局部差异正则化的边界判别投影

何进荣[1),2)] 闭应洲[2)] 丁立新[3)] 刘 斌[1)]

[1)](西北农林科技大学信息工程学院 陕西 杨凌 712100)

[2)](广西师范学院科学计算与智能信息处理广西高校重点实验室 南宁 530023)

[3)](武汉大学计算机学院 武汉 430070)

摘 要 高维是大数据的一个重要特点,数据降维是处理高维数据的有效手段.数据降维算法的设计,关键在于保持原始高维数据集中蕴含的判别信息和几何结构,使得生成的低维特征表示既能刻画原始高维数据的分布形状,又能以更低的计算成本服务于后续的分类任务.边界判别投影算法是一种有监督的线性降维算法,通过最大化不同类别样本点之间的最小距离和最小化同类样本点之间的最大距离,来获取最优判别投影方向.为了保持样本点的几何结构,提高边界判别投影算法的泛化能力,在边界判别投影模型中融入了样本点的局部差异性信息.通过最大化投影之后样本点之间的局部差异来保持数据集的多样性,即在数据降维过程中,局部邻域内相距较远的样本点在投影之后应该保持较远的距离,从而防止在投影过程中原始数据集中蕴含的相似关系和拓扑结构发生扭曲.在图嵌入框架下,数据集的相似信息、判别信息和局部差异信息可以采用正则化的迹差准则进行数据建模.在优化求解时,为了降低散度矩阵特征分解的时间复杂度,通过对数据矩阵进行 QR 分解来加速计算.人脸图像数据集上的分类实验验证了局部差异正则化的边界判别投影算法在判别特征提取方面的有效性.

关键词 数据降维;边界判别投影;数据分类;局部差异;图嵌入

中图法分类号 TP18 **DOI 号** 10.11897/SP.J.1016.2018.00780

Local Variation Regularized Margin Discriminant Projection

HE Jin-Rong[1),2)] BI Ying-Zhou[2)] DING Li-Xin[3)] LIU Bin[1)]

[1)](*College of Information Engineering, Northwest A&F University, Yangling, Shaanxi* 712100)

[2)](*Science Computing and Intelligent Information Processing of Guangxi Higher Education Key Laboratory, Guangxi Teachers Education University, Nanning* 530023)

[3)](*School of Computer, Wuhan University, Wuhan* 430070)

Abstract High dimensionality is one of important properties of big data, and dimensionality reduction is an effective method to deal with high-dimensional data. The key point of dimensionality reduction algorithm design is preserving the discriminant information and geometric structure contained in original high-dimensional data set, such that the obtained low-dimensional feature representations not only can characterize the distributional shape of original high-dimensional data set, but also are helpful to corresponding classification tasks with lower computational costs. Margin discriminant projection is a supervised linear dimensionality reduction algorithm, which seeks for optimal discriminant projection directions by maximizing the minimum distance between samples that belong to different classes, and simultaneously minimizing the maximum distance between samples that belong to the same class. In order to preserve the local geometric structure of original high-dimensional sample points, and improve the generalization ability of margin discriminant projection,

收稿日期:2016-05-25;在线出版日期:2016-12-02.本课题得到西北农林科技大学博士科研启动基金(2452015302)、杨凌示范区科技计划项目(2016NY-31)、科学计算与智能信息处理广西高校重点实验室基金(GXSCIIP201406)资助.**何进荣**,男,1984 年生,博士,讲师,主要研究方向为机器学习与数据挖掘.E-mail:hejinrong@163.com.**闭应洲**(通信作者),男,1967 年生,博士,教授,主要研究领域为智能计算与软件工程.E-mail:byzhou@163.com.**丁立新**(通信作者),男,1967 年生,博士,教授,主要研究领域为智能计算与机器学习.E-mail:lxding@whu.edu.cn.**刘 斌**,男,1981 年生,博士,讲师,主要研究方向为并行计算与机器学习.

local variation information of original high-dimensional samples is encoded in margin discriminant projection model, then the diversities of samples in the original high-dimensional data set can be preserved by maximizing local variations of projected sample points, which means that, in the process of dimensionality reduction, the sample points that are far apart in the local neighborhood should keep a larger distance after the projection, so as to prevent the distortion of the similarity relation and the topological structure contained in the original high-dimensional data set. Under the graph embedding framework, the similarity information, discriminant information and local variation information of original data set can be modeled by the regularized trace difference criterion, which has a closed form solution. By this way, the margin discriminant projection method is extended to the local variation regularized margin discriminant projection. In some real-world applications, such as image data, the dimensionality of data samples is often very large after represented as vectors. In such case, the computational cost of eigen-decomposition on scatter matrix is high. In order to reduce the time complexity of the eigen-decomposition of scatter matrix, as for implementation, an efficient eigen-decomposition algorithm to solve local variation regularized margin discriminant projection optimization problem is derived, in which the QR decomposition technique on data matrix is used to improve the computational efficiency. Since local variation regularized margin discriminant projection only considers margin samples and uses QR decomposition to accelerate calculation, the computational complexity is much lower than original margin discriminant projection. The experimental performance of classification tasks on face image data sets confirm the effectiveness of the proposed local variation regularized margin discriminant projection algorithm on discriminant feature extraction. Compared with margin discriminant projection, local variation regularized margin discriminant projection considers local variation information of data set, and is more flexible, since it is formulated as trace difference regularization framework, which can be adaptive to specific data set. Thus, it can generate low-dimensional features with better discriminant ability. Since local variation can describe intrinsic geometric structure of data set, and the projection directions obtained from regularized margin discriminant projection are orthogonal, it can preserve geometric structure of data set, which leads to wide application and robustness of the proposed method.

Keywords　dimensionality reduction; margin discriminant projection; data classification; local variation; graph embedding

1　引　言

数据降维是高维数据分析的重要手段，其目标是将高维观测数据映射至低维子空间，使得高维数据空间中“相似”的样本点映射至低维子空间中也“相似”. 反之，高维数据空间中“不相似”的样本点映射至低维子空间中也“不相似”. 这种样本间相似性和不相似性关系的保持可以防止高维数据集在降维过程中几何结构和判别结构发生改变，而几何结构和判别结构正是数据挖掘和机器学习中描述数据集蕴含信息的重要内在属性. 数据降维既可以作为一个单独过程进行探索性数据分析，将高维样本点投影至 2 维或 3 维空间来观察数据集的内在分布结构；也可以作为分类等其他机器学习任务的前驱过程，通过样本维数约减，实现信息压缩，提高数据存储、传输和计算的效率[1]，同时去除冗余属性和噪声，生成解释性和判别性更强的数据特征表示[2]. 根据高维数据到低维表示数据映射关系的不同，数据降维算法可分为线性和非线性两类. 由于线性降维算法可直接求得高维数据到低维表示之间的投影矩阵，并直接将新增样本点投影至低维空间，无需重新学习，从而避免了所谓的“外样本问题”[3]，降低了降维过程的计算复杂度；其次，线性降维过程在理论上等价于马氏距离意义下的度量学习[4]，具有简单直观的几何解释；另外线性降维算法是非线性降维的

基础，通过核技巧可扩展至处理非线性情形[5]. 因此，线性降维算法近年来引起许多研究者的关注.

线性数据降维算法可形式化的描述如下：给定含有 n 个 d 维样本点的数据集，并将其表示为数据矩阵 $\boldsymbol{X}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n\}\in R^{d\times n}$，假设将原有样本点维数约减至 r，其中 $r<d$，并称 r 为目标维数. 线性降维算法通过构造优化准则函数 $J_{\boldsymbol{X}}(\cdot)$ 来得到投影矩阵 $\boldsymbol{V}=[\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_r]\in R^{d\times r}$，并记 $\boldsymbol{Y}=\boldsymbol{V}^{\mathrm{T}}\boldsymbol{X}$，此处 $\boldsymbol{Y}=\{\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_n\}\in R^{r\times n}$ 称为高维数据矩阵 $\boldsymbol{X}$ 对应的低维表示矩阵. 如何构造优化准则函数 $J_{\boldsymbol{X}}(\cdot)$ 是线性降维算法研究的核心问题. 在数据建模时，优化准则函数 $J_{\boldsymbol{X}}(\cdot)$ 的构造依赖于数据分析的目标任务，作为预处理过程的数据降维，生成的低维表示要服务于后续数据分析任务，因此在构造优化准则函数时，应该采用逆向思维. 首先需要确定对数据集哪一方面特性感兴趣，然后构造相应的优化准则函数进行信息提取，在信息提取的过程中，尽可能保持高维数据集固有的内在结构特性不变. 例如，主成分分析(Principle Component Analysis，PCA)通过方差最大化来保持数据集的全局几何结构[6]，线性判别分析(Linear Discriminant Analysis，LDA)通过类内散度极小化和类间散度极大化来保持不同类别数据的可分性[7]，局部保持投影(Locality Preserving Projection，LPP)通过极小化局部邻域内样本点之间的距离来保持数据的局部邻近关系[8]，多维尺度扩展(Multi-dimensional Scaling，MDS)通过最小化重构误差保持数据之间的欧氏距离[9]，MFA 通过局部类内散度极小化和局部类间散度极大化来保持不同类别数据的局部可分性[10]，低秩相似保持投影(Low-rank Similarity Preserving Projections，LSPP)算法通过保持正交投影过程中数据集中显著性特征的相似性和局部性来实现无监督的特征学习[11]. 通常，降维算法中优化准则函数 $J_{\boldsymbol{X}}(\cdot)$ 的构造是受数据分析的目标任务驱动的. 对于数据分类任务，需要保持高维数据集的判别特性；对于数据可视化任务，通常对应于数据流形展开，将每个样本点看作是图上的结点，样本点之间的距离作为结点之间边上的权重，通过最短路径来近似计算数据流形的测地距离[12]，在图嵌入过程中保持测地距离不变，即可实现数据流形展开. 如果在构造优化准则函数 $J_{\boldsymbol{X}}(\cdot)$ 时，考虑了数据集中的样本标记信息，或者样本点之间的相似性关系，这类方法称为有监督算法，否则为无监督算法. 假设第 i 个样本点 x_i 对应的类别标记为 $l(x_i)\in\boldsymbol{C}$，这里 $\boldsymbol{C}=\{c_1,c_2,\cdots,c_N\}$ 表示样本点类别标记的集合.

在有监督线性降维算法设计中，边界是度量不同类别样本点可分性的常用方法，被广泛应用于机器学习和数据挖掘算法设计中. 不同类别样本点之间的边界的定义有多种. 例如，在支撑向量机(Support Vector Machine，SVM)[13]方法中，两类样本点之间的边界被定义所有样本点到分类决策超平面的距离的最小值. 在最大边界准则(Maximum Margin Criterion，MMC)[14]中，边界被定义为两类样本点的均值向量之间的距离减去这两类样本点的类内散度. 由于不同类别的样本点可看作是分布在不同的流形上，为了描述具有流形结构的不同类样本点之间的边界，流形划分判别分析(Manifold Partition Discriminant Analysis，MPDA)[15]同时考虑了所有样本点对之间的差异性和分片区域的一致性，首先将数据流形划分为一系列不重叠的线性子空间，然后针对划分之后的不同数据流形构造切空间，最后采用一阶泰勒展开生成数据流形的低维分段线性表示. 正则化的稀疏保持半监督降维(Regularized Sparsity Preserving Semi-Supervised Dimension Reduction，Reg-S3DR)[16]通过构建 k 连通图实现流形划分，不同流形之间通过 k 最近邻样本点相连，然后在此连通图上计算所有样本点之间的测地距离，最后对每个样本点的图距离向量进行低维嵌入. 通过流形划分和测地距离嵌入的降维方法可以在判别特征提取的同时，有效挖掘高维数据集中蕴含的非线性流形结构，然而当数据集规模较大时，这些方法计算复杂度较高. 在边界判别投影(Margin Discriminant Projection，MDP)[17]中，首先定义了同类边界样本点和异类边界样本点，然后将边界定义为所有异类边界样本点之间的距离之和减去所有同类边界样本点之间的距离之和. 与其他基于边界思想的数据降维算法相比，MDP 算法中定义的边界概念具有直观的几何解释，且易于计算. 由于 MDP 算法在构造优化准则函数时，只考虑了边界样本点，因此当训练集中样本类别个数较少时，在构建样本点的权重矩阵时，只有边界样本点之间的权重非零，而其他所有样本点之间的权重均为 0，因此权重矩阵是一个稀疏矩阵，这为数值处理提供了快速计算的便利. 此外，MDP 算法对数据集的分布没有特定假设，具有更广泛的数据适用性. 而 LDA 算法通常假设数据集中每类样本点是高斯分布的，这些假设在实际应用中不一定满足. 然而，MDP 算法会扭曲数据的局部真实几何结构. 在图嵌入框架[10]下，

MDP 算法以异类样本之间的距离最大化和同类样本之间的距离最小化为优化目标，此时可能导致同类样本点被投影至同一点上从而聚集在一起，即原始样本点之间的局部拓扑关系发生了扭曲. 因此 MDP 算法优化准则函数构造时只考虑了原始高维数据集的判别结构，而忽略了几何拓扑结构.

为了克服这一不足之处，借鉴稳定正交的局部判别嵌入(Stable Orthogonal Local Discriminant Embedding，SOLDE)算法[18]的思想，本文在前期工作 MDP 算法的基础上，提出了正则化边界判别投影(Regularized Margin Discriminant Projection，RMDP)算法. RMDP 在 MDP 的基础上，考虑了数据集的局部差异性，使得在投影过程中，属于同一类别的样本点之间具有一定程度的差异性，并通过最大化局部差异性来保持数据的多样性，从而最大限度保持样本点类内几何结构，提高降维算法的泛化性能. 不同于 SOLDE 算法中的迹比建模方法，RMDP 算法考虑了类间可分性、类内相似性和局部差异性这 3 个目标，并将其归结为迹差优化问题. 与 SOLDE 算法相比，RMDP 算法可直接用于小样本问题，且求解效率更高，在人脸图像数据分类实验中具有更优越的性能.

2　MDP 算法简介

2.1　边界的定义

边界描述判别特征提取算法的常用概念，与 LDA 和 MMC 中类内散度与类间散度的定义不同，MDP 采用边界样本点来定义类内散度和类间散度. 边界样本点是描述高维数据集判别结构的有力工具，本节首先介绍 RMDP 算法中边界的定义.

定义 1.　给定样本点 $\boldsymbol{x}_i$ 与 $\boldsymbol{x}_j$，它们之间的距离记作 $d(\boldsymbol{x}_i,\boldsymbol{x}_j)$，并定义如下：

$$d(\boldsymbol{x}_i,\boldsymbol{x}_j)=\|\boldsymbol{x}_i-\boldsymbol{x}_j\|_2.$$

定义 2.　第 i 类样本点集合 C_i 和第 j 类样本点集合 C_j 中距离最近的样本点称为异类边界样本点，记做 $\boldsymbol{x}_j^i$ 与 $\boldsymbol{x}_i^j$，即：

$$\{\boldsymbol{x}_j^i\in c_i,\boldsymbol{x}_i^j\in C_j:d(\boldsymbol{x}_j^i,\boldsymbol{x}_i^j)\le d(\boldsymbol{x}_i,\boldsymbol{x}_j),\ \forall\boldsymbol{x}_i\in C_i,\forall\boldsymbol{x}_j\in C_j\}.$$

定义 3.　假设 C_i 表示第 i 类样本集合，则 C_i 的同类边界样本点定义为 C_i 中距离最远的样本点，记做 $\boldsymbol{x}_a^i$ 与 $\boldsymbol{x}_b^i$，即：

$$\{\boldsymbol{x}_a^i,\boldsymbol{x}_b^i\in C_i:d(\boldsymbol{x}_a^i,\boldsymbol{x}_b^i)\ge d(\boldsymbol{x}_i,\boldsymbol{x}_j),\forall\boldsymbol{x}_i,\boldsymbol{x}_j\in C_i\}.$$

异类边界样本点和同类边界样本点统称为边界样本点.

定义 4.　第 i 类样本点集合 C_i 和第 j 类的样本点集合 C_j 之间的类间距离 $d(C_i,C_j)$ 定义为

$$d(C_i,C_j)=d(\boldsymbol{x}_j^i,\boldsymbol{x}_i^j).$$

定义 5.　第 i 类样本点集合 C_i 中样本点的类内距离 $d(C_i)$ 定义为

$$d(C_i)=d(\boldsymbol{x}_a^i,\boldsymbol{x}_b^i).$$

定义 6.　假设训练样本集 $\boldsymbol{X}$ 共有 N 类样本点，即 $\boldsymbol{X}=\{C_i\}_{i=1}^N$. 训练样本集的边界 J 定义为

$$J=\sum_{i\ne j}d(C_i,C_j)-\sum_{i=1}^N d(C_i).$$

2.2　MDP 算法原理

如图 1 所示，MDP 算法以最大化训练数据集的边界为优化准则函数，从而使得生成的低维表示具有更强的判别能力. 图 1 中不同几何形状的结点表示属于不同类别的样本点，投影之后的低维样本点采用虚线结点表示. MDP 算法的目标是寻找最优投影方向，使得该方向上投影之后的低维样本点之间的边界尽可能的大. 由定义 6 可知，最大化训练样本集的边界对应于极大化异类边界样本点之间的距离，同时极小化同类边界样本点之间的距离，从而导致训练样本集中类内紧致性和类间可分性得到了进一步强化，这为下一步的数据分类任务提供了更加易于区分的低维表示.

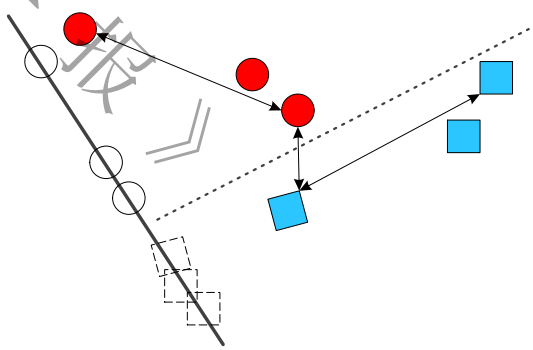


图 1　MDP 算法的基本思想

假设 $\boldsymbol{V}$ 表示投影矩阵，则 $\boldsymbol{y}_i=\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}_i$ 是高维样本点 $\boldsymbol{x}_i$ 经过投影之后的低维表示样本，于是任意两个低维表示样本点之间的距离记为

$$\delta(\boldsymbol{y}_i,\boldsymbol{y}_j)=\|\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}_j\|_2.$$

假设投影之后低维表示样本点的类间距离和类内距离分别记作 $\delta(c_i,c_j)$ 和 $\delta(c_i)$，则由定义 4 和定义 5 可知：

$$\delta(c_i,c_j)=\|\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}_j^i-\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}_i^j\|_2,$$
$$\delta(c_i)=\|\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}_a^i-\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}_b^i\|_2.$$

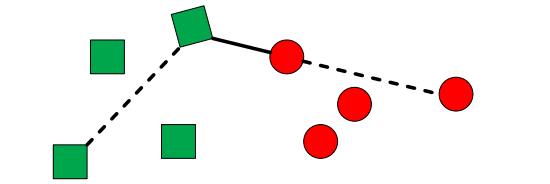
于是，MDP 算法的优化目标可以表示如下：

$$\max_{\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}=I}\sum_{i\neq j}\delta(c_i,c_j) \tag{1}$$

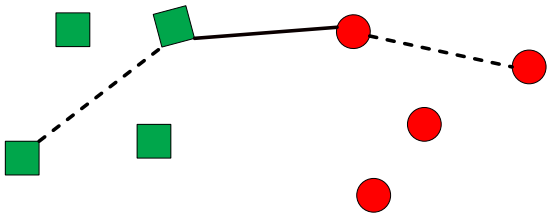
$$\min_{\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}=I}\sum_{i=1}^{C}\delta(c_i) \tag{2}$$

3 RMDP 算法的数学模型

在数据降维算法的优化准则函数构造中，保持原始高维数据集的局部差异性对于描述数据集所蕴含的局部流形结构和全局几何结构具有重要意义[19]. 同时，样本点之间的差异性在一定程度上也反映了数据集的判别结构，局部邻域内的样本点可能属于不同的类别，也可能属于同一个类别中的不同子类. 为了描述数据集的这种特性，在优化准则函数构造时，RMDP 算法同时考虑了样本之间的相似性、判别性和差异性，并采用图嵌入准则进行数据建模，其基本思想如图 2 所示. 图 2 中共有 8 个样本点，每类 4 个样本点，经过投影变换之后，我们希望实线表示的类间距离增大，虚线表示的类内距离减小的同时，增大这 8 个样本点之间的距离. 对比图 1 和图 2 可知，RMDP 算法可以实现高维数据集几何结构和判别结构的有效挖掘.



(a) 变换前



(b) 变换后

图 2 RMDP 算法的基本思想

3.1 相似图嵌入模型

与原始的 MDP 算法类似，RMDP 算法的判别图定义为 $G^{(S)}=\{\boldsymbol{X},\boldsymbol{W}^{(S)}\}$，主要用于描述数据集中不同类别样本点之间的可分性，其中判别图的权重矩阵 $\boldsymbol{W}^{(S)}$ 定义为

$$\boldsymbol{W}_{ij}^{(S)}=\begin{cases}1, & \text{样本点为 } \boldsymbol{x}_a^c,\boldsymbol{x}_b^c\\ 0, & \text{其他}\end{cases} \tag{3}$$

这里 $\boldsymbol{x}_a^c$ 和 $\boldsymbol{x}_b^c$ 分别表示第 c 类样本点集合中的同类边界样本点. 为了保持原始高维数据集样本之间的相似性，在数据降维过程中，我们希望属于同一类别的高维样本点在投影之后，它们之间的欧式距离尽可能地小. 即

$$\min\sum_{i,j}\|\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}_j\|^2\boldsymbol{W}_{ij}^{(S)} \tag{4}$$

目标函数式(4)与 MDP 算法中的目标函数式(2)等价，该二次优化问题又可改写为

$$\min tr(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{L}^{(S)}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}) \tag{5}$$

其中 $\boldsymbol{L}^{(S)}=\boldsymbol{D}^{(S)}-\boldsymbol{W}^{(S)}$ 是相似图的 Laplacian 矩阵，且 $\boldsymbol{D}^{(S)}$ 是对角元素为 $\boldsymbol{D}_{ii}^{(S)}=\sum_{j=1}^{n}\boldsymbol{W}_{ij}^{(S)}$ 的对角阵.

3.2 判别图嵌入模型

判别图定义为 $\boldsymbol{G}^{(D)}=\{\boldsymbol{X},\boldsymbol{W}^{(D)}\}$，主要用于描述属于不同类别的样本点之间的邻接关系，其中权重矩阵 $\boldsymbol{W}^{(D)}$ 定义如下：

$$\boldsymbol{W}_{ij}^{(D)}=\begin{cases}1, & \text{样本点为 } \boldsymbol{x}_j^i,\boldsymbol{x}_i^j\\ 0, & \text{其他}\end{cases} \tag{6}$$

这里 $\boldsymbol{x}_j^i$ 和 $\boldsymbol{x}_i^j$ 是数据集中的异类边界样本点，并采用异类边界样本点之间的距离来描述低维空间中样本点之间的判别性，于是建立如下形式的判别图嵌入模型：

$$\max\sum_{i,j}\|\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}_j\|^2\boldsymbol{W}_{ij}^{(D)} \tag{7}$$

目标函数式(7)与 MDP 算法中的目标函数式(1)等价. 类似的，该问题可以改写为

$$\max tr(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{L}^{(D)}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}) \tag{8}$$

这里 $\boldsymbol{L}^{(D)}=\boldsymbol{D}^{(D)}-\boldsymbol{W}^{(D)}$ 表示判别图的 Laplacian 矩阵，$\boldsymbol{D}^{(D)}$ 是对角元素为 $\boldsymbol{D}_{ii}^{(D)}=\sum_{j=1}^{n}\boldsymbol{W}_{ij}^{(D)}$ 的对角阵.

3.3 差异图嵌入模型

差异图定义为 $\boldsymbol{G}^{(LV)}=\{\boldsymbol{X},\boldsymbol{W}^{(LV)}\}$，用于描述局部邻域内样本点的分散程度，距离较大的样本点之间具有较大的差异性，于是可定义如下的权重矩阵 $\boldsymbol{W}^{(LV)}$[18]：

$$\boldsymbol{W}_{ij}^{(LV)}=\begin{cases}\exp\left(-\dfrac{t}{\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2}\right), & \boldsymbol{x}_i\in N_k(\boldsymbol{x}_j)\vee\\ & \boldsymbol{x}_j\in N_k(\boldsymbol{x}_i)\\ 0, & \text{其他}\end{cases} \tag{8}$$

这里 $N_k(\boldsymbol{x}_i)$ 表示由距离样本点 $\boldsymbol{x}_i$ 最近的 k 个样本点构成的集合. 在降维过程中引入差异图主要是为了防止局部邻域内的高维样本点被投影至同一低维表示，从而保持原始数据集中蕴含的几何结构和拓扑结构，于是差异图嵌入模型可定义为

$$\max\sum_{i,j}\|\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}_j\|^2\boldsymbol{W}_{ij}^{(LV)} \tag{9}$$

类似的，写成矩阵形式为

$$\max tr(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{L}^{(LV)}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}) \tag{10}$$

这里 $\boldsymbol{L}^{(LV)}=\boldsymbol{D}^{(LV)}-\boldsymbol{W}^{(LV)}$ 表示差异图的 Laplacian 矩阵，$\boldsymbol{D}^{(LV)}$ 是对角元素为 $\boldsymbol{D}_{ii}^{(LV)}=\sum_{j=1}^{n}\boldsymbol{W}_{ij}^{(LV)}$ 的对角阵.

由式(9)可知，差异图嵌入模型具有以下性质. 首先，差异图嵌入模型可以保持局部邻域内样本点之间的不相似性，即保持局部邻域内样本点之间的差异性尽可能不变，距离越大的样本点之间具有更大的差异性保持要求，这样的差异性与样本点之间的类别标记无关，它描述的是数据集自身固有属性，在多模态数据分类[20]中表现为样本的多样性或者子类结构. 另外注意到，当同类边界样本点是局部近邻样本点中相距较远的点时，差异图嵌入模型与相似图嵌入模型存在矛盾，此时通过数据集局部差异最大化可以防止相似性最小化过程中破坏数据集的几何拓扑结构；其次，由式(8)可知，差异图嵌入模型中的权重函数是关于样本点之间欧氏距离的单调递增函数，即局部邻域内距离较远的样本点之间具有较大的差异图权重，在差异图嵌入过程中，它们对应的低维表示也应该具有较远的距离. 如图 3 所示，差异图权重函数和高斯权重函数在图像形态上具有相似性，而单调性正好相反. 随着样本点之间距离的增大，差异图权重函数先是缓慢上升，然后快速上升，最后上升速度又开始下降，逐渐趋于平衡. 这与差异图嵌入模型的目标是吻合的，随着样本点之间距离的减小，对应的局部差异性权重趋近于 0，反之，距离较远的样本点(可能是离群点)对应的差异图权重也较大，因此在极大化差异图嵌入模型式(9)的过程中将会被施加更大的惩罚，从而导致对应的低维表示也相距较远，这在一定程度上说明了 RMDP 模型对数据集中的离群点具有稳健性；另外，虽然判别图嵌入模型和相似图嵌入模型在数据降维过程中都具有一定的判别能力，但是它们不能保持数据集的几何拓扑结构. 然而，在差异图嵌入模型中，局部邻域内的样本点有可能属于相同的类别，也有可能属于不同的类别，此时通过将局部邻域内的样本点投影至较远的距离，且距离越大的样本点投影之后的距离也更远，以此来避免数据降维过程中原始数据集的几何结构和拓扑关系被破坏，同时进一步挖掘数据集的判别结构，防止降维算法出现过拟合.

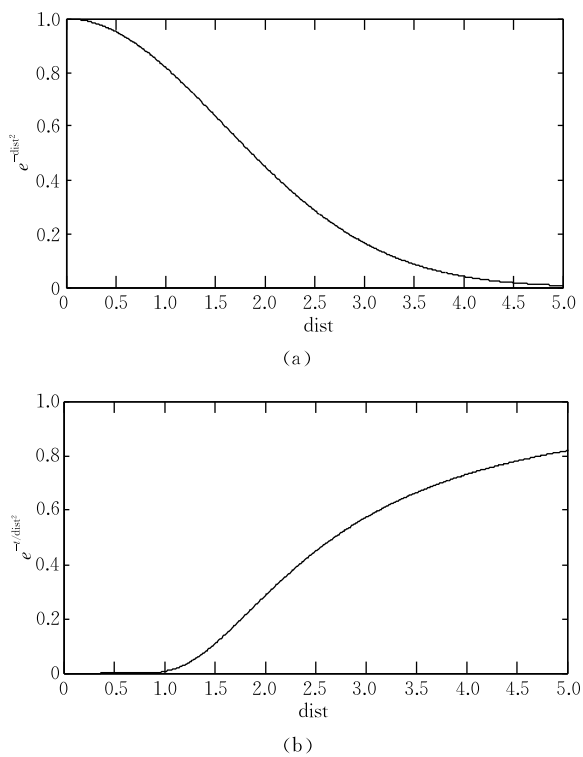


图 3　高斯权重函数(a)和差异图权重函数(b)

3.4　RMDP 模型

将目标函数式(5)、式(8)和式(10)融合在一起，可得 RMDP 算法的数学模型：

$$\max_{\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}=\boldsymbol{I}} tr(\boldsymbol{V}^{\mathrm{T}}(\beta\boldsymbol{X}\boldsymbol{L}^{(P)}\boldsymbol{X}^{\mathrm{T}}-(1-\beta)\boldsymbol{X}\boldsymbol{L}^{(S)}\boldsymbol{X}^{\mathrm{T}})\boldsymbol{V}) \tag{11}$$

其中 $\boldsymbol{L}^{(P)}=\alpha\boldsymbol{L}^{(LV)}+(1-\alpha)\boldsymbol{L}^{(D)}$. 为了对式(11)进行简化，可令 $\boldsymbol{L}=\beta\boldsymbol{L}^{(P)}-(1-\beta)\boldsymbol{L}^{(S)}$，于是式(11)可简化为

$$\max_{\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}=\boldsymbol{I}} tr(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{L}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}) \tag{12}$$

RMDP 模型(12)中含有两个正则化参数 α 和 β，其中 α 可看作是类间判别信息和局部差异信息之间的折中，而 β 可看作是类内相似性(对应于极小化目标函数)与极大化目标函数之间的折中. α 越大，表明局部差异性信息在降维中起的作用越大；而 β 越小，说明类内相似性在降维中起的作用越大. RMDP 模型(12)是一个典型的二次优化问题，该问题的求解等价于对散度矩阵 $\boldsymbol{X}\boldsymbol{L}\boldsymbol{X}^{\mathrm{T}}$ 进行特征分解. 当 $\alpha=0$ 时，RMDP 模型(12)退化为 MDP 模型；当 $\beta=0$ 时，RMDP 模型(12)退化为 LPP 模型.

4　RMDP 算法的求解步骤

来源于实际应用领域的观测数据在向量化之后通常具有较高的维数，例如基因表达数据[21]、图像视频数据和网页文本数据[22]等，其维数 d 通常大于 10^4. 此时散度矩阵 $\boldsymbol{X}\boldsymbol{L}\boldsymbol{X}^{\mathrm{T}}$ 的大小为 $d\times d$，对其进行特征分解的计算复杂度与维数 d 呈立方关系，数值计算中具有较高的内存和时间开销. 为了降低计算复杂度，RMDP 采用 QR 分解进行加速计算.

定理 1. 令 $\boldsymbol{X}=\boldsymbol{QR}$ 表示数据矩阵 $\boldsymbol{X}$ 的 QR 分解，这里 $\boldsymbol{Q}\in R^{d\times m}$ 是列正交矩阵，即 $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}=\boldsymbol{I}$，$\boldsymbol{R}\in R^{m\times n}$ 是上三角矩阵，$m=\mathrm{rank}(\boldsymbol{X})$ 表示数据矩阵 $\boldsymbol{X}$ 的秩，显然 $m<d$. 假设构造矩阵 $\boldsymbol{RLR}^{\mathrm{T}}\in R^{m\times m}$，对其特征分解可得 $\boldsymbol{RLR}^{\mathrm{T}}\boldsymbol{U}=\boldsymbol{U}\boldsymbol{\Lambda}$，这里 $\boldsymbol{\Lambda}$ 是对角阵，其对角线上元素为矩阵 $\boldsymbol{RLR}^{\mathrm{T}}$ 的特征值，$\boldsymbol{U}$ 是正交矩阵，其列向量为特征值 $\boldsymbol{\Lambda}_{ii}$ 对应的特征向量. 则 $\boldsymbol{\Lambda}$ 和 $\boldsymbol{QU}$ 构成了散度矩阵 $\boldsymbol{XLX}^{\mathrm{T}}$ 的特征分解，此时 $\boldsymbol{V}=\boldsymbol{QU}$ 是投影矩阵.

证明. 根据 $\boldsymbol{RLR}^{\mathrm{T}}\boldsymbol{U}=\boldsymbol{U}\boldsymbol{\Lambda}$ 和 $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}=\boldsymbol{I}$，可得

$$\boldsymbol{RLR}^{\mathrm{T}}(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q})\boldsymbol{U}=\boldsymbol{U}\boldsymbol{\Lambda} \tag{13}$$

式(13)两边同时左乘 $\boldsymbol{Q}$，得

$$\boldsymbol{QRLR}^{\mathrm{T}}(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q})\boldsymbol{U}=\boldsymbol{QU}\boldsymbol{\Lambda} \tag{14}$$

式(14)又可以改写为

$$(\boldsymbol{QR})\boldsymbol{L}(\boldsymbol{QR})^{\mathrm{T}}(\boldsymbol{QU})=(\boldsymbol{QU})\boldsymbol{\Lambda} \tag{15}$$

替换式(15)中的 $\boldsymbol{X}=\boldsymbol{QR}$，得

$$\boldsymbol{XLX}^{\mathrm{T}}(\boldsymbol{QU})=(\boldsymbol{QU})\boldsymbol{\Lambda}$$

从上式呈现的结构中可以看出，令 $\boldsymbol{V}=\boldsymbol{QU}$，则 $\boldsymbol{V}$ 是散度矩阵 $\boldsymbol{XLX}^{\mathrm{T}}$ 的特征向量构成的矩阵，而 $\boldsymbol{\Lambda}$ 是对应的特征值矩阵.

由于 $\boldsymbol{RLR}^{\mathrm{T}}$ 是对称矩阵，因此 $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}=\boldsymbol{I}$，于是

$$\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}=(\boldsymbol{QU})^{\mathrm{T}}\boldsymbol{QU}=\boldsymbol{U}^{\mathrm{T}}(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q})\boldsymbol{U}=\boldsymbol{I}$$ 证毕.

又因为 $m<d$，所以 $\boldsymbol{RLR}^{\mathrm{T}}\in R^{m\times m}$ 的规模将远小于 $\boldsymbol{XLX}^{\mathrm{T}}\in R^{d\times d}$，对 $\boldsymbol{RLR}^{\mathrm{T}}$ 进行特征分解的计算复杂度将远低于对 $\boldsymbol{XLX}^{\mathrm{T}}$ 进行特征分解，于是 RMDP 算法具有较高的计算效率. 综上所述，RMDP 算法计算步骤总结如下.

算法 1. RMDP 算法.

输入：由训练样本构成的数据矩阵 $\boldsymbol{X}$，标签集合 C 和预先设置的目标维数 r

输出：投影矩阵 $\boldsymbol{V}$ 和低维表示矩阵 $\boldsymbol{Y}$

1. 根据由训练样本所构成的数据矩阵 $\boldsymbol{X}$，分别计算相似图权重矩阵 $\boldsymbol{W}^{(S)}$、判别图权重矩阵 $\boldsymbol{W}^{(D)}$ 和差异图权重矩阵 $\boldsymbol{W}^{(LV)}$.

2. 根据式(12)，计算 RMDP 模型目标函数中的 Laplacian 矩阵 $\boldsymbol{L}$.

3. 对数据矩阵 $\boldsymbol{X}$ 进行 QR 分解，并将分解结果记作 $\boldsymbol{X}=\boldsymbol{QR}$.

4. 根据步 2 结果计算 $\boldsymbol{RLR}^{\mathrm{T}}$，并对其进行特征分解，并将前 r 个最大特征值所对应的特征向量记作 $\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_r$，将它们作为列向量构成的矩阵记作 $\boldsymbol{U}=[\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_r]$.

5. 计算最优投影矩阵 $\boldsymbol{V}=\boldsymbol{QU}$.

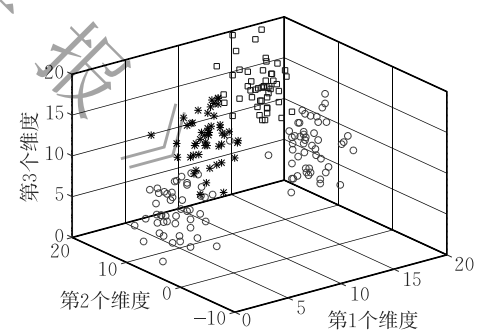
6. 计算训练样本集对应的低维表示矩阵：$\boldsymbol{Y}=\boldsymbol{V}^{\mathrm{T}}\boldsymbol{X}$.

5 实验结果与分析

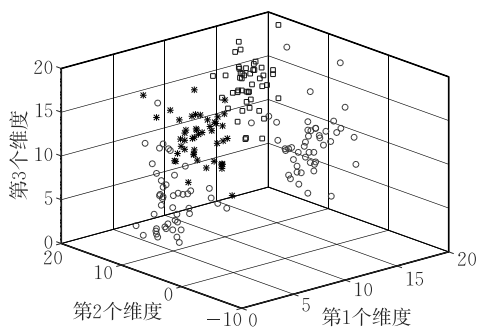
为了验证 RMDP 算法的有效性，实验环节采用 PCA、LDA、LPP、MFA、MDP、SOLDE 和 MPDA 算法进行比较.

5.1 人工数据集上的 2 维嵌入结果

实验中采用的人工数据集如图 4 所示，分别命名为 3D3Cluster 和 3D3ClusterOutlier. 其中，3D3Cluster 是 3 维空间中随机生成的 200 个样本点，共有 3 类样本点，且同类样本点中存在子类；3D3ClusterOutlier 通过在 3D3Cluster 数据集中添加离群点得到，即随机选取 10%的样本点，将其各个分量的属性值添加较大的扰动值. 为了比较不同降维算法的特点，实验中将原始 3 维样本点投影至的 2 维平面，嵌入结果如图 6 和图 7 所示，其中正则化参数 α 和 β 均设置为 0.01 和 0.99. 图 5 讨论了 3D3Cluster 人工数据集上 RMDP 算法中正则化参数选择问题，α 取值越小，β 取值越大，越有利于保持数据集的几何结构和判别结构，此时对应的 RMDP 模型(11)中差异图嵌入模型(10)和相似图嵌入模型(5)权重较小，而判别图嵌入模型(8)权重较大. 最优情形下的 α 取值范围为[0，0.1]，β 取值范围为[0.9，1]，且二维嵌入结果对 β 的取值比较敏感.



(a) 3D3Cluster人工数据集



(b) 3D3ClusterOutlier人工数据集

图 4 3 维人工数据集散点图示例

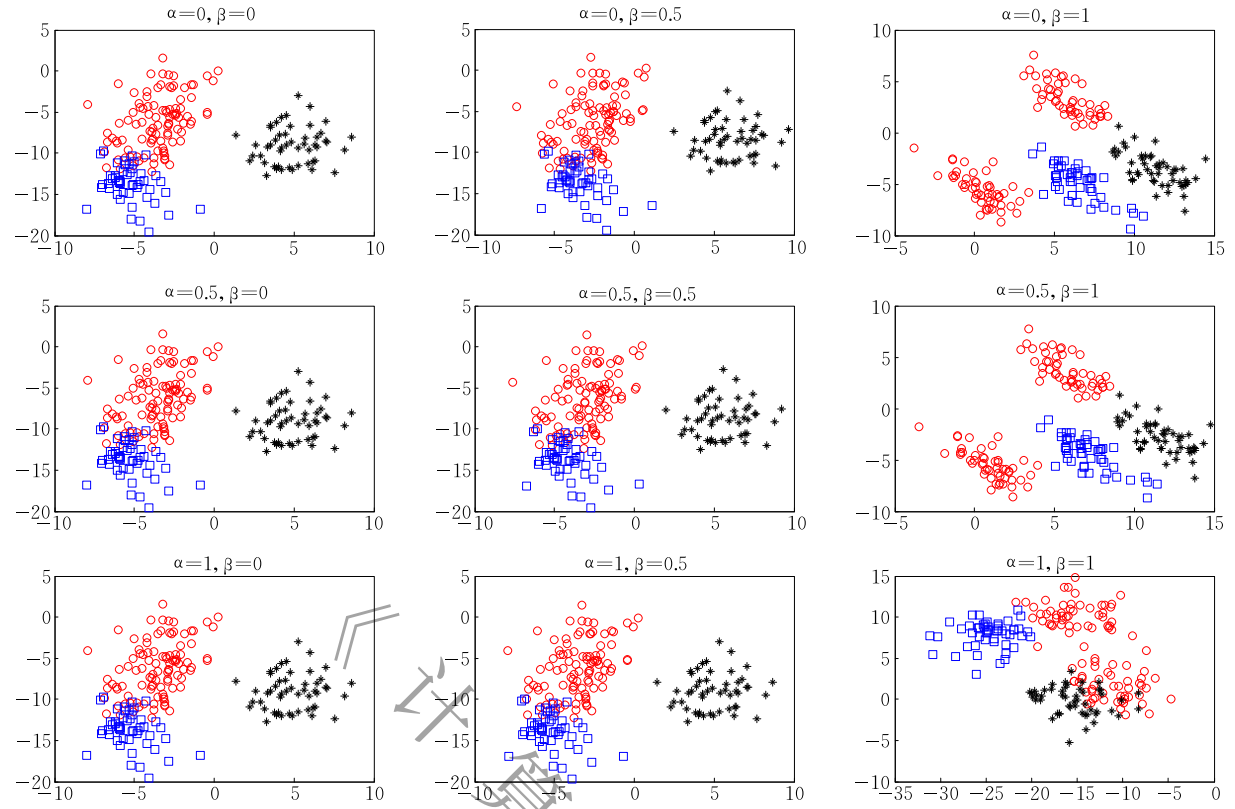


图 5　3D3Cluster 人工数据集上不同正则化参数设置下的 RMDP 嵌入结果

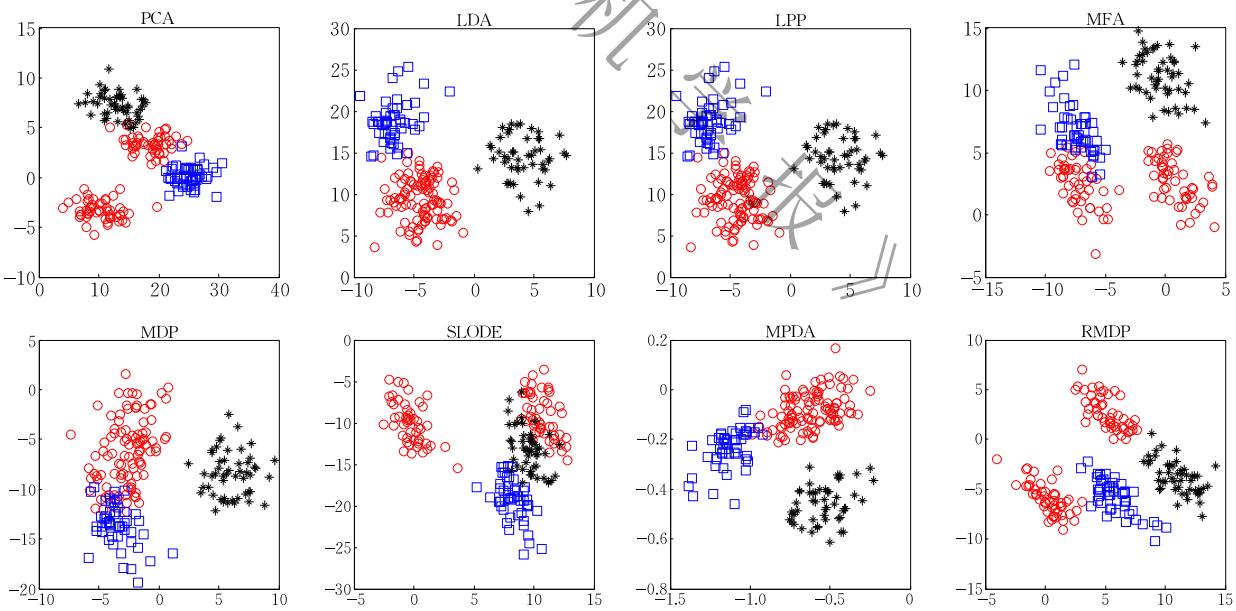


图 6　3D3Cluster 人工数据集上的 2 维嵌入结果比较

在数据降维过程中，我们希望在保持原始样本点几何结构的同时，最大限度挖掘出数据集的判别结构. 由图 6 可知，LDA、LPP、MDP 和 MPDA 等算法不能保持同类样本点中蕴含的子类结构，即样本点的几何结构被破坏；虽然 PCA、MFA、SLODE 和 RMDP 均保持了样本点中的子类结构，但是 RMDP 算法得到的嵌入结果中，不同类别样本点之间的可分性是最好的. 同时结合图 7 可知，当数据集中含有离群点或噪声时，MFA、SOLDE 和 RMDP 仍然能够得到较好的嵌入结果. 由于 RMDP 在数据建模中引入了局部差异最大化准则，在数据降维过程中，不仅可以较好的保持原始数据集的几何结构，而且使得降维算法具有一定的鲁棒性.

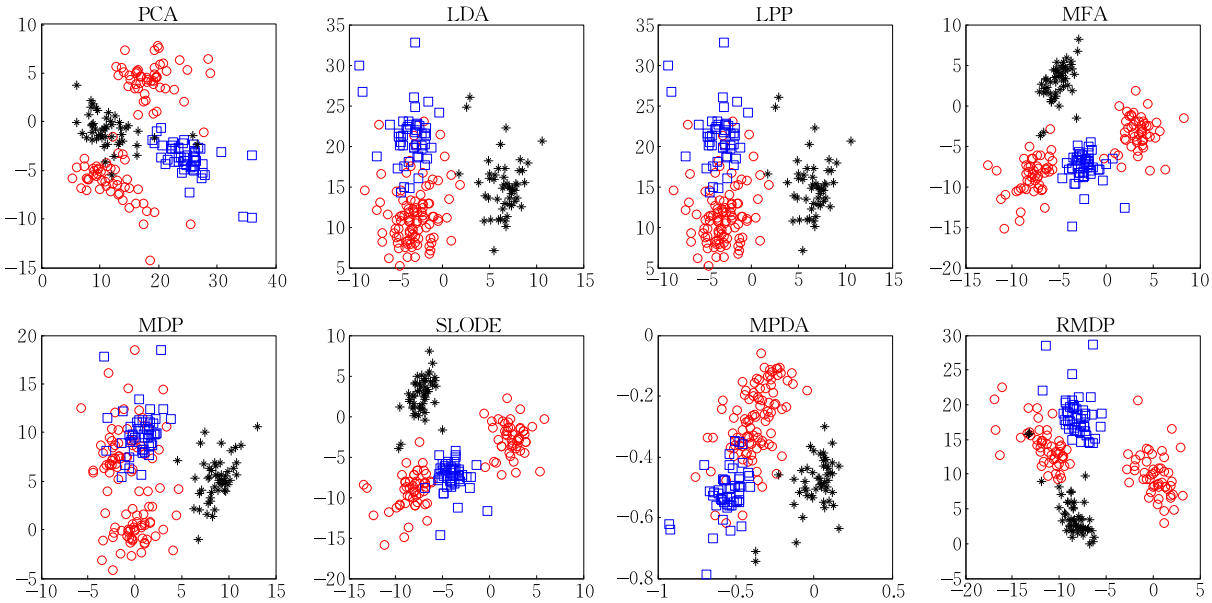


图 7 3D3ClusterOutlier 人工数据集上的 2 维嵌入结果比较

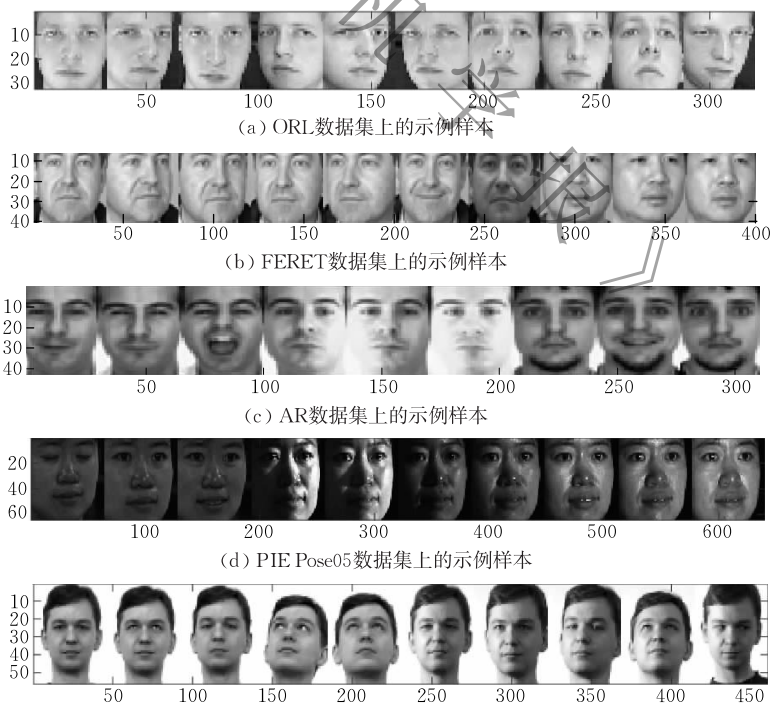
5.2 人脸数据集上的分类实验

5.2.1 实验设置

实验中选用ORL①、FERET②、AR③、PIE Pose05④和 Altkom⑤ 数据集，数据集说明如表 1 所示，各个数据集中的部分样本图像如图 8 所示.

表 1 数据集描述

数据集	d	n	C
ORL	32×32	400	40
FERET	40×40	1400	200
AR	42×30	1400	100
PIE Pose05	64×64	3332	68
Altkom	56×46	1200	80



(a) ORL数据集上的示例样本

(b) FERET数据集上的示例样本

(c) AR数据集上的示例样本

(d) PIE Pose05数据集上的示例样本

(e) Altkom数据集上的示例样本

图 8 实验中选用的数据集示例图像

本文采用的实验方法和步骤简述如下：首先进行数据集划分，在每类样本集中随机抽取 $L(L=3,4,5)$个样本作为训练数据集，其余的样本作为测试数据集；然后在训练数据集上使用数据降维算法，得

① http://www.uk.research.att.com/facedatabase.html
② http://www.itl.nist.gov/iad/humanid/feret/
③ http://rvl1.ecn.purdue.edu/~aleix/aleix_face_DB.html
④ http://www.cad.zju.edu.cn/home/dengcai/Data/FaceData.html
⑤ http://www.iis.ee.ic.ac.uk/icvl/code.htm

到对应的低位表示和投影矩阵 $\boldsymbol{V}$；其次采用投影矩阵 $\boldsymbol{V}$ 将测试数据投影至低维子空间；最后在低维子空间中，针对测试数据集采用最近邻算法进行分类. 在数据降维算法评价环节，实验中针对每个数据集进行了 20 次随机划分，并在不同目标维数上进行投影，最后输出了投影之后低维表示数据上的最近邻平均分类准确率和对应的标准差. 取得最高平均分类准确率的目标维数称为最优嵌入维数，对应的平均分类准确率称为最优分类准确率. 实验中将这两个指标作为数据降维算法的评价指标.

实验中的参数均由经验给出，其中近邻参数 k 设定为 3，MFA 中同类近邻参数 k_1 设定为 2，异类近邻参数 k_2 设定为 10. 另外实验中采用 PCA 预处理来克服散度矩阵奇异性问题，其中 LDA 和 MFA 中主成分贡献率设置为 0.95，SOLDE 中主成分贡献率设置为 0.9. 在局部差异图权重矩阵计算过程中，参数 t 的取值根据如下公式设置[23]：

$$t=\frac{1}{k^2}\sum_{j=1}^{k}\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2 \tag{16}$$

这里 k 是数据建图时的近邻参数. 由式(16)可知，t 可看作是某个样本点到所有近邻样本点距离平方和的均值，由当前样本点到其局部邻域内样本点的距离决定，相当于对局部邻域内所有样本点之间距离的尺度进行规整化，参数 t 随着样本点邻域的变化而自适应调整，使得每个样本点局部邻域内样本点之间的差异性权重取值范围大致保持一致. t 值的大小是动态自适应调整的，该邻域内的样本点越分散，则对应的 t 值也越大，由此生成的差异图权重应该越小，这样才能很好的保持局部邻域内样本点之间的差异性.

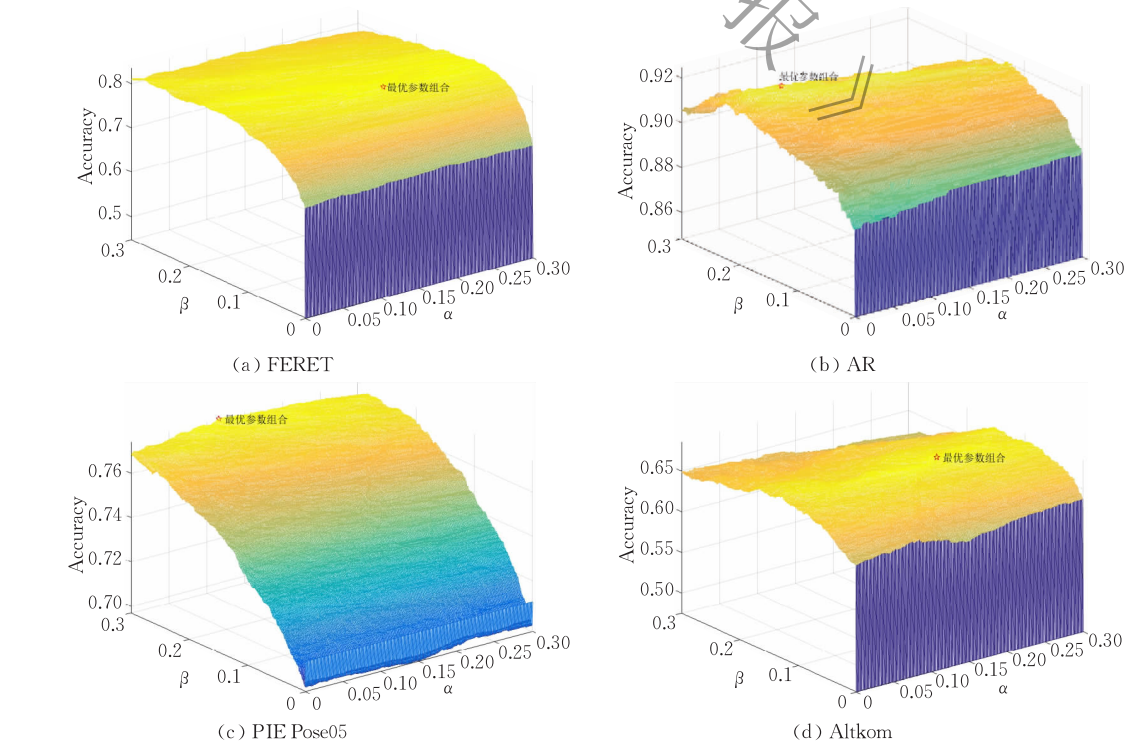
5.2.2　正则化参数灵敏度分析

RMDP 算法中正则化参数的设置对人脸图像数据集的分类准确率有较大影响，其中 α 用以衡量数据集中局部差异性保持的程度，β 用于衡量样本不相似性保持的程度. 实验中针对每个数据集，在[0,0.3]区间内以 0.001 为步长进行搜索，表 2 给出了正则化参数的经验设置范围.

表 2　RMDP 算法中的正则化参数设置

数据集	正则化参数 α	正则化参数 β
ORL	0.200～0.300	0.200～0.210
FERET	0.001～0.300	0.050～0.200
AR	0.030～0.150	0.170～0.200
PIE Pose05	0.080～0.300	0.250～0.300
Altkom	0.003～0.300	0.008～0.120

为了进一步分析正则化参数 α 和 β 对降维算法判别能力的影响，图 9 显示了 FERET、AR、PIE Pose05 和 Altkom 数据集上 RMDP 算法降维之后特征表示的分类准确率随着正则化参数 α 和 β 的变化趋势. 从图 9 可以看出，各个数据集上的分类准确率对正则化参数 β 更为敏感，随着 β 的增大，分类准



(a) FERET　(b) AR　(c) PIE Pose05　(d) Altkom

图 9　不同正则化参数设置下 RMDP 算法分类准确率变化趋势

确率先上升之后趋于稳定.

5.2.3 分类结果比较与分析

ORL、FERET、AR、PIE Pose05 和 Altkom 的人脸图像数据集上的分类结果如表 3 至表 7 所示.表中分别报告了每个人脸图像数据集上不同降维算法所对应的的最优分类准确率和标准差以及相应的最优嵌入维数(如表中括号内数字所示).图 10 至图 14 分别显示了各个数据集上不同降维算法的平均分类准确率随着目标维数的变化情况.

由表 3 至表 7 可知,在每个数据集上,RMDP 的最优分类准确率均高于其他数据降维算法,且对应

表 3 ORL 数据集上分类结果比较

方法	3 个标记样本	4 个标记样本	5 个标记样本
PCA	77.89±2.60(50)	83.79±2.81(49)	87.68±2.46(48)
LDA	87.55±2.14(39)	91.63±2.25(39)	94.08±2.23(39)
LPP	83.73±2.50(30)	88.40±2.66(30)	91.30±1.62(30)
MFA	88.38±2.50(42)	91.21±2.26(60)	94.08±1.98(56)
MDP	90.16±2.66(57)	94.00±2.09(55)	96.40±1.66(55)
SOLDE	87.25±2.11(29)	91.77±1.93(24)	94.80±1.78(42)
MPDA	87.18±2.56(36)	93.42±1.94(38)	95.70±1.58(39)
RMDP	**91.59±2.13(43)**	**95.13±1.54(46)**	**96.68±1.86(36)**

表 4 FERET 数据集上分类结果比较

方法	3 个标记样本	4 个标记样本	5 个标记样本
PCA	32.88±1.30(60)	38.19±1.56(60)	41.44±1.91(60)
LDA	37.65±1.96(55)	35.59±2.00(57)	32.86±2.03(60)
LPP	6.34±1.14(30)	5.56±0.75(30)	5.49±1.20(30)
MFA	41.73±2.12(60)	53.90±2.43(47)	66.84±2.09(39)
MDP	76.22±1.45(25)	81.96±1.40(25)	84.73±1.74(29)
SOLDE	64.93±1.68(21)	73.13±1.80(24)	77.06±1.96(36)
MPDA	78.85±1.44(16)	84.00±0.89(19)	84.95±1.25(19)
RMDP	**84.21±1.03(33)**	**87.54±1.07(42)**	**89.54±0.84(40)**

表 5 AR 数据集上分类结果比较

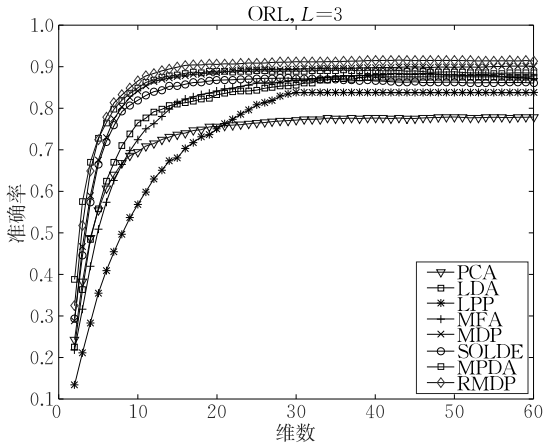
方法	3 个标记样本	4 个标记样本	5 个标记样本
PCA	42.00±1.41(60)	48.77±1.34(60)	54.05±2.05(60)
LDA	84.64±1.36(60)	88.65±1.08(60)	89.87±0.97(60)
LPP	68.12±1.38(30)	73.59±1.77(30)	75.99±1.84(30)
MFA	85.83±1.38(60)	90.72±1.21(60)	92.95±1.19(60)
MDP	87.46±1.31(60)	92.05±0.85(60)	94.24±0.82(60)
SOLDE	63.47±5.52(50)	78.97±2.90(54)	86.65±2.27(58)
MPDA	81.96±1.68(60)	88.95±1.28(60)	92.07±1.42(60)
RMDP	**90.77±1.14(60)**	**94.49±0.98(58)**	**96.24±0.61(60)**

表 6 PIE Pose05 数据集上分类结果比较

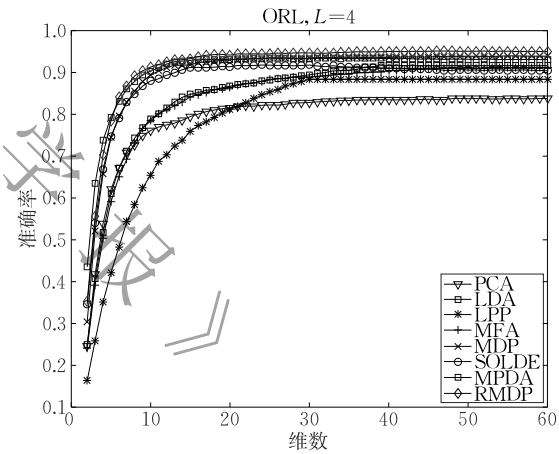
方法	3 个标记样本	4 个标记样本	5 个标记样本
PCA	33.00±1.05(60)	39.49±0.90(60)	44.89±1.02(60)
LDA	72.26±1.96(60)	77.61±1.64(60)	79.84±1.84(60)
LPP	74.75±1.79(30)	80.93±1.71(30)	84.05±1.21(30)
MFA	71.43±1.79(60)	78.03±1.59(60)	81.79±1.13(60)
MDP	76.32±2.01(60)	82.46±1.41(60)	85.98±0.97(60)
SOLDE	70.75±2.06(57)	79.24±1.51(60)	83.99±1.00(31)
MPDA	71.53±1.80(60)	79.17±1.49(60)	83.38±0.93(60)
RMDP	**77.41±2.02(60)**	**83.23±1.42(60)**	**86.42±0.97(59)**

表 7 Altkom 数据集上分类结果比较

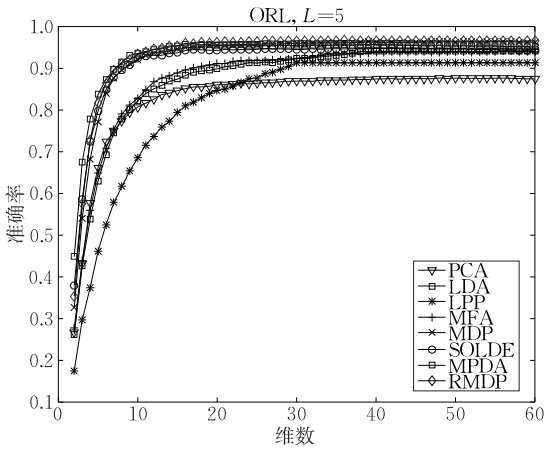
方法	3 个标记样本	4 个标记样本	5 个标记样本
PCA	18.11±1.45(60)	20.52±1.02(60)	22.23±1.41(60)
LDA	50.59±2.11(57)	60.48±1.49(59)	66.97±2.29(60)
LPP	29.98±1.86(30)	35.63±2.31(30)	41.13±2.43(30)
MFA	55.31±2.16(58)	62.57±2.19(56)	69.41±2.09(60)
MDP	57.49±1.90(59)	67.44±1.74(50)	74.46±1.87(50)
SOLDE	38.04±1.90(59)	50.38±1.58(39)	59.33±1.82(28)
MPDA	45.03±1.68(58)	55.06±1.43(29)	63.76±2.33(26)
RMDP	**63.32±1.64(58)**	**73.75±1.35(47)**	**80.69±1.43(60)**



(a) 3个标记样本

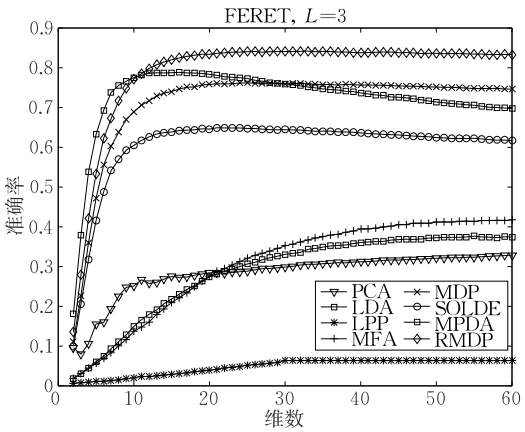


(b) 4个标记样本

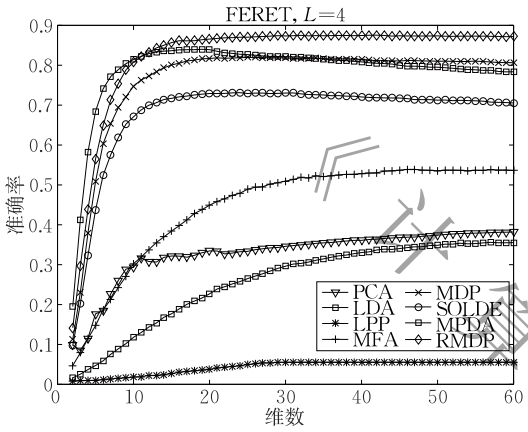


(c) 5个标记样本

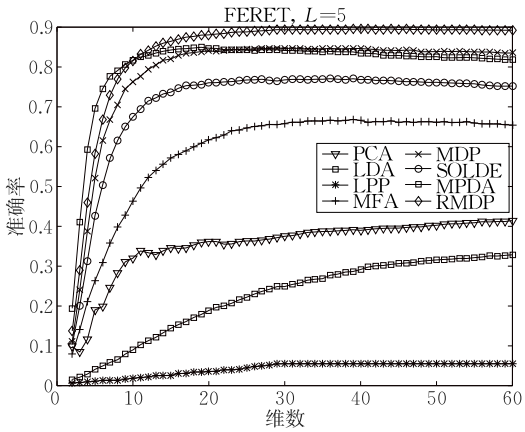
图 10 ORL 数据集上不同目标维数下的分类结果比较



(a) 3个标记样本

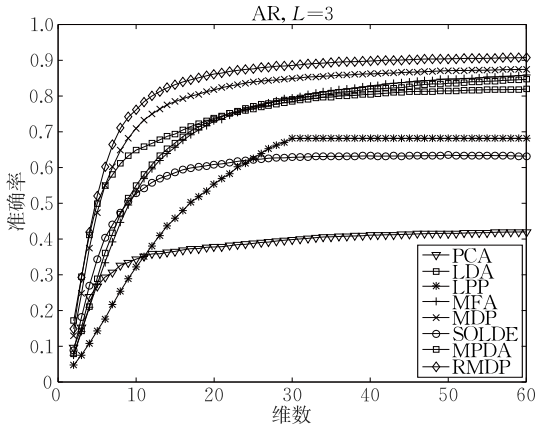


(b) 4个标记样本

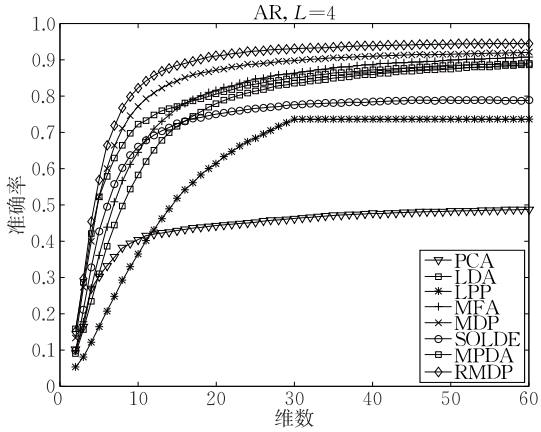


(c) 5个标记样本

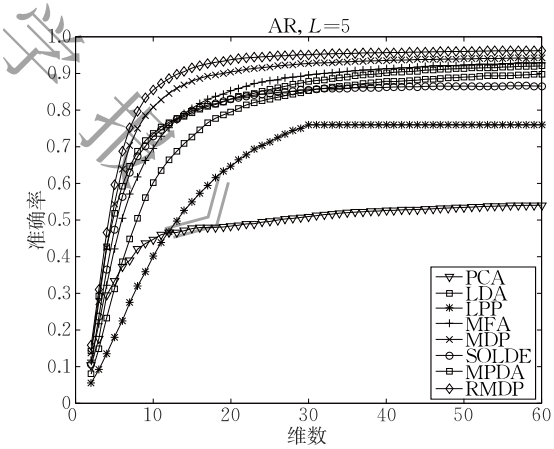
图 11　FERET 数据集上不同目标维数下的分类结果比较



(a) 3个标记样本



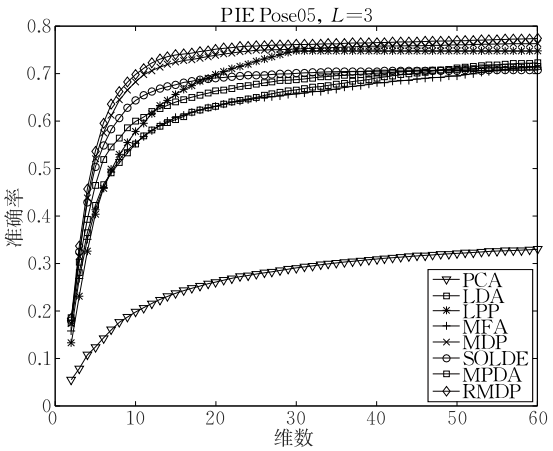
(b) 4个标记样本



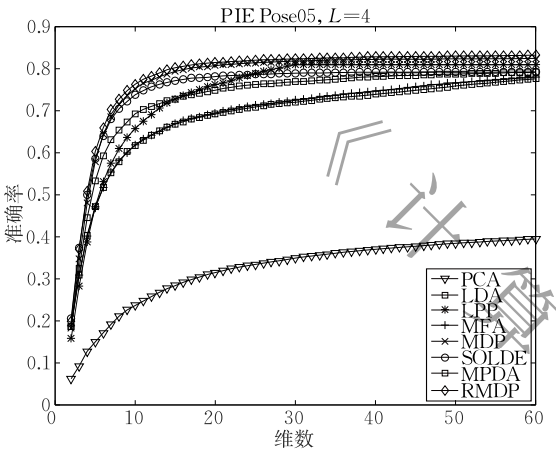
(c) 5个标记样本

图 12　AR 数据集上不同目标维数下的分类结果比较

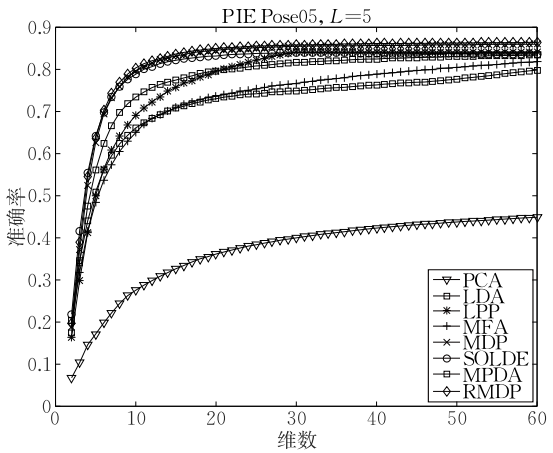
的标准差也相对较小，这在一定程度上说明 RMDP 算法的稳定性. 尽管 RMDP 和 SOLDE 在数据建模时都考虑样本点之间的局部差异性，但是 RMDP 算法建模准则与 SOLDE 不同，且计算效率更高. 这是因为 SOLDE 是采用迹比优化准则建模，在数值计算时需要采用 PCA 进行预处理，以此来避免出现矩阵奇异性问题，这样可能会导致数据集中一些有用的判别信息在 PCA 预降维时丢失. 而 RMDP 算法是采用迹差优化准则进行数据建模，矩阵奇异性问题对数值计算没有影响，所以不需要采用 PCA 进行预处理，这样就保证了降维过程中原始高维数据集中蕴含的判别信息不会因为 PCA 预处理而丢失. 另外，由于 RMDP 在数据降维时将判别图嵌入、相似图嵌入和差异图嵌入融合在一起进行优化建模，综合考虑了高维数据集各种内在结构信息，而传统的 PCA、LDA、LPP、MFA、MDP 和 MPDA 算法只考虑了其中的一种或者两种内在结构信息，因此 RMDP 在数据降维实验上最优分类准确率高于其他方法.



(a) 3个标记样本

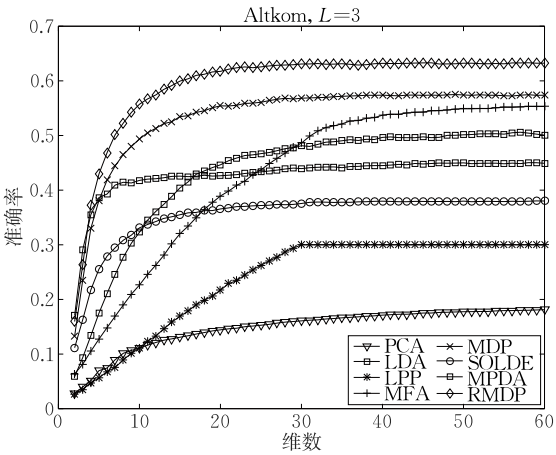


(b) 4个标记样本

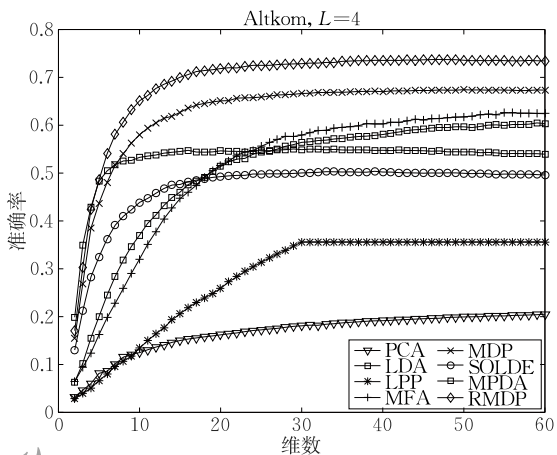


(c) 5个标记样本

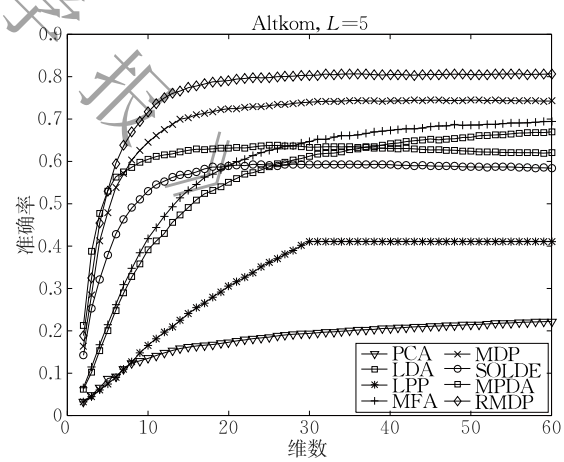
图 13 PIE Pose05 数据集上不同目标维数下的分类结果比较



(a) 3个标记样本



(b) 4个标记样本



(c) 5个标记样本

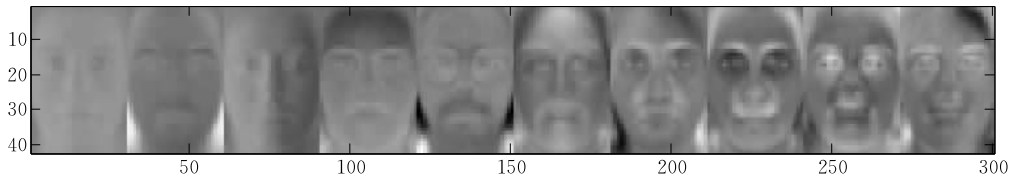
图 14 Altkom 数据集上不同目标维数下的分类结果比较

RMDP 是由 MDP 算法改进而来，由于在 MDP 模型中融合了局部差异信息，并采用正则化方法将其归结为易于求解的迹差优化问题，正则化参数可根据数据集的特性进行设置，因此 RMDP 算法更加灵活，且具有更好的判别能力.

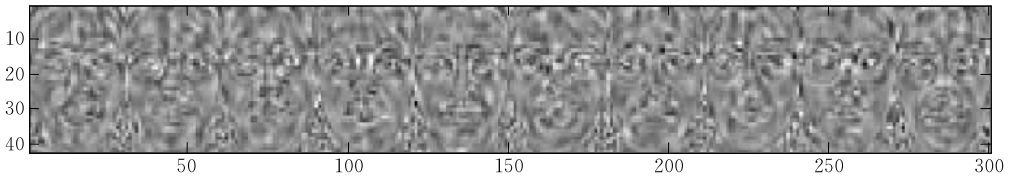
同时注意到，RMDP 和 MDP 算法在 ORL、AR、PIE Pose05 和 Altkom 数据集上的最优平均分类准确率高于其它算法，这表明最大化类间最小距离和最小化类内最大距离的降维算法设计思想在判别特征提取中具有重要作用. RMDP 算法在 Altkom、FERET和 AR 数据集上的平均分类准确率显著高于其他算法，结合图 8 中的示例样图像本可以看出，RMDP 算法对人脸图像中的姿态和表情变化具有一定的鲁棒性. 这在一定程度上说明，采用边界样本

点进行数据建模，对数据集的内部变化具有一定的稳健性.

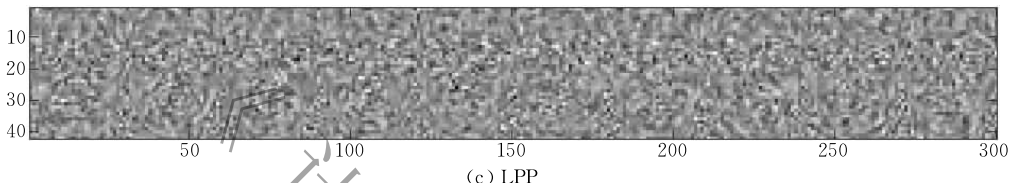
为了直观分析各个数据降维算法所生成投影向量的实际物理意义，图 15 显示了当 $L=3$ 时，PCA、



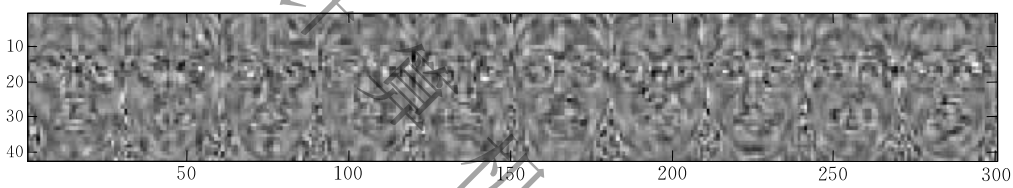
(a) PCA



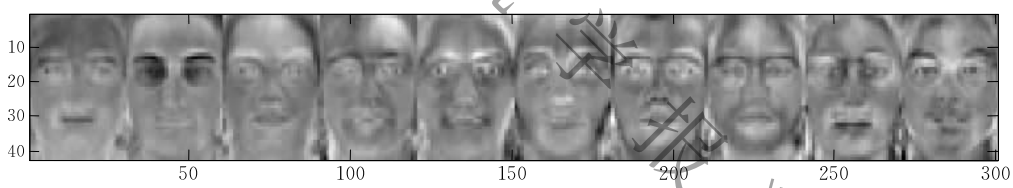
(b) LDA



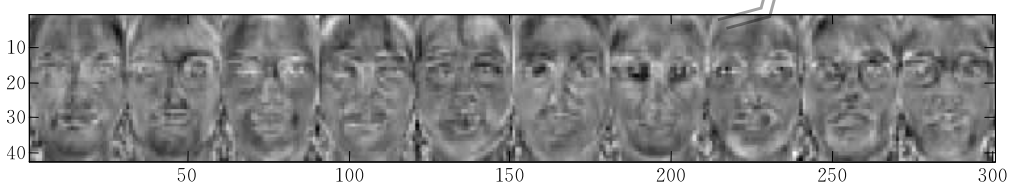
(c) LPP



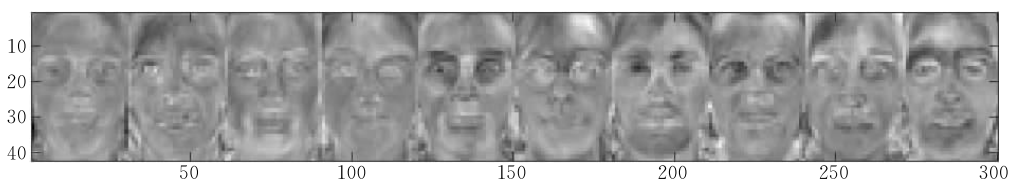
(d) MFA



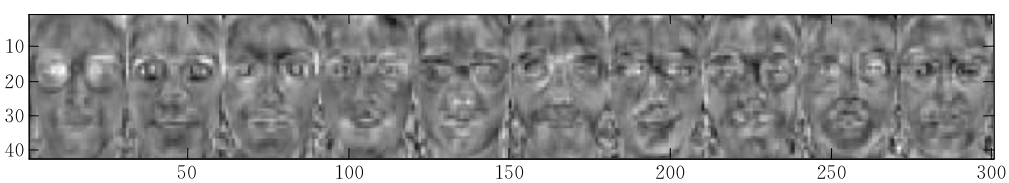
(e) MDP



(f) SOLDE



(g) MPDA



(h) RMDP

图 15　不同算法的特征脸(取前 10 个)比较

LDA、LPP、MFA、MDP、SOLDE、MPDA 和 RMDP 算法在 AR 人脸图像数据集上的前 10 个最优判别特征脸(投影矩阵 $\boldsymbol{V}$ 中前 10 个列向量对应的图像).由图 15 可知,PCA、MDP 和 MPDA 算法所生成的特征脸图像比较光滑,人脸轮廓形状明显,重点突出了眼睛、鼻子和嘴巴等人脸部位像素信息;而 LPP、LDA 和 MFA 算法所生成的特征脸图像比较粗糙,其中包含了更多的人脸细节信息;含有局部差异信息的 SOLDE 和 RMDP 算法的特征脸图像则可看作是人脸图像特征脸的整体轮廓信息和局部细节信息的折中.

最后,实验中比较了各个算法在 AR 数据集上的时间消耗,此时目标维数设定为 60,结果如表 6 所示.RMDP 算法的时间消耗与 MDP 相当,但远小于 SOLDE 和 MPDA.

表 8 AR 数据集($L=3$)上的时间消耗比较

方法	时间/s	方法	时间/s
PCA	0.03	MDP	0.10
LDA	0.05	SOLDE	8.63
LPP	0.05	MPDA	10.41
MFA	0.29	RMDP	0.13

6 结 论

为了在挖掘高维数据集中蕴含的判别结构的同时,最大限度的保持数据集的几何结构,本文将局部差异信息融入边界判别投影算法中,并将其归结为正则化的迹差模型.由于局部差异可以描述数据集的类内几何结构,防止降维过程中同类样本点被投影至同一点上,从而最大限度保持原始数据集的拓扑结构,提高降维算法的泛化能力.另外由于正则化边界判别投影算法采用迹差准则进行优化建模,因此生成的投影向量是相互正交的,可以较好的保持原始高维数据中蕴含的全局几何结构,使得该算法具有更高的适用性和鲁棒性.另外算法建模时只考虑了边界样本点,优化求解中采用 QR 分解加速计算,从而降低了计算复杂性.在未来的算法研究中,可将该算法采用核方法进行推广,或者扩展至半监督情形下,以提高算法对不同数据集的适用性.

参 考 文 献

[1] Li Guo-Jie. The scientific value of big data research. Communications of the CCF, 2012, 8(9): 8-15(in Chinese)
(李国杰. 大数据研究的科学价值. 中国计算机学会通讯, 2012, 8(9): 8-15)

[2] Wang J. Geometric Structure of High-Dimensional Data and Dimensionality Reduction. Beijing: Higher Education Press, 2012

[3] Bengio Y, Paiement J F, Vincent P, et al. Out-of-sample extensions for lle, isomap, mds, eigenmaps, and spectral clustering//Proceedings of the Advances in Neural Information Processing Systems 16. Cambridge, UK, 2004: 177-184

[4] Xiang S, Nie F, Zhang C. Learning a mahalanobis distance metric for data clustering and classification. Pattern Recognition, 2008, 41(12): 3600-3612

[5] Ham J, Lee D D, Mika S, et al. A kernel view of the dimensionality reduction of manifolds//Proceedings of the 21st International Conference on Machine Learning. Alberta, Canada, 2004: 47

[6] Jolliffe I T. Principal Component Analysis. 2nd Edition. New York, USA: Springer, 2002

[7] Fukunaga K. Introduction to Statistical Pattern Recognition. 2nd Edition. New York, USA: Academic Press, 1990

[8] He X, Niyogi P. Locality preserving projections//Proceedings of the 16th Advances in Neural Information Processing Systems. Vancouver, Canada, 2003: 153-160

[9] Cox T, Cox M. Multidimensional Scaling. London, UK: Chapman & Hall, 1994

[10] Yan S, Xu D, Zhang B, et al. Graph embedding and extensions: A general framework for dimensionality reduction. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(1): 40-51

[11] Zhang Z, Yan S, Zhao M. Similarity preserving low-rank representation for enhanced data representation and effective subspace learning. Neural Networks, 2014, 53(5): 81-94

[12] Tenenbaum J B, De Silva V, Langford J C. A global geometric framework for nonlinear dimensionality reduction. Science, 2000, 290(5500): 2319-2323

[13] Cortes C, Vapnik V. Support-vector networks. Machine Learning, 1995, 20(3): 273-297

[14] Li H, Jiang T, Zhang K. Efficient and robust feature extraction by maximum margin criterion. IEEE Transactions on Neural Networks, 2006, 17(1): 157-165

[15] Zhou Y, Sun S. Manifold partition discriminant analysis. IEEE Transactions on Cybernetics, 2016, 99(3): 1-11

[16] Fan M, Zhang X, Lin Z, et al. A regularized approach for geodesic-based semisupervised multimanifold learning. IEEE Transactions on Image Processing, 2014, 23(5): 2133-2147

[17] He Jin-Rong, Ding Li-Xin, Li Zhao-Kui, Hu Qing-Hui. Margin discriminant projection for dimensionality reduction. Journal of Software, 2014, 25(4): 826-838(in Chinese)
(何进荣, 丁立新, 李照奎, 胡庆辉. 基于边界判别投影的数据降维. 软件学报, 2014, 25(4): 826-838)

[18] Gao Q, Ma J, Zhang H, et al. Stable orthogonal local discriminant embedding for linear dimensionality reduction. IEEE Transactions on Image Processing, 2013, 22(7): 2521-2531

[19] Gao Q, Gao F, Zhang H, et al. Two-dimensional maximum local variation based on image euclidean distance for face recognition. IEEE Transactions on Image Processing, 2013, 22(10): 3807-3817

[20] Sugiyama M. Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. The Journal of Machine Learning Research, 2007, 8(1): 1027-1061

[21] Dai J J, Lieu L, Rocke D. Dimension reduction for classification with gene expression microarray data. Statistical Applications in Genetics and Molecular Biology, 2006, 5(1): 1-21

[22] Aggarwal C C, Zhai C X. Mining Text Data. Germany: Springer Science & Business Media, 2012

[23] Gou J, Yi Z. Locality-based discriminant neighborhood embedding. The Computer Journal, 2013, 56(9): 1063-1082

HE Jin-Rong, born in 1984, Ph.D., lecturer. His research interests include machine learning and data mining.

BI Ying-Zhou, born in 1967, Ph. D., professor. His research interests include intelligent computing and software engineering.

DING Li-Xin, born in 1967, Ph. D., professor. His research interests include intelligent computing and machine learning.

LIU Bin, born in 1981, Ph. D., lecturer. His research interests include parallel computing and machine learning.

Background

Dimensionality reduction is an important preprocessing technique for high-dimensional data analysis. By projecting the high-dimensional data into low-dimensional representation, it can be used for data compression, to reduce costs of data acquisition and storage and improve the efficiency of data transmission, query and computation. Margin discriminant projection aims to maximize inter-class distances and minimize intra-class distances, which may lead to similar samples are projected onto the same point together, so the local topology relationships are interrupted and local geometric structure is distorted. How to preserve global and local structure of data sets is a challenge problem in dimensionality reduction. The proposed regularized margin discriminant projection (RMDP) is proposed, which is an extension of our previous MDP algorithm. RMDP consider local variation in samples, which is defined as weighted distance in a neighborhood of each sample. In optimization model, the diversity of data sets can be preserved by maximizing local variation, which describes the real geometrical structure of data sets. The objectives in inter-class seperability, intra-class similarity and local variation are incorporated as trace difference optimization problem. Therefore, RMDP can be directly used for small sample problem and can be solved efficiently, which has better performance in face recognition experiments.

This work is supported in part by the Science Computing and Intelligent Information Processing of Guangxi Higher Education Key Laboratory (No. GXSCIIP201406), the Yangling Demonstration Zone Science and Technology Plan Project (No. 2016NY-31), and the Doctoral Scientific Research Starting Foundation of Northwest A&F University (No. 2452015302).