

控制与决策

Control and Decision

动态加权非参数判别分析

高云龙, 王志豪, 丁柳, 潘金艳, 王德鑫

引用本文:

高云龙, 王志豪, 丁柳, 等. 动态加权非参数判别分析[J]. 控制与决策, 2020, 35(8): 1866–1872.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2018.1716>

您可能感兴趣的其他文章

Articles you may be interested in

一种参数优化VMD多尺度熵的轴承故障诊断新方法

A new fault diagnosis approach for bearing based on multi-scale entropy of the optimized VMD

控制与决策. 2020, 35(7): 1631–1638 <https://doi.org/10.13195/j.kzyjc.2018.1598>

基于深度去噪核映射的长期预测模型

Deep denoising kernel mapping-based long-term prediction model

控制与决策. 2019, 34(5): 989–996 <https://doi.org/10.13195/j.kzyjc.2017.1451>

基于主元和判别集成分析的模拟电路故障诊断

Fault diagnosis of analog circuit by integrating principal component analysis with linear discriminant analysis

控制与决策. 2015(7): 1321–1324 <https://doi.org/10.13195/j.kzyjc.2014.0774>

自适应局部图嵌入加权罚支持向量机

Adaptive local graph embedding weighted-penalty SVM

控制与决策. 2015(2): 203–214 <https://doi.org/10.13195/j.kzyjc.2014.0017>

基于概率类和不相关判别的半监督局部Fisher方法

Semi-supervised local Fisher method based on probability class and uncorrelated discriminant

控制与决策. 2015(1): 32–38 <https://doi.org/10.13195/j.kzyjc.2013.1673>

动态加权非参数判别分析

高云龙^{1†}, 王志豪¹, 丁柳¹, 潘金艳², 王德鑫¹

(1. 厦门大学 航空航天学院, 福建 厦门 361102; 2. 集美大学 信息工程学院, 福建 厦门 361021)

摘要: 线性判别分析(LDA)是最经典的子空间学习和有监督判别特征提取方法之一. 受到流形学习的启发, 近年来众多基于LDA的改进方法被提出. 尽管出发点不同, 但这些算法本质上都是基于欧氏距离来度量样本的空间散布度. 欧氏距离的非线性特性带来了如下两个问题: 1) 算法对噪声和异常样本点敏感; 2) 算法对流形或者是多模态数据集中局部散布度较大的样本点过度强调, 导致特征提取过程中数据的本质结构特征被破坏. 为了解决这些问题, 提出一种新的基于非参数判别分析(NDA)的维数约减方法, 称作动态加权非参数判别分析(DWNDA). DWNDA采用动态加权距离来计算类间散布度和类内散布度, 不仅能够保留多模态数据集的本质结构特征, 还能有效地利用边界样本点对之间的判别信息. 因此, DWNDA在噪声实验中展现出对噪声和异常样本的强鲁棒性. 此外, 在人脸和手写体数据库上进行实验, DWNDA方法均取得了优异的实验结果.

关键词: 非参数判别分析; 特征提取; 动态加权距离; 局部散布度; 判别信息; 鲁棒性

中图分类号: TP181

文献标志码: A

Dynamic weighted nonparametric discriminant analysis

GAO Yun-long^{1†}, WANG Zhi-hao¹, DING Liu¹, PAN Jin-yan², WANG De-xin¹

(1. College of Aeronautics and Astronautics, Xiamen University, Xiamen 361102, China; 2. College of Information Engineering, Jimei University, Xiamen 361021, China)

Abstract: Linear discriminant analysis(LDA) is one of the most classical subspace learning and supervised learning methods. Inspired by manifold learning, many improved methods based on LDA have been proposed in recent years. Although the motivations of these methods are different, they are all based on the Euclidean distance to measure the spatial dispersion of the samples. The non-linear characteristic of Euclidean distance brings about two problems: 1) these methods are too sensitive to noise and outlier; 2) the essential structure would be destructed, due to the overemphasis of the points which has a large local dispersion in manifold or multimodal datasets. To solve these problems, a new dimension reduction method based on nonparametric discriminant analysis(NDA) is proposed, called a dynamic weighted nonparametric discriminant analysis(DWNDA). Then DWNDA uses the dynamic weighted distance to calculate the within-class and between-class scatters. It can not only retain the essential geometrical structure of multimodal datasets, but also make better use of the discriminant information between marginal point pairs. Hence, the DWNDA shows better robustness to noise and outlier than other methods, which is also demonstrated in experiments. Besides, the DWNDA also shows excellent performance for face and handwriting classification.

Keywords: nonparametric discriminant analysis; feature extraction; dynamic weighted distance; local dispersion; discriminative information; robustness

0 引言

随着信息技术的迅速发展, 获取的数据往往具有数量大、维数高等特点. 针对这些海量的高维数据, 如何从中有效提取所需信息, 发现其内在规律, 是模式识别等领域所面临的基本问题, 而维数约简成为解决此类问题的有效方法之一.

线性判别分析(linear discriminant analysis, LDA)

是最经典的有监督维数约简方法. LDA通过最大化Rayleigh商的方式, 实现最大化类间散布度, 实现类间可分性强, 且最小化类内散布度, 从而使得投影空间中内聚程度大. 这种基于Rayleigh商的特征提取方法能够突出数据集之间的判别信息, 因此在众多模式识别问题上取得优异的性能. 但是, LDA在计算类内散布度的过程中, 存在全局正态分布的模型假设, 而当

收稿日期: 2018-12-18; 修回日期: 2019-04-01.

基金项目: 国家自然科学基金项目(61203176); 福建省自然科学基金项目(2013J05098, 2016J01756).

责任编辑: 阳春华.

[†]通讯作者. E-mail: gaoyl@xmu.edu.cn.

数据集不符合正态分布时,算法的性能将受到极大影响.

为了解决这个问题,近年来不同学者提出了大量的改进算法,例如:RLDA^[1],OCM^[2],LBMMC^[3],NM-FDA^[4],KLDA^[5],HMLDA^[6]等.此外,文献[7]考虑了同类样本的多模态分布特征,而不依赖于全局高斯分布的模型假设;文献[8-9]在考虑同类样本多模态分布特征的基础上,突出了边缘样本的影响,以提高不同类的样本之间的可分性.文献[10]在文献[8]的基础上,将模型从处理二分类问题扩展为多分类问题.文献[11-14]在处理含复杂分布模态的数据集上性能优异,对判别信息的提取能力强,因而受到了广泛的关注.

上述算法都是基于流行学习的方法.在特征提取过程中,这些算法考虑了数据的局部结构特征保留能力,对流形数据和多模态数据展现出很好的性能.但是,这些方法都是基于欧氏距离来度量样本的空间散布度,其非线性变化特征会带来两个问题:一是对大距离样本点对的过度强调作用会造成算法对噪声和异常样本点敏感;二是对于流形或者是多模态数据集,欧氏距离对局部散布度较大的数据过度强调,导致特征提取过程中数据簇本质结构特征遭到破坏.

为解决这些问题,本文在NDA^[10]基础上,提出一种动态加权非参数判别分析方法(dynamic weighted nonparametric discriminant analysis, DWNDA).该算法在计算类内散布度的过程中,考虑了同类样本的分布模态,以便于提取同类样本的内在几何结构特征;而在计算类间散布度的过程中,算法突出了边缘样本点的影响力,从而降低噪声样本的影响.

1 相关算法

1.1 线性判别分析(LDA)

LDA是一种有监督维数约简方法,目的是寻找低维线性投影子空间,使投影后样本的类间散布度和类内散布度的比值达到最大.LDA的模型如下所示:

$$\begin{cases} \max_W J_{LDA}(W) = \frac{\text{tr}(W^T S_b W)}{\text{tr}(W^T S_w W)}; \\ \text{s.t. } W^T W = I. \end{cases} \quad (1)$$

而类间散布度和类内散布度分别为

$$S_b = \sum_{j=1}^C n_j (u_j - u)(u_j - u)^T, \quad (2)$$

$$S_w = \sum_{j=1}^C \sum_{i=1}^{n_j} (x_i - u_j)(x_i - u_j)^T. \quad (3)$$

其中: W 为投影矩阵, C 为数据集中的类别数, $u =$

$\frac{1}{n} \sum_{i=1}^n x_i$ 为 n 个样本点的均值, $u_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_i^j$ 为第 j 个类别的 n_j 个样本的均值.根据拉格朗日乘法可得, W 的最优解为矩阵 $S_w^{-1} S_b$ 的前 d (d 为降维后数据集的新维度)个最大的特征值对应的特征向量所组成的矩阵.

LDA通过最大化类间散布度、最小化类内散布度的方式来构建投影空间,保证了子空间中的不同类别具有较强的可分性,但LDA对数据集存在高斯分布的模型假设,因而在处理非高斯分布的数据集时,算法性能较差;其次,数据经LDA投影后的数据集维度至多只能是 $C - 1$ 维(C 为类别数),这使得在处理高维数据时,算法可能会丢失一些重要特征.而对于高维小样本数据集而言,类内散布度矩阵往往不能满秩;并且,在计算类间散布度时LDA只考虑了各类的中心,未能有效地利用不同类别边界样本之间的判别信息.针对这些问题,众多改进算法被提出,其中非参数判别分析(nonparametric discriminant analysis, NDA)在处理上述问题时表现优异,因而受到了广泛的关注.

1.2 非参数判别分析(NDA)

NDA与LDA的差异主要体现在类内散布矩阵和类间散布矩阵的计算方式上.在NDA中,类内和类间非参数散布矩阵被定义为

$$\begin{aligned} S_w &= \sum_{j=1}^C \sum_{i=1}^{n_j} \sum_{p=1}^{k_1} (x_i^j - N_p(x_i^j, j))(x_i^j - N_p(x_i^j, j))^T, \quad (4) \\ S_b &= \sum_{j=1}^C \sum_{q=1, q \neq j}^C \sum_{p=1}^{k_2} \sum_{i=1}^{n_j} \{w(j, q, p, i) \times \\ &\quad (x_i^j - N_p(x_i^j, q))(x_i^j - N_p(x_i^j, q))^T\}. \quad (5) \end{aligned}$$

其中

$$w(j, q, p, i) = \frac{\min\{d^a(x_i^j, N_p(x_i^j, j)), d^a(x_i^j, N_p(x_i^j, q))\}}{d^a(x_i^j, N_p(x_i^j, j)) + d^a(x_i^j, N_p(x_i^j, q))}, \quad (6)$$

k_1 和 k_2 分别为类内和类间的kNN参数,用来确定样本点的局部邻域包含的样本点个数; $N_p(x_i^j, j)$ 表示从第 j 类中第 i 个样本 x_i^j 在第 j 类中的第 p 近邻样本点; $d(x_1, x_2)$ 表示 x_1 和 x_2 之间的欧氏距离; $w(j, q, p, i)$ 为加权函数.为了突出边缘样本点对之间的判别信息, $w(j, q, p, i)$ 随着样本点到边界距离的增加,具有快速衰减特性.在式(6)中,通过引入参数 a ($a \geq 0$)来调节 $w(j, q, p, i)$ 的快速衰减特性.

与LDA相比,NDA在计算 S_w 时不存在假设分布先验,引入了数据分布的局部结构特征,有利于算法提取流形数据集中复杂的几何结构.同时,NDA在计算 S_b 时突出了边缘样本点对之间的可分性,这一措施提高了子空间中不同类别间的可分性.

在实际应用中,NDA依然存在一些缺点,表现在以下两个方面:1)NDA在降维过程中无法有效挖掘数据的本质结构.例如:对于样本空间中两个具有相同几何结构,不同尺度的数据簇,式(4)和(5)定义的 S_w 和 S_b ,尺度较大的数据簇比尺度较小的数据簇占优势地位.这一特点导致NDA在降维过程中无法有效挖掘数据的本质结构.2)NDA对判别信息的提取极易受到边缘噪声样本的影响.式(6)给出的加权方式使得边界样本点对之间的判别信息得到强调,但同时也增大了边缘噪声样本的影响.

2 动态加权非参数判别分析建模

在实际应用中,数据集复杂的分布特征往往难以被提取.维数约简算法对数据集的模型假设使得算法在处理相应的数据集(数据集的分布基本符合模型假设)时具有较好的表现,但同时也使得算法在处理其他数据集时性能明显下降,算法的实用性不足.为了解决这一问题,将思路从对样本点按所属类别的分布特征进行归一化,转化为对样本点按局部分布特征进行归一化.基于上述分析,为了有效地提取流形数据的本质结构特征,在计算类内、类间散布度的过程中,按照样本点的局部分布特征进行归一化,建立自适应类内、类间散布度计算方式.这里的自适应性主要体现在样本点之间的距离计算方式随着样本点的局部分布特征而自适应调整.

对于每个样本点 x_i^j ,从第 j 类中选择 x_i^j 的前 k 个近邻样本组成局部邻域 $\{N_j(x_i^c, c) | j = 1, 2, \dots, k\}$,而 $N_p(x_i^j, j)$ 为 x_i^j 在第 j 类的第 p 近邻样本点.局部邻域半径 $r_{x_i^j}$ 可采用下式计算:

$$r_{x_i^j} = d(N_p(x_i^j, j), x_i^j). \quad (7)$$

基于kNN概率密度估计,样本点 x_i^j 的概率密度为

$$\hat{P}_{\text{kNN}}(x_i^j, k) = \frac{k}{n_j V_k(x_i^j)}. \quad (8)$$

其中: n_j 为第 j 类样本点的个数; $V_k(x_i^j)$ 为 x_i^j 的 k 个近邻样本点组成局部邻域的超球体的体积,其计算公式为 $V_k(x_i^j) = \frac{\pi^{d'/2}}{\Gamma(d'/2 + 1)} r_{x_i^j}^{d'}$, d' 为数据的维度.由式(8)可以看出, x_i^j 的概率密度与 $r_{x_i^j}^{d'}$ 成反比.因此对于同一个数据簇,可以用 $1/r_{x_i^c}^{d'}$ 代替样本点的先验概率.

基于 $r_{x_i^j}^2$,本文的动态加权类内散布度和类间散布度分别定义为

$$S_w = \sum_{j=1}^C \sum_{i=1}^{n_c} \sum_{p=1}^{k_1} \frac{1}{r_{x_i^j}^2} (x_i^j - N_p(x_i^j, j))(x_i^j - N_p(x_i^j, j))^T, \quad (9)$$

$$S_b = \sum_{j=1}^C \sum_{q=1}^C \sum_{p=1}^{k_2} \sum_{i=1}^{n_c} \left\{ \frac{w(j, q, p, i)}{r_{x_i^j}^2} \times (x_i^j - N_p(x_i^j, q))(x_i^j - N_p(x_i^j, q))^T \right\}, \quad (10)$$

其中 $w(j, q, p, i)$ 按式(6)定义.基于式(8)和(9),本文的动态加权非参数判别分析(DWNDA)建模为

$$\begin{cases} \max_W J_{\text{DWNDA}}(W) = \frac{\text{tr}(W^T S_b W)}{\text{tr}(W^T S_w W)}; \\ \text{s.t. } W^T W = I. \end{cases} \quad (11)$$

基于上述分析,本文提出的DWNDA方法通过利用动态加权距离来计算类内和类间散布度,不仅能够降低稀疏区域样本点的作用力,降低噪声样本点的影响,而且还能降低不同类交界处样本点的作用力,减弱不同模态间的相互作用力,增大不同类间的可分性,从而提高算法对数据本质结构的识别能力.

3 算法求解

本文将DWNDA建模为式(10)所示的迹比优化问题.针对这一问题,本文建立如下的构造辅助函数的方法求解优化问题(10):

$$\begin{cases} H(\lambda) = \max_W \text{tr}(W^T S_b W - \lambda W^T S_w W); \\ \text{s.t. } W^T W = I. \end{cases} \quad (12)$$

令式(11)中 $\lambda = \lambda_t$,对于任意给定的 W_t ,有

$$\lambda_t = \frac{\text{tr}(W_t^T S_b W_t)}{\text{tr}(W_t^T S_w W_t)}. \quad (13)$$

如果问题(11)存在一个 W_{t+1} 满足

$$\begin{cases} \text{tr}(W_{t+1}^T S_b W_{t+1} - \lambda_t W_{t+1}^T S_w W_{t+1}) \geq 0, \\ W_{t+1}^T W_{t+1} = I, \end{cases} \quad (14)$$

其中 S_w 为半正定矩阵,则对于任意 W 有

$$\text{tr}(W^T S_w W) \geq 0. \quad (15)$$

结合(14)与(15)有

$$\frac{\text{tr}(W_{t+1}^T S_b W_{t+1})}{\text{tr}(W_{t+1}^T S_w W_{t+1})} \geq \lambda_t. \quad (16)$$

通过当前的 W_t 来确定当前的 λ_t ,构造新的辅助函数,再求解出新的投影矩阵 W_{t+1} .DWNDA正是通过这种迭代的方式,求解出原始模型的最终解投影矩阵 W^* .上述分析指出,模型可以通过构造辅助函数

的方式,迭代求解原模型的极值点. 这里存在3个问题: 1)解的存在性问题; 2)辅助函数法求解; 3)算法的收敛性问题.

3.1 解的存在性问题

令优化问题(10)对应的最优解为 W^* , 其对应的目标函数值为 λ^* , 则有

$$\max_W J_{DWNDA}(W) = \frac{\text{tr}(W^{*T} S_b W^*)}{\text{tr}(W^{*T} S_w W^*)} = \lambda^*. \quad (17)$$

对于任意满足约束条件 $W^T W = I$ 的 W , 有

$$\frac{\text{tr}(W^T S_b W)}{\text{tr}(W^T S_w W)} \leq \lambda^*. \quad (18)$$

因为 S_w 为半正定矩阵, 所以有

$$\text{tr}(W^T S_b W) - \lambda^* \text{tr}(W^T S_w W) \leq 0. \quad (19)$$

由此得到如下3个结论:

1) 当 $\lambda_t \leq \lambda^*$ 时, 优化问题(11)至少存在一个 W_{t+1} 满足约束(14), 此时有

$$\max_W \text{tr}(W^T S_b W - \lambda_t W^T S_w W) \geq 0.$$

2) 当 $\lambda_t = \lambda^*$ 时, 优化问题(11)存在 W_{t+1} 且 $W_{t+1} = W^*$ 满足约束(14), 此时有

$$\max_W \text{tr}(W^T S_b W - \lambda_t W^T S_w W) = 0.$$

3) 当 $\lambda_t > \lambda^*$ 时, 满足约束条件(14)的 W_{t+1} 不存在, 此时有

$$\max_W \text{tr}(W^T S_b W - \lambda_t W^T S_w W) < 0.$$

上述3个结论中, 结论1)给出了优化问题(11)至少存在一个 $W_{t+1} = W^*$ 满足约束条件(14)的前提条件. 该条件对下文设计优化算法的初始解的选择具有指导意义; 结论2)给出了原始模型取到最优解的必要条件; 结论3)给出了利用辅助函数求解原模型时, 算法迭代的终止判断条件.

3.2 辅助函数法求解

对于任意 $\lambda_t \leq \lambda^*$, 通过构造辅助函数的方式和拉格朗日乘子法, 原模型优化问题转化为

$$L(W, \Lambda) = \text{tr}(W^T S_b W - \lambda_t W^T S_w W) + \langle W^T W - I, \Lambda \rangle. \quad (20)$$

将式(20)分别对 W 、 Λ 求导, 有

$$\begin{cases} \frac{\partial L(W, \Lambda)}{\partial W} = (S_w - \lambda_t S_b)W + \Lambda W, \\ \frac{\partial L(W, \Lambda)}{\partial \Lambda} = W^T W - I. \end{cases} \quad (21)$$

令一阶导数为零, W 的极值点可通过对矩阵 $S_w - \lambda_t S_b$ 进行特征分解求得. 不妨设新求得的解为 W_{t+1} , 基于结论1)有

$$\text{tr}(W_{t+1}^T S_b W_{t+1} - \lambda W_{t+1}^T S_w W_{t+1}) \geq 0.$$

可得 $\lambda_{t+1} = \text{tr}(W_{t+1}^T S_b W_{t+1}) / \text{tr}(W_{t+1}^T S_w W_{t+1})$ 就是优化问题(11)中比 λ_t 更优的解.

3.3 收敛性问题

对于优化问题(11), 取当前迭代解 $\lambda = \lambda_t$, 不妨设极值点在 $W = W_{t+1}$ 处取得. 根据结论2), 原模型的最优解在 $H(\lambda) = 0$ 处取得. 在 $\lambda = \lambda_t$, 即 $W = W_{t+1}$ 处线性近似为

$$H'(\lambda) = \text{tr}(W_{t+1}^T S_b W_{t+1} - \lambda W_{t+1}^T S_w W_{t+1}). \quad (22)$$

基于牛顿迭代法, 对于线性近似函数, 令 $H'(\lambda) = 0$, 有

$$\begin{aligned} \text{tr}(W_{t+1}^T S_b W_{t+1} - \lambda W_{t+1}^T S_w W_{t+1}) = 0 \Rightarrow \\ \lambda = \frac{W_{t+1}^T S_b W_{t+1}}{W_{t+1}^T S_w W_{t+1}}. \end{aligned} \quad (23)$$

根据牛顿法可得, 优化问题(11)在极值点附近具有平方收敛性.

基于上述分析, DWNDA 的迭代优化算法步骤如下.

step 1: 初始化动态加权类内散布矩阵 S_W , 动态加权类间散布矩阵 S_b , 最大迭代次数 t_{\max} 和目标函数收敛阈值 ε , 令迭代次数 $t = 0$;

step 2: 通过式(13)更新辅助参数 λ_t ;

step 3: 对矩阵 $S_w - \lambda_t S_b$ 进行特征分解, 并将前 d 大的特征值对应的特征向量组成的矩阵作为 W_{t+1} ;

step 4: 若 $\max_W \text{tr}(W^T S_b W - \lambda_t W^T S_w W) < \xi$, 且 $t < t_{\max}$, 则令 $t = t + 1$, 返回到 step 2, 反之算法迭代停止.

4 实验分析

4.1 图像识别

实验选用 Yale A、Yale B、Fei 和 Umist 四个人脸数据集, 将图像大小调成为 32×32 像素, 并采用 PCA 降维到原始信息的 95%. 每个数据集均随机选取一部分图片作为训练样本, 其余的图片作为测试样本. 选用 PCA、LDA、MFA^[15]、NDA^[10] 以及 KLDA^[5] 作为比较算法, 用 kNN 算法对训练样本进行分类. 为了降低随机取样带来的影响, 每组重复试验 100 次, 计算平均识别准确率. 各算法的最优参数设置如表 1 所示. 表 2 给出了 6 种算法在各数据集上的平均最优识别率和对应的特征维数. 图 1~图 4 分别给出了 6 种算法在 4 个数据集上的识别率与投影维度的变化曲线.

如表 2 所示, 在各个数据集上 DWNDA 的识别准确率都远远高于 PCA 和 LDA. 这是因为 PCA 和 LDA 对数据分布存在模型假设, 当假设不成立时, 算法的性能会大大下降.

表1 各算法参数

算法	参数	Yale A 数据库	Yale B 数据库	Fei 数据库	Umist 数据库
MFA	k_1	3	5	3	2
	k_2	9	10	6	6
NDA	$k(k_1 = k_2)$	2	5	3	3
	a	0.13	0.17	0.13	0.14
DWNDA	$k(k_1 = k_2)$	2	5	2	3
	a	0.07	0.3	0.08	0.22

表2 各算法最优识别准确率

算法	Yale A 数据库	Yale B 数据库	Fei 数据库	Umist 数据库	平均最优识别率
PCA	79.43(39)	59.73(70)	81.44 (72)	88.49(70)	77.27
LDA	84.38(13)	85.47(36)	89.04 (38)	95.56(13)	88.61
MFA	86.10(14)	83.66(67)	87.28 (41)	95.17(31)	88.05
NDA	90.86(22)	86.20(60)	88.68 (33)	95.73(14)	90.37
KLDA	91.81(15)	87.81(33)	89.24 (44)	95.73(14)	91.15
DWNDA	93.33(27)	90.03(54)	90.60(36)	96.89(26)	92.71

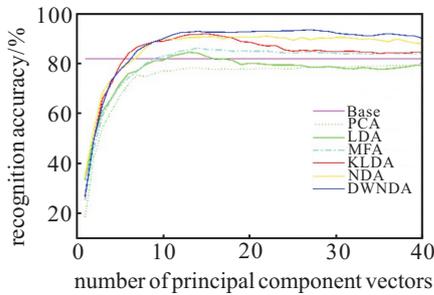


图1 不同算法在Yale A数据库的识别准确率

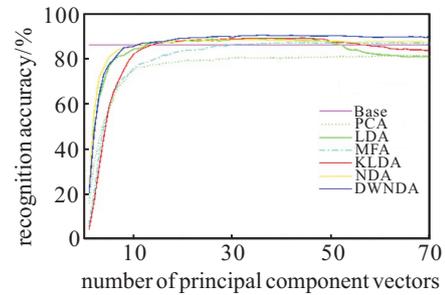


图3 不同算法在Fei数据库的识别准确率

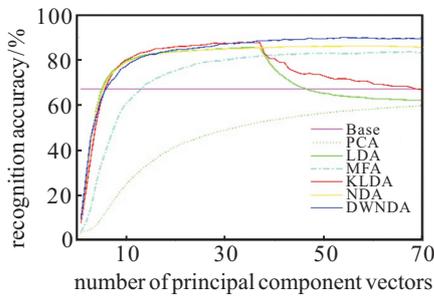


图2 不同算法在Yale B数据库的识别准确率

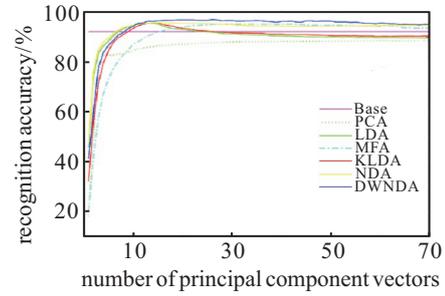


图4 不同算法在Umist数据库的识别准确率

此外, DWNDA在各数据集上的识别准确率均高于NDA和MFA. 数据分布的复杂多样造成采取固定的参数模型(如kNN, 高斯函数)对数据分布的本质结构特征的提取能力不足. NDA和MFA算法正是通过固定的参数模型来构建流型数据集的图描述, 因而它们的性能不如DWNDA. 另外, 与这两种算法相比, DWNDA利用了数据的局部特征, 动态地调整局部邻域内样本点之间的距离度量, 有效地提高了算法对分布特征的描绘能力, 同时降低了边缘噪声样本和分布稀疏区域样本点的影响力, 因而拥有更优的性能.

KLDA采用了核校准的方法, 使得在投影空间中不同类别的样本分布在不同区域, 即维数约简过程突出了不同类别间的可分性. 如表2所示, KLDA在4个

数据集上都取得了比PCA、LDA、MFA三种算法更高的识别精度, 而与NDA相比, 除了在Umist数据集上两种算法效果基本一致, 在其他3个数据集上, KLDA算法的识别精度都高于NDA. 而DWNDA算法在4个数据集上的识别准确率均优于KLDA.

4.2 手写体识别

为验证算法的鲁棒性与抗噪能力, 本实验选取BinAlpha和MNIST两个数据集中的“C”“P”“X”“Z”“0”和“3”六类数据. 此外, 还在训练集中添加了噪声样本. 实验采用PCA预处理数据集, 将各类数据约简到统一维度. 在进行维数约简时, 算法从各类随机选择20个样本作为训练集, 其余19个样本则作为测试样本.

表3给出了LDA、NDA和DWNDA三种算法在含

噪声和不含噪声的数据集上的测试结果. 在不含噪声的情况下 NDA 较 LDA 有 0.88% 的性能提升, 而 DWNDA 较 LDA 有 2.63% 的性能提升. 因为数据集在样本空间中不具有多模态分布, 所以 NDA 和 DWNDA 在计算类内散布度的过程中所采取的一些措施对算法性能的提升影响不大. 但与 LDA 相比, NDA 和 DWNDA 依然具有性能上的优势, 这说明边缘样本点对之间的可分性对算法的性能影响极大, 而 DWNDA 所采取类间散布度计算方式在突出小距离边缘样本点对之间影响力的同时, 降低了边缘噪声样本点的影响. 因此 DWNDA 较 LDA 与 NDA 具有更好的表现. 另外, 在含噪声的情况下, LDA、NDA、DWNDA 都具有一定程度的性能下降, 其中 LDA 下降 8.77%, NDA 下降 7.02%, DWNDA 下降 5.26%. 这从侧面显示了 DWNDA 的强鲁棒性与抗噪能力.

表 3 鲁棒性测试-算法识别准确率的比较

准确率 / %	PCA+LDA	PCA+NDA	PCA+DWNDA
不含噪声	84.21	85.09	86.84
包含噪声	75.44	78.07	81.58
性能下降	8.77	7.02	5.26

4.3 深度学习实验对比

为验证 DWNDA 算法的实用性, 实验采用人脸分类网络模型^[16], 分别对原数据和经过 DWNDA 降维后的数据进行分类处理, 对比两种分类效果. 实验结果表明, DWNDA 能有效地提高网络训练的效率.

1) 模型的分​​类能力不受影响. 实验采用相同的网络, 迭代过程均取 80 个 epoch, 共 1 600 个 step. 由表 4 可见, DWNDA 预处理并没有导致分类网络的性能下降. 原数据集和经 DWNDA 降维后的数据集分别作为网络输入时, 网络在测试集上的分类准确率基本相同. 这说明 DWNDA 能在大幅减少输入数据的维度、减少网络所需神经元、缩小网络规模的同时, 保留数据的本质特征, 体现了 DWNDA 算法对数据本质结构特征的提取能力.

表 4 最终准确率比较

数据集	non-DWNDA / %	DWNDA / %	差异
Fei	98.917	98.950	+0.033
Umist	96.197	96.183	-0.014
Yale A	96.150	96.183	+0.033
Yale B	98.100	98.017	-0.083

2) 加快迭代过程. 如表 5 所示, 经 DWNDA 预处理后, 网络的单步迭代过程所需时间下降 60% 以上. 迭代过程加速明显.

表 5 单步迭代平均用时比较

数据集	non-DWNDA	DWNDA	减少时间量	比值 / %
Fei	0.420	0.142	0.278	33.8
Umist	0.405	0.137	0.268	33.8
Yale A	0.388	0.155	0.233	39.9
Yale B	0.440	0.138	0.302	31.4

3) 优化效果增强. 图 5~图 8 分别展示了在 4 个人脸数据库上未使用预处理和使用 DWNDA 将数据集分别降低到 20、40、60、80 维的情况下, 网络识别准确率的迭代变化情况. 可以看出: 经 DWNDA 处理的 Fei、Umist、Yale A 数据集用于迭代, 准确率提升效果十分出色, 优于未经 DWNDA 处理的数据集; 经 DWNDA 处理的 Yale B 前期收敛较慢, 后期网络优化

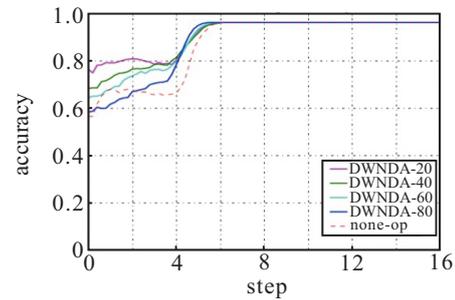


图 5 不同算法在 Yale A 数据库的识别准确率

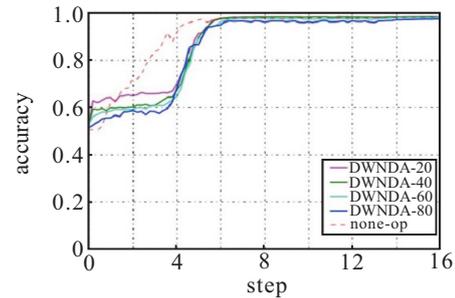


图 6 不同算法在 Yale B 数据库的识别准确率

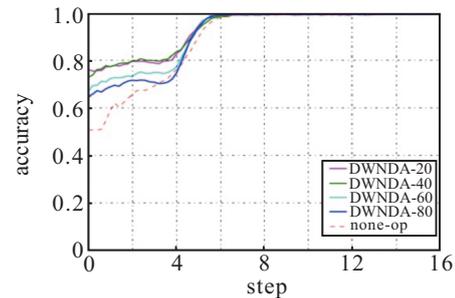


图 7 不同算法在 Fei 数据库的识别准确率

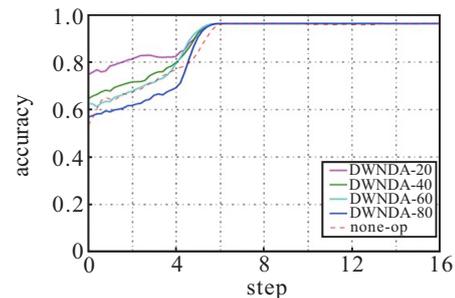


图 8 不同算法在 Umist 数据库的识别准确率

明显,同步到达最大准确率.同时,使用DWNDA算法处理样本集,在合适的维度内,维度下降得越多,优化效果越明显.这表明高维样本经过DWNDA降维后再进行深度学习训练,将从整体上提升优化能力,加速网络的收敛.

5 结论

本文根据局部领域半径提出了基于动态加权距离的方式计算类内、类间散布度,并基于此提出了一种新的维数约简算法,称作动态加权非参数判别分析算法(DWNDA).此外,本文还给出了DWNDA相应的迭代优化算法.DWNDA在计算类内散布矩阵时,考虑了不同模态的分布特征的差异性,根据样本点的局部邻域半径动态地进行调整,从而有效地提取数据分布的本质结构特征;在计算类间散布矩阵时,通过分析边缘样本点局部散布度,动态地降低边缘噪声样本的影响,进而提高算法的分类性能和所提取特征结构的泛化性能.DWNDA在人脸和手写体数据集上的实验结果验证了所提出算法的有效性和实用性.

参考文献(References)

- [1] Guo Y Q. Regularized linear discriminant analysis and its application in microarrays[J]. *Biostatistics*, 2007, 8(1): 86-100.
- [2] Park H, Jeon M, Rosen J B. Lower dimensional representation of text data based on centroids and least squares[J]. *BIT Numerical Mathematics*, 2003, 43(2): 427-448.
- [3] 皋军, 黄丽莉, 王士同. 基于局部子域的最大间距判别分析[J]. *控制与决策*, 2014, 29(5): 827-832.
(Gao J, Huang L L, Wang S T. Local sub-domains based maximum margin criterion[J]. *Control and Decision*, 2014, 29(5): 827-832.)
- [4] 范玉刚, 李平, 宋执环. 基于非线性映射的Fisher判别分析[J]. *控制与决策*, 2007, 22(4): 384-388.
(Fan Y G, Li P, Song Z H. Fisher discriminant analysis based on nonlinear mapping[J]. *Control and Decision*, 2007, 22(4): 384-388.)
- [5] Zheng S, Ding C. Kernel alignment inspired linear discriminant analysis[C]. *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*. Nancy, 2014: 401-416.
- [6] Zheng S, Nie F P, Ding C, et al. A harmonic mean linear discriminant analysis for robust image classification[C]. *Proceedings of the 28th International Conference on Tools with Artificial Intelligence*. San Jose: IEEE, 2016: 402-409.
- [7] Masashi S. Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis[J]. *Journal of Machine Learning Research*, 2007, 8(1): 1027-1061.
- [8] Li Z F, Liu W, Lin D H, et al. Nonparametric subspace analysis for face recognition[C]. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. San Diego: IEEE, 2005: 961-966.
- [9] Yan S C, Xu D, Zhang B Y. Graph embedding and extensions: A general framework for dimensionality reduction[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, 29(1): 40-51.
- [10] Li Z F, Lin D H, Tang X O. Nonparametric discriminant analysis for face recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(4): 755-761.
- [11] Harandi M, Salzmann M, Hartley R. Dimensionality reduction on SPD manifolds: The emergence of geometry-aware methods[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 40(1): 48-62.
- [12] Cao G Q, Iosifidis A, Gabbouj M. Multi-view nonparametric discriminant analysis for image retrieval and recognition[J]. *IEEE Signal Processing Letters*, 2017, 24(10): 1537-1541.
- [13] Gyamfi K S, Brusey J, Hunt A, et al. Linear classifier design under heteroscedasticity in linear discriminant analysis[J]. *Expert Systems with Applications*, 2017, 79: 44-52.
- [14] Wang H, Wang Y T, Zhou Z, et al. CosFace: Large margin cosine loss for deep face recognition[C]. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018: 5265-5274.
- [15] Xu D, Yan S C, Tao D C, et al. Marginal fisher analysis and its variants for human gait recognition and content-based image retrieval[J]. *IEEE Transactions on Image Processing*, 2007, 16(11): 2811-2821.
- [16] Schroff F, Kalenichenko D, Philbin J. FaceNet: A unified embedding for face recognition and clustering[C]. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Boston: IEEE, 2015: 815-823.

作者简介

高云龙(1979—),男,副教授,博士,从事机器学习、时间序列分析等研究, E-mail: gaoyl@xmu.edu.cn;

王志豪(1993—),男,硕士生,从事机器学习和模式识别的研究, E-mail: zhwang@stu.xmu.edu.cn;

丁柳(1993—),女,硕士生,从事模式识别和维数约简的研究, E-mail: 838824384@qq.com;

潘金艳(1978—),女,副教授,博士,从事人工智能和机器学习等研究, E-mail: jypan@jmu.edu.cn;

王德鑫(1996—),男,本科生,从事机器学习和计算机视觉的研究, E-mail: 34520152201338@stu.xmu.edu.cn.

(责任编辑: 孙艺红)