss}_{\text{分布差异}}$  的性能都有一定提升, 体现了可分度损失项的作用。(3) 在所有数据集上  $\text{loss}_{\text{转置投影}} + \text{loss}_{\text{可分度}} + \text{loss}_{\text{分布差异}}$  相比  $\text{loss}_{\text{可分度}}$  的性能都有明显提升, 体现了基于包络样本进行 Fisher 准则建模

相比于基于原样本进行 Fisher 准则建模效果更好。

### 4.4 算法对比

#### 4.4.1 基于结构化数据集的对比分析

对于结构化数据集, 各方法在所有降维维度下的最高分类准确率结果见表 4, 同时还测试了不进行任何特征提取直接分类的分类准确率(列“None”), 各方法的分类准确率-降维维度折线图见图 4。

观察表 4 可知, 所提方法在 8 个结构化数据集集中的 7 个数据集达到了前三水平。在 Ionosphere 数据集上, 所提方法相比次优方法 SPFDR 提升了 3%, 在 Control 数据集上所提方法相比次优方法 SPFDR 提升 0.5%, 在 Sonar 数据集上所提方法相比次优方法 EGCFs 有将近 0.2% 的提升。在 Dermatology、German 以及 Isolet 数据集上所提方法相比于最优方法性能差距并不明显。此

表 4 各算法在结构化数据集上的平均准确率±标准差

单位: %

Table 4 Average accuracy ± standard deviation of each algorithm on structured datasets

unit: %

数据集	None	TRLDA	RSLDA	RSMLDA	GRLDA	ROBSLDA+	SWULDA	SPFDR	EGCFS	TPELDA
Ionosphere	87.43±2.45	89.43±2.13	89.71±2.87	88.76±1.67	90.19±2.01	89.71±2.61	89.24±2.58	<u>90.95±2.07</u>	88.29±1.49	<b>94.19±1.45</b>
Control	96.33±1.09	96.94±1.02	97.56±0.75	95.00±1.92	96.39±1.02	97.72±0.41	96.33±1.09	<u>98.11±0.99</u>	97.83±1.32	<b>98.67±1.02</b>
Urban	73.32±2.21	73.86±2.83	<u>79.60±2.27</u>	43.51±3.07	74.90±1.81	<b>81.34±1.98</b>	73.47±2.32	72.23±3.66	76.14±2.73	78.81±2.15
Sonar	83.39±3.81	84.52±4.82	83.39±3.81	79.52±4.10	84.52±3.06	85.48±3.95	84.68±4.25	77.58±6.15	<u>85.97±2.41</u>	<b>86.13±3.66</b>
Dermatology	94.31±1.97	94.40±1.96	<b>96.88±1.08</b>	94.5±1.67	95.87±1.80	95.69±0.97	94.59±2.78	94.68±1.72	96.06±1.56	<u>96.33±0.97</u>
German	68.00±2.71	68.80±2.09	70.70±2.40	<b>71.5±1.65</b>	68.97±1.70	69.90±2.93	68.03±2.90	67.67±2.13	68.00±2.71	<u>71.17±2.69</u>
Isolet	87.35±1.08	87.35±1.08	<b>92.52±0.82</b>	90.49±1.70	87.88±0.91	89.96±1.05	88.31±1.71	84.55±1.44	87.35±1.08	<u>91.97±1.10</u>
Madelon	64.94±2.28	<u>82.69±1.57</u>	64.94±2.28	56.40±1.04	<b>82.90±1.22</b>	65.88±2.01	82.31±1.57	65.44±2.22	64.94±2.28	77.72±1.74

注:最优结果用粗体显示,次优结果用下划线显示。

外,所提方法在 Ionosphere 数据集上相比其他对比方法基本都有 4%~6% 左右的提升,在 Control 数据集上相比 TR LDA、GRLDA 以及 SWULDA 有 2% 左右的提升,在 Urban 数据集上相比 TR LDA、GRLDA、SPFDR 以及 SWULDA 有 4%~6% 左右的提升,在 Sonar 数据集上相比 RSLDA 有接近 3% 的提升,在 Isolet 数据集上相比 TR LDA、GRLDA、ROBSLDA+、SWULDA 以及 EGCFS 有 2%~4% 左右的提升,在 Madelon 数据集上相比 RSLDA、ROBSLDA+、SPFDR 以及 EGCFS 有 12% 左右的提升。

观察图 4 可知,在 Ionosphere 数据集上,当降维维度大于 6 时,所提方法优于其他所有对比方法。在 Control 数据集上,当降维维度大于 9 时,所提方法优于其他所有对比方法。在 Urban 数据集上,基本在所有降维维度中,所提方法优于 TR LDA、GRLDA、SWULDA、RSMLDA、SPFDR 以及 EGCFS。在 Sonar 数据集上,基本在所有降维维度中,所提方法优于 RSLDA、RSMLDA 以及 SPFDR。在 Dermatology 数据集上,基本在所有降维维度中,所提方法优于 TR LDA、GRLDA、SWULDA、RSMLDA 以及 EGCFS。在 German 数据集上,基本在所有降维维度中,所提方法优于 TR LDA、GRLDA、SWULDA、ROBSLDA+、SPFDR 以及 EGCFS。在 Isolet 数据集上,基本在所有降维维度中,所提方法优于除了 RSLDA 以外的其余所有对比方法。在 Madelon 数据集

上,基本在所有降维维度中,所提方法优于 RSLDA、RSMLDA、ROBSLDA+、SPFDR 以及 EGCFS。综上所述,这些结果体现了所提方法在这些结构化数据集上有一定的效果。

此外,在 Urban 数据集以及 Madelon 数据集上,所提方法相比表现最好方法性能分别弱了 2.5% 以及 5% 左右,有一定差距。虽然所提方法相比表现最好方法性能有一定差距,但是相比不进行降维直接分类性能有 5.49% 以及 12.78% 的提升,这充分说明所提方法在这两个数据集上有效,造成性能不为最优的原因可能为:最优方法所作的假设可能更加符合这两个数据集的特性,所以性能有更大提升。且对于 Madelon 数据集,其可能还存在如下原因: Madelon 为人工数据集,只有 20 个特征有效,其余 480 个特征为手动添加的干扰特征,在这种情况下,很难保证通过欧式距离获取到的相似样本具有很高的相似度,这将导致生成的包络样本并未富集太多相似样本间的关联信息,所以性能提升有限。

#### 4.4.2 基于图像数据集的对比分析

对于图像数据集,各方法在所有降维维度下的最高分类准确率结果见表 5,同时还测试了不进行任何特征提取直接分类的分类准确率(列“None”),各方法的分类准确率-降维维度折线图见图 5。

表 5 各算法在图像数据集上的平均准确率±标准差

单位: %

Table 5 Average accuracy ± standard deviation of each algorithm on image datasets

unit: %

数据集	None	TRLDA	RSLDA	RSMLDA	GRLDA	ROBSLDA+	SWULDA	SPFDR	EGCFS	TPELDA
Jaffe(8)	96.62±2.10	96.62±2.10	<b>99.25±0.79</b>	72.41±8.27	96.62±2.10	98.80±1.01	96.62±2.10	71.73±8.54	96.62±2.10	<u>99.02±1.23</u>
Jaffe(10)	<u>96.37±2.26</u>	<u>96.37±2.26</u>	<b>99.20±1.28</b>	90.00±5.02	<u>96.37±2.26</u>	99.20±1.14	<u>96.37±2.26</u>	89.03±7.46	<u>96.37±2.26</u>	99.20±0.88
Jaffe(12)	95.81±2.80	96.02±2.73	<b>99.46±1.04</b>	91.40±4.89	95.91±2.57	99.03±1.29	95.81±2.98	93.12±3.56	95.81±2.80	<u>99.14±1.32</u>
Coil20(10)	91.79±2.38	92.26±2.53	<u>93.91±2.35</u>	54.31±6.06	93.42±2.01	88.27±2.71	91.79±2.38	58.73±5.81	93.43±2.12	<b>94.25±2.60</b>
Coil20(15)	95.45±1.79	96.16±2.05	<u>96.90±1.76</u>	80.12±6.81	96.85±1.68	94.14±2.60	95.45±1.79	84.38±3.49	96.67±1.70	<b>97.35±1.42</b>
Coil20(20)	95.63±1.58	96.26±1.62	96.81±1.79	86.75±3.23	<u>96.88±1.58</u>	95.32±1.71	95.63±1.58	87.83±2.45	96.76±1.51	<b>97.17±1.57</b>
Umist(12)	71.70±3.37	71.88±4.42	<u>86.24±1.95</u>	52.75±4.59	72.33±4.14	73.61±5.19	73.58±2.85	43.94±3.36	72.51±3.75	<b>88.75±3.42</b>
Umist(15)	69.31±0.49	71.05±1.14	<u>84.04±1.51</u>	60.44±3.62	71.16±0.62	72.95±2.42	72.65±1.61	49.93±2.47	70.73±0.88	<b>89.42±1.59</b>
Umist(18)	61.02±0.29	64.05±0.76	<u>79.07±1.03</u>	52.65±2.68	63.26±0.38	65.58±1.18	66.47±2.00	41.81±3.50	68.28±0.98	<b>85.77±2.96</b>

注:最优结果用粗体显示,次优结果用下划线显示。

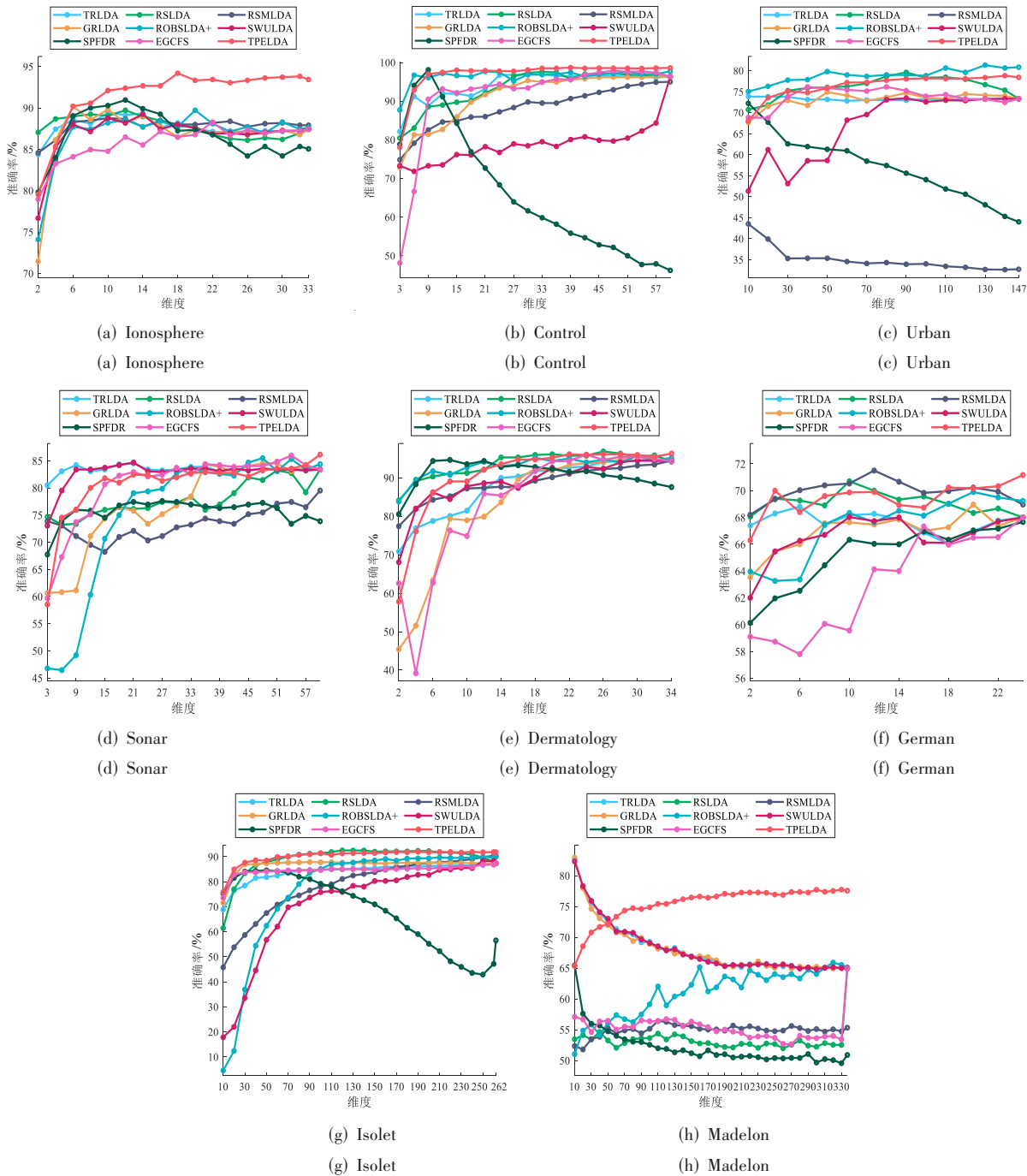


图4 结构化数据集上的分类准确率-维度折线图

Figure 4 Accuracy versus dimension curves on structured datasets

观察表 5 可知,所提方法在 Jaffe 数据集上排名第二,弱于 RSLDA,但两者性能差距并不明显。在 Coil20 以及 Umist 数据集上,所提方法为最优方法。在 Jaffe 数据集上,所提方法相比于 TRLDA、GRLDA、SWULDA 以及 EGCFS 有 3% 左右的提升。在 Coil20 数据集上相比 TRLDA 以及 SWULDA 有 1%~2% 左右的提升;当从每类中选取 10 个样本作为训练样本的情况下,相比于

ROBSLDA+ 有 6% 左右的提升,其余情况下有 2%~3% 左右的提升。值得注意的是,在 Umist 数据集上,所提方法相比于第二好的方法 RSLDA 有 2%~6% 左右的提升;当从每类中选取 12、15 以及 18 个样本作为训练样本的情况下,相比于 TRLDA、GRLDA、ROBSLDA+、SWULDA 以及 EGCFS 有 15%~17%、17%~18% 以及 17%~22% 左右的提升。

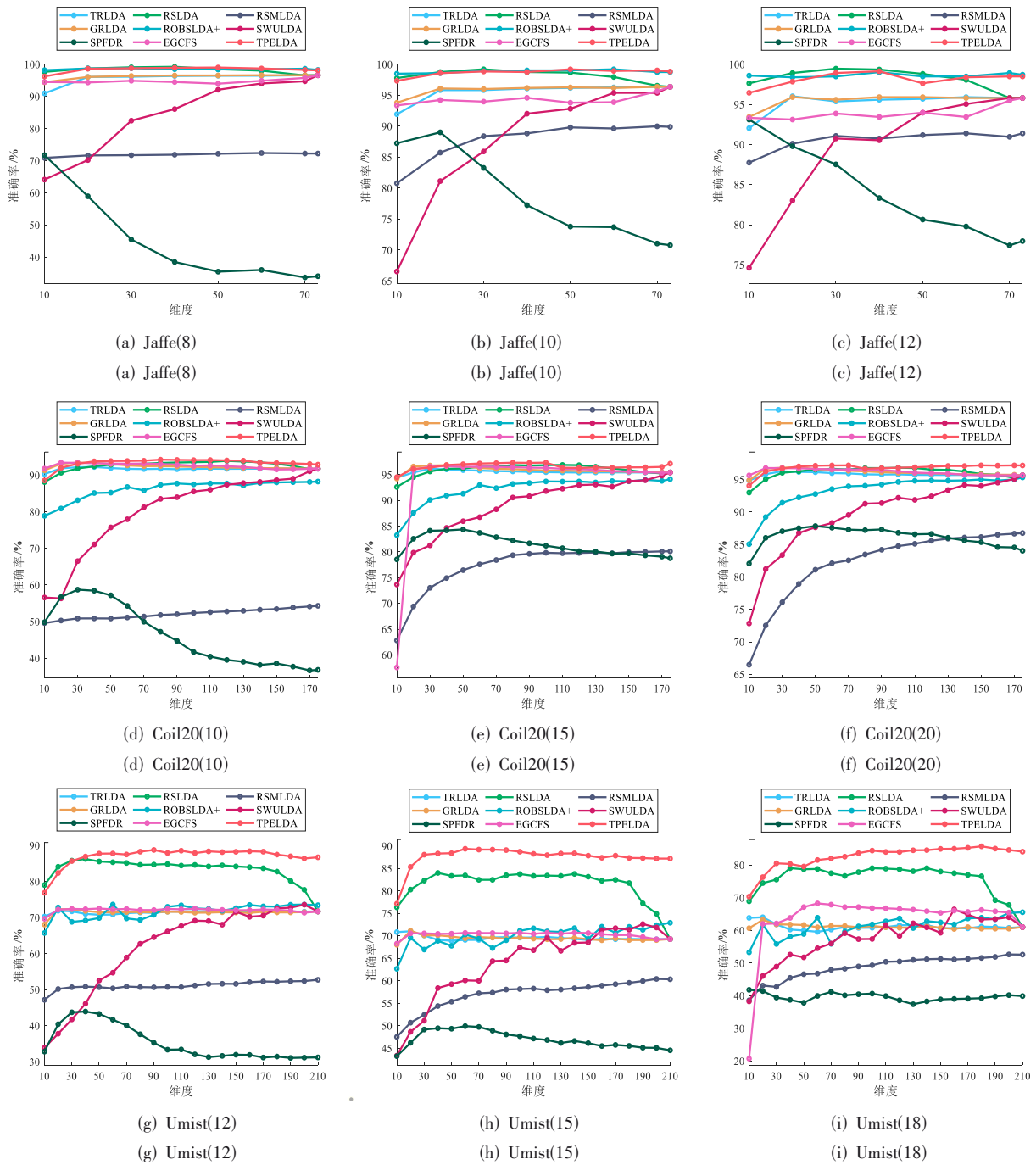


图5 图像数据集上的分类准确率-维度折线图

Figure 5 Accuracy versus dimension curves on image datasets

观察图5可知,对于Jaffe数据集,在所有降维维度中,所提方法优于TRLDA、GRLDA、SWULDA、RSMLDA、SPFDR以及EGCFS。对于Coil20和Umist数据集,基本在所有降维维度中,所提方法优于其余所有对比方法。综上所述,这些结果体现了所提方法在这些图像数据集上有一定效果,尤其对于Umist数据集,提升明显。

#### 4.4.3 非参数检验

为评估所提方法与其他对比方法性能差异的统计

显著性,进行弗里德曼检验分析。各方法的平均排名如表6。

通过表6可以发现所提方法的平均排名最小。弗里德曼检验结果见表7,显著性水平 $\alpha$ 设置为0.05。

由表7可知 $p$ 值远小于 $\alpha$ ,说明各方法间的差异在统计学上显著。Wilcoxon符号秩检验用于考察成对算法间的差异,检验结果如表8,显著性水平 $\alpha$ 设置为0.05。表中小于 $\alpha$ 的 $p$ 值加粗,表明这两种方法性能有显著差异。

表 6 各算法弗里德曼秩比较表

Table 6 Friedman rank comparison of each algorithm

数据集	None	TRLDA	RSLDA	RSMLDA	GRLDA	ROBSLDA+	SWULDA	SPFDR	EGCFS	TPELDA
Ionosphere	10	6	4.5	8	3	4.5	7	2	9	1
Control	8.5	6	5	10	7	4	8.5	2	3	1
Urban	8	6	2	10	5	1	7	9	4	3
Sonar	7.5	5.5	7.5	9	5.5	3	4	10	2	1
Dermatology	10	9	1	8	4	5	7	6	3	2
German	8.5	6	3	1	5	4	7	10	8.5	2
Isolet	8	8	1	3	6	4	5	10	8	2
Madelon	8	2	8	10	1	5	3	6	8	4
Jaffe	7.5	4	1	10	5	3	7.5	9	6	2
Coil20	6.5	5	2	10	3	8	6.5	9	4	1
Umist	8	6	2	9	7	4	3	10	5	1
平均排名	8.23	5.77	<u>3.36</u>	8	4.68	4.14	5.95	7.55	5.50	<b>1.82</b>

注:最优结果用粗体显示,次优结果用下划线显示。

表 7 各算法弗里德曼检验统计量

Table 7 Friedman test statistics of each algorithm

观测数	卡方值( $\chi^2$ )	自由度	$p$ 值
11	46.95	9	$4.01 \times 10^{-7}$

由表 8 可知,所提方法显著优于除了 RSLDA 外的 TR LDA、RSMLDA、GRLDA、ROBSLDA+、SWULDA 以及两种基于流形学习的降维方法 SPFDR 以及 EGCFS,而 RSLDA 只显著优于 RSMLDA、SPFDR 和 EGCFS。所以综合来看,在给定的这些数据集中,所提方法在这些对比方法中最优,体现了其有效性。

此外,TPELDA 性能优于基于流形学习的 SPFDR 以及 EGCFS,这与前文的分析一致,虽然基于流形学习的降维方法通过构建图也能引入样本间关联信息,但这种利用是“隐式”的,构建图的目的是以结构保持为主,所以性能并未超过直接使用转置投影进行关联信息“显式”挖掘的 TPELDA。

表 8 各算法 Wilcoxon 符号秩检验结果

Table 8 Wilcoxon signed-rank test results of each algorithm

方法	None	TRLDA	RSLDA	RSMLDA	GRLDA	ROBSLDA+	SWULDA	SPFDR	EGCFS	TPELDA
None	—	<b>0.002 0</b>	<b>0.003 9</b>	0.071 0	<b>0.001 0</b>	<b>0.006 8</b>	<b>0.007 8</b>	0.174 8	<b>0.015 6</b>	<b>0.001 0</b>
TRLDA	<b>0.002 0</b>	—	0.123 0	<b>0.042 0</b>	0.064 5	0.206 1	0.591 8	<b>0.024 4</b>	0.431 6	<b>0.042 0</b>
RSLDA	<b>0.003 9</b>	0.123 0	—	<b>0.002 0</b>	0.174 8	0.375 0	0.101 6	<b>0.013 7</b>	<b>0.037 1</b>	0.140 6
RSMLDA	0.071 3	<b>0.042 0</b>	<b>0.002 0</b>	—	<b>0.032 2</b>	<b>0.009 8</b>	<b>0.032 2</b>	0.700 2	<b>0.042 0</b>	<b>0.002 0</b>
GRLDA	<b>0.001 0</b>	0.064 5	0.174 8	<b>0.032 2</b>	—	0.413 1	0.206 1	<b>0.009 8</b>	1	<b>0.042 0</b>
ROBSLDA+	<b>0.006 8</b>	0.206 1	0.375 0	<b>0.009 8</b>	0.413 1	—	0.240 2	<b>0.009 8</b>	0.206 1	<b>0.018 6</b>
SWULDA	<b>0.007 8</b>	0.591 8	0.101 6	<b>0.032 2</b>	0.206 1	0.240 2	—	<b>0.042 0</b>	0.492 2	<b>0.024 4</b>
SPFDR	0.174 8	<b>0.024 4</b>	<b>0.013 7</b>	0.700 2	<b>0.009 8</b>	<b>0.009 8</b>	<b>0.042 0</b>	—	<b>0.032 2</b>	<b>0.001 0</b>
EGCFS	<b>0.015 6</b>	0.431 6	<b>0.037 1</b>	<b>0.042 0</b>	1	0.206 1	0.492 2	<b>0.032 2</b>	—	<b>0.001 0</b>
TPELDA	<b>0.001 0</b>	<b>0.042 0</b>	0.140 6	<b>0.002 0</b>	<b>0.042 0</b>	<b>0.018 6</b>	<b>0.024 4</b>	<b>0.001 0</b>	<b>0.001 0</b>	—

注:小于 $\alpha$ 的 $p$ 值加粗显示。

#### 4.5 降维结果可视化

在本节中,对各方法在 Umist 数据集上的降维结果利用  $t$ -SNE ( $t$ -distributed Stochastic Neighbor Embedding)<sup>[43]</sup>降至 2 维以进行可视化分析。图 6 展示了 Umist 数据集中所属 10 个类别的样本分布。

观察图 6(a)以及图 6(j)可知,所提方法将原样本进行映射后同类样本相互靠近,异类样本相互远离,即可分度变大,所以分类结果得到提升,符合预期。图 6(b)(e)(g)(i)中的样本可分度相比原样本图 6(a)并无明显改善,对应了 TR LDA、GRLDA、SWULDA 以及 EGCFS 在该数据集上分类准确率无明显提升的结果。此外,对于图 6(h),虽然样本可分度有一定提升,但仔细观察可以发现部分类别样本发生重叠,这也对应了 SPFDR 在该数据集上性能表现并不理想。值得注意的是,图 6(c)(d)(f)当中的同类样本分布比图 6(j)当中的同类样本分布更为紧凑,但 RSLDA、RSMLDA 以及

ROBSLDA+在该数据集上的性能并未超过所提方法,可能有如下原因:(1)虽然图6(c)(d)(f)当中的同类样本分布更为紧凑,但异类样本分布不如图6(j)分散。(2)可分度并不是衡量分布好坏的唯一标准,可分度越高

也并不代表泛化性能越好,应当以实际性能表现为准。(3)可视化结果通过另一种降维方法  $t$ -SNE 再次降维获得,可能并不能真实反应样本在子空间中的实际分布,应当更加关注方法的实际定量性能表现情况。

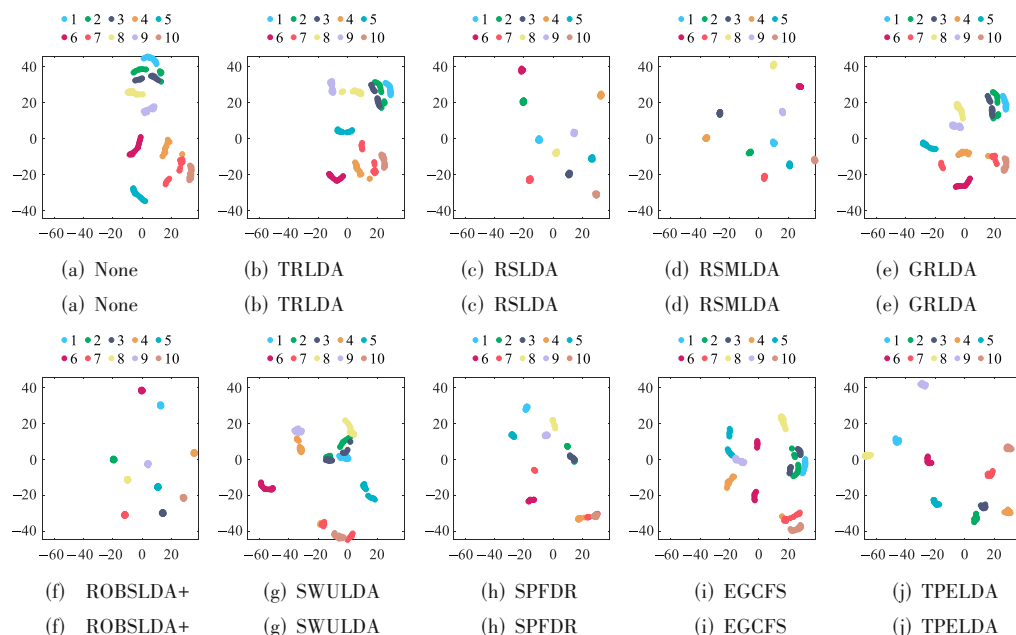


图6 Umist数据集上各算法  $t$ -SNE 可视化结果

Figure 6  $t$ -SNE visualization results of each algorithm on the Umist dataset

#### 4.6 收敛曲线可视化

图7展示了在Sonar以及Coil20数据集上的实际收

敛曲线。其中算法1、算法2以及算法3的迭代次数都设置为50。

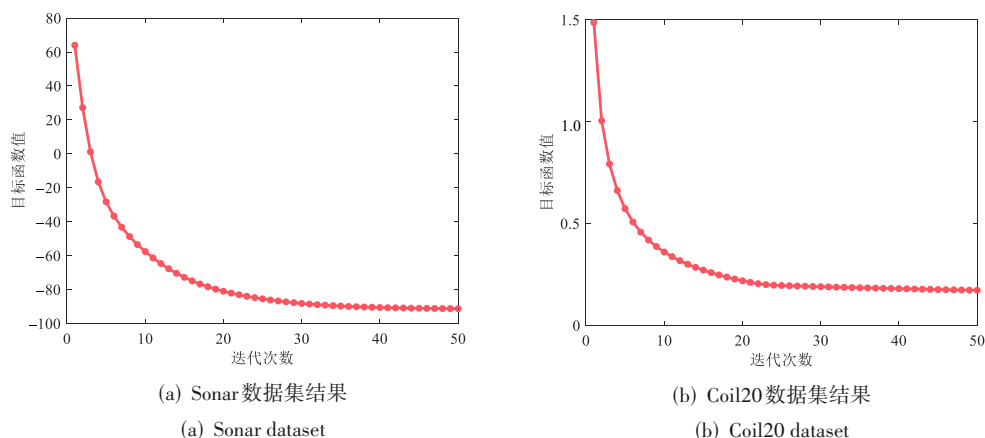


图7 所提方法收敛曲线

Figure 7 Convergence curves of the proposed method

由图7可知,利用算法3求解所提方法,能够很好地收敛。Sonar数据集在40步左右收敛,Coil20数据集在30步左右收敛。实际应用中,可以适当减小各算法迭代次数以防止过拟合同时减少计算时间。

#### 4.7 计算时间分析

本节测量所提方法的实际训练用时。实验在 Con-

trol、German以及Madelon数据集上进行,它们的规模(样本数量 $\times$ 特征数量)分别为 $600 \times 60$ 、 $1\,000 \times 20$ 以及 $2\,600 \times 500$ ,分别代表了小型、中型以及大规模的数据集。表9记录了所提方法中算法1平均每轮迭代用时  $t_1$ 、算法2平均每轮迭代用时  $t_2$  以及算法3排除执行算法1、算法2的平均每轮迭代用时  $t_3$ ,单位均为  $s$ 。

表9 所提方法在所选数据集上的 $t_1, t_2, t_3$ 耗时 单位:sTable 9 Computation time  $t_1, t_2, t_3$  of the proposed method on the selected datasets unit: s

数据集	$t_1$	$t_2$	$t_3$
Control	0.000 132 75	0.003 681 24	0.307 990 87
German	0.000 126 39	0.005 572 50	0.876 952 70
Madelon	0.000 129 41	0.077 917 47	17.560 425 44

观察表9可知,三个数据集的 $t_1$ 基本相同,这是由于 $t_1$ 只与近邻样本数的设置有关,而与样本数量、特征数量无关,与前文的时间复杂度分析结论相对应。 $t_2$ 以及 $t_3$ 随着数据集规模的增大而增大,符合预期,可以发现所提方法训练耗时主要集中在 $t_3$ 。对于大规模的Madelon数据集,算法1、算法2以及算法3迭代次数都设置为几十次的情况下,训练时间大致在几分钟到十几分钟这个范围,在实际运用中能够接受。通过以下途径,可以进一步提高算法的计算效率:(1)可以在满足精度的前提下提前终止优化算法的执行,而非设置固定迭代次数。(2)可以只抽取部分特征空间中的样本与 $\mathbf{Q}$ 构成 $\mathbf{Z}$ 。(3)优化代码,避免重复计算以及尽量减少

循环遍历操作而将其转化为矩阵或向量操作。

#### 4.8 实际场景应用效果

本节将所提方法运用于两个实际场景以评估其效果。对应数据集分别为Mex以及Arcene,均来源于UCI Machine Learning Repository1。Mex数据集包含30个测试对象在进行7种运动时的数据,使用四种传感器进行记录,用于运动识别。本实验中,选取其中两种加速度传感器记录的数据进行实验,即两个模态。首先使用离散余弦变换(Discrete Cosine Transform, DCT)从单个传感器记录的数据中提取特征,以8s为窗口进行滑动提取,总共有X、Y以及Z三个空间方向,对每个方向保留60个DCT系数,所以单个传感器对应的样本特征数为180,同时从单个传感器中可以获取1570个样本。将两个传感器的样本进行特征拼接,以构成多模态数据集,即样本数为1570,特征数为360。Arcene数据集记录200个质谱样本,每个样本拥有10000个特征,为高维稀疏数据集,特征反映了人类血清中具有特定质量值的蛋白质的丰度,用于癌症检测。表10记录了这两个数据集上的实验结果。

表10 各算法在实际场景中的平均准确率±标准差

单位:%

Table 10 Average accuracy ± standard deviation in real-world scenarios

unit: %

数据集	None	TRLDA	RSLDA	RSMLDA	GRLDA	ROBSLDA+	SWULDA	SPFDR	EGCFS	TPELDA
Mex	89.98±1.08	91.74±0.96	93.29±1.07	88.47±1.29	87.24±1.38	93.82±1.51	92.06±1.18	82.48±1.35	<b>94.01±1.06</b>	<b>95.05±0.73</b>
Arcene	83.50±4.04	84.50±2.09	85.17±3.55	64.50±5.93	<u>86.67±4.58</u>	65.00±5.21	84.33±2.85	59.67±5.82	86.50±3.37	<b>87.17±4.31</b>

注:最优结果用粗体显示,次优结果用下划线显示。

由表10可知,所提算法在这两个应用场景中都达到了最优表现。相比于不进行特征提取性能分别有5.07%以及3.67%的提升,相比于第二优秀的算法性能分别有1.04%以及0.50%的提升。此实验结果表明了本文方法在实际复杂场景如多模态数据、高维稀疏数据中也具有良好表现,能够用于实际应用。

## 5 结束语

LDA是一种有效的降维方法,但是其存在未考虑相似样本间关联信息的缺陷。

为解决该问题,本文提出转置投影包络线性判别分析,其在考虑相似样本间关联信息的基础上利用Fisher准则进行投影降维。该方法通过转置投影将原始样本转化为包含了自身信息以及与其相似样本间关联信息的包络样本,基于包络样本进行Fisher准则建模学习降维子空间,并引入分布差异惩罚项解决测试样本转换为包络样本的问题。通过联合优化,所提方法学习到的子空间保证了映射后包络样本的可分度,且能直接运用于原始样本,改善其分布。实验部分采用多个数据集用于方法验证比较,发现其性能相比其他相关算法有一定改善,提升范围在2.25%至13.19%之

间,实验结果表明本文改进思路有效和必要。

尽管本文方法被验证有效,但仍有待进一步改进。首先,本文方法在大规模数据集上的训练速度仍不够快,未来需要设计出具有更快收敛速度的优化算法,以提高训练效率。其次,可以考虑设计出更好的包络样本变换算法,以富集更多相似样本间的关联信息,提高模型性能。

#### 参考文献

- [1] Lu Bin, Wang Fuwang, Chen Junxiang, et al. Improving two-dimensional linear discriminant analysis with L1 norm for optimizing EEG signal[J]. Information Sciences, 2025, 690: 121585.
- [2] 刘中华,周静波,陈毅,等.距离保持投影非线性降维技术的可视化与分类[J].电子学报,2009,37(8):1820-1825. Liu Zhonghua, Zhou Jingbo, Chen Yi, et al. Non-linear dimensionality reduction techniques of distance-preserving projection for visualization and classification[J]. Acta Electronica Sinica, 2009, 37(8): 1820-1825. (in Chinese)
- [3] Yang Xuanhao, Che Hangjun, Leung M F, et al. Self-paced regularized adaptive multi-view unsupervised fea-

- ture selection[J]. *Neural Networks*, 2024, 175: 106295.
- [4] Wang Xiang, Zhu Junxing, Xu Zichen, et al. Local nonlinear dimensionality reduction via preserving the geometric structure of data[J]. *Pattern Recognition*, 2023, 143: 109663.
- [5] 尤殿龙, 郭松, 赵春慧, 等. 面向分类的流特征在线特征选择算法[J]. *电子学报*, 2020, 48(2): 321-332.
- You Dianlong, Guo Song, Zhao Chunhui, et al. Online feature selection with streaming features for classification[J]. *Acta Electronica Sinica*, 2020, 48(2): 321-332. (in Chinese)
- [6] Greenacre M, Groenen P J F, Hastie T, et al. Principal component analysis[J]. *Nature Reviews Methods Primers*, 2022, 2(1): 100.
- [7] Tenenbaum J B, De Silva V, Langford J C. A global geometric framework for nonlinear dimensionality reduction[J]. *Science*, 2000, 290(5500): 2319-2323.
- [8] Zhao Shuping, Zhang B, Yang Jian, et al. Linear discriminant analysis[J]. *Nature Reviews Methods Primers*, 2024, 4(1): 70.
- [9] Dufrenois F, Jbilou K. A support vector machine formulation for linear and kernel discriminant analysis[J]. *Neurocomputing*, 2025, 622: 129273.
- [10] Li Fan, Li Yongming, Shen Yinghua, et al. Deep fuzzy envelope sample generation mechanism for imbalanced ensemble classification[J]. *IEEE Transactions on Fuzzy Systems*, 2024, 32(3): 1248-1262.
- [11] Ma Jie, Cheng Huan, Chen Hong, et al. Envelope rotation forest: A novel ensemble learning method for classification[J]. *Neurocomputing*, 2025, 618: 129059.
- [12] Mahadi M, Ballal T, Moinuddin M, et al. Regularized linear discriminant analysis using a nonlinear covariance matrix estimator[J]. *IEEE Transactions on Signal Processing*, 2024, 72: 1049-1064.
- [13] Chu Delin, Liao Lizhi, Ng M K. Sparse orthogonal linear discriminant analysis[J]. *SIAM Journal on Scientific Computing*, 2012, 34(5): A2421-A2443.
- [14] Wang Jingyu, Wang Lin, Nie Feiping, et al. A novel formulation of trace ratio linear discriminant analysis[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 33(10): 5568-5578.
- [15] Ye Qiaolin, Yang Jie, Zheng Hao, et al. Convergence analysis on trace ratio linear discriminant analysis algorithms[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2025, 36(2): 3878-3881.
- [16] Zhou Yang, Sun Shiliang. Manifold partition discriminant analysis[J]. *IEEE Transactions on Cybernetics*, 2017, 47(4): 830-840.
- [17] Nie Feiping, Wang Zheng, Wang Rong, et al. Adaptive local linear discriminant analysis[J]. *ACM Transactions on Knowledge Discovery from Data*, 2020, 14(1): 9.
- [18] Qu Jinlong, Zhao Xiaowei, Xiao Yun, et al. Adaptive manifold graph representation for two-dimensional discriminant projection[J]. *Knowledge-Based Systems*, 2023, 266: 110411.
- [19] Feng Wenyi, Wang Zhe, Cao Xiqing, et al. Discriminative sparse subspace learning with manifold regularization[J]. *Expert Systems with Applications*, 2024, 249: 123831.
- [20] Zhu Yufei, Lai Zhihui, Gao Can, et al. Generalized robust linear discriminant analysis for jointly sparse learning[J]. *Applied Intelligence*, 2024, 54(19): 9508-9523.
- [21] Zhou Jianhang, Zhang Qi, Zeng Shaoning, et al. Latent linear discriminant analysis for feature extraction via isometric structural learning[J]. *Pattern Recognition*, 2024, 149: 110218.
- [22] Li Zhengxin, Nie Feiping, Wu Danyang, et al. Sparse trace ratio LDA for supervised feature selection[J]. *IEEE Transactions on Cybernetics*, 2024, 54(4): 2420-2433.
- [23] Wen Jie, Fang Xiaozhao, Cui Jinrong, et al. Robust sparse linear discriminant analysis[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2019, 29(2): 390-403.
- [24] Li Shuyi, Zhang Hengmin, Ma Ruijun, et al. Linear discriminant analysis with generalized kernel constraint for robust image classification[J]. *Pattern Recognition*, 2023, 136: 109196.
- [25] Liu Jingjing, Feng Manlong, Xiu Xianchao, et al. Towards robust and sparse linear discriminant analysis for image classification[J]. *Pattern Recognition*, 2024, 153: 110512.
- [26] Yang Xiaojun, Cao Chuanjie, Zhou Keyi, et al. A novel linear discriminant analysis based on alternate ratio sum minimization[J]. *Information Sciences*, 2025, 689: 121444.
- [27] Li Xuelong, Zhang Yunxing, Zhang Rui. Self-weighted unsupervised LDA[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, 34(3): 1627-1632.
- [28] Ikotun A M, Ezugwu A E, Abualigah L, et al. K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data[J]. *Information Sciences*, 2023, 622: 178-210.
- [29] Lam B S Y, Choy S K, Yu C K W. Linear discriminant analysis with trimmed and difference distribution modeling[J]. *Knowledge-Based Systems*, 2024, 299: 112093.
- [30] Frické M H. Data-information-knowledge-wisdom (DIKW) pyramid, framework, continuum[M]//Schintler L A, Mc-

- Neely C L. Encyclopedia of big data. Cham: Springer, 2018: 1-4.
- [31] Gazzaniga M S, Mangun G R, Blakemore S J. The Cognitive Neurosciences[M]. 5th ed. Cambridge: MIT Press, 2014.
- [32] Xia Shuyin, Dai Xiaochuan, Wang Guoyin, et al. An efficient and adaptive granular-ball generation method in classification problem[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(4): 5319-5331.
- [33] Peng Xiaoli, Wang Ping, Xia Shuyin, et al. VPGb: A granular-ball based model for attribute reduction and classification with label noise[J]. Information Sciences, 2022, 611: 504-521.
- [34] 李帆, 张小恒, 李勇明, 等. 基于包络学习和分级结构一致性机制的不平衡集成算法[J]. 电子学报, 2024, 52(3): 751-761.  
Li Fan, Zhang Xiaoheng, Li Yongming, et al. Imbalanced ensemble algorithm based on envelope learning and hierarchical structure consistency mechanism[J]. Acta Electronica Sinica, 2024, 52(3): 751-761. (in Chinese)
- [35] Li Yongming, Liu Chengyu, Wang Pin, et al. Envelope multi-type transformation ensemble algorithm of Parkinson speech samples[J]. Applied Intelligence, 2023, 53(12): 15957-15978.
- [36] Schölkopf B, Smola A, Müller K R. Nonlinear component analysis as a kernel eigenvalue problem[J]. Neural Computation, 1998, 10(5): 1299-1319.
- [37] Gretton A, Borgwardt K M, Rasch M J, et al. A kernel two-sample test[J]. The Journal of Machine Learning Research, 2012, 13: 723-773.
- [38] Wright S J. Coordinate descent algorithms[J]. Mathematical Programming, 2015, 151(1): 3-34.
- [39] Nie Feiping, Yuan Jianjun, Huang Heng. Optimal mean robust principal component analysis[C]//Proceedings of the 31st International Conference on Machine Learning. Cambridge: JMLR.org, 2014: II-1062-II-1070.
- [40] Wang Jingyu, Wang Hongmei, Nie Feiping, et al. Ratio sum versus sum ratio for linear discriminant analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(12): 10171-10185.
- [41] Yi Jihai, Duan Huiyu, Wang Jikui, et al. Structure preserved fast dimensionality reduction[J]. Applied Soft Computing, 2024, 162: 111817.
- [42] Zhang Rui, Zhang Yunxing, Li Xuelong. Unsupervised feature selection via adaptive graph learning and constraint[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(3): 1355-1362.
- [43] Van Der Maaten L, Hinton G. Visualizing data using t-SNE[J]. Journal of Machine Learning Research, 2008, 9(86): 2579-2605.

#### 作者简介



**李勇明** 男, 1976年9月生, 四川绵阳人。重庆大学微电子与通信工程学院教授、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



**张小恒** 男, 1980年10月生, 四川达州人。博士、副教授。主要研究领域为医学信号处理、机器学习、行人轨迹预测。  
E-mail: 7818320@qq.com



**赵文强** 男, 2002年7月生, 四川南充人。重庆大学微电子与通信工程学院硕士研究生。主要研究领域为机器学习。  
E-mail: 2310534870@qq.com



**王品** 女, 1979年11月生, 江苏盐城人。博士、副教授、硕士生导师。主要研究领域为图像处理与识别。  
E-mail: wangpin@cqu.edu.cn



**李帆** 男, 1993年4月生, 湖北汉川人。博士。主要研究领域为非平衡数据处理、机器学习。中国电子学会会员编号: E190197979M。  
E-mail: 979940181@qq.com