文章编号: 1001-0920 (2012) 10-0000-00

局部学习支持向量机

陶剑文[1,2], 王士同[1]

(1. 江南大学 信息工程学院，江苏 无锡 214122；2. 浙江工商职业技术学院 工学院，浙江 宁波 315012)

摘 要: 针对传统支持向量机不能较好地利用数据空间局部信息的问题, 提出一种基于局部学习的支持向量机. 通过同时最小化局部内散度和最大化局部间散度信息来寻求一个最优的分类决策函数. 为了更好地反映数据的局部几何特征, 该方法采用适于局部学习的测地线距离来度量数据点对间的相似性. 另外, 通过引入一个能同时控制间隔误差上界和支持向量下界的参数 μ, 以进一步提升学习泛化能力. 最后, 人造和实际数据集实验验证了所提出方法的有效性.

关键词: 局部学习；流形学习；支持向量机；散度
中图分类号: TP181 **文献标识码:** A

Local learning based support vector machine

TAO Jian-wen[1,2], *WANG Shi-tong*[1]

(1. School of Information Engineering，Southern Yangtze University，Wuxi 214122，China；2. School of Information Engineering，Zhejiang Business Technology Institute，Ningbo 315012，China. Correspondent：TAO Jian-wen，E-mail：tjw@zjbti.net.cn)

Abstract: The classic support vector machine(SVM) can not efficiently exploit the local information of data points, which is useful for pattern recognition. Therefore, a so-called local learning based support vector machine is presented to address those problems mentioned above, which makes full use of the local information such as intra-locality scatter and inter-locality scatter of the data sets to search an optimal decision function by minimizing the intra-locality scatter and simultaneously maximizing the inter-locality scatter. Meanwhile, the proposed method adopts geodesic distance metric to measure the distance between data, which can reflect the true local geometry of data space. In addition, an additional parameter μ is introduced to control both the super bound on the fraction of margin errors and the lower bound on the fraction of support vectors, thus improving the generalization capacity of the proposed method. Finally, extensive experiments show the effectiveness of the proposed method on the artificial and real world problems.

Key words: local learning；manifold learning；support vector machine；scatter

1 引 言

模式分类旨在通过有限的训练样本学习一个分类器, 且该分类器须对未来数据具有良好的泛化能力[1]. 目前, 已有多种用于模式分类的方法提出, 其中支持向量机 (SVM)[2]及其相关变体是实现模式分类的主流方法之一[3]. 尽管SVM及其变体已在机器学习和模式识别领域得到了广泛而成功地应用[1], 但在通过最大化间隔分割二类时却没有考虑类内数据的局部信息 (或数据散度), 从而导致 SVM 的优化解不具有鲁棒性[3-5]. 为了克服 SVM 的缺陷, 一直以来, 在分类问题中利用数据分布信息是一个很重要的研究主题[6]. 近来, Zafeiriou 等[7]基于 Fisher 线性判别分析 (FLDA) [8]的思想, 提出一种最小类方差 SVM(MCVSVM), 通过类内散度来正则化 SVM, 即在确保类内散度最小化的同时实现类间间隔最大化, 从而使 MCVSVM 的优化解具有一定的鲁棒性.

流形学习[9-12]旨在通过数据的局部信息直接发现数据的全局的非线性几何结构, 所学习的流形是一个嵌入在高维输入空间的本质低维空间. 为了关注数据的局部流形信息, Wang 等[3]将局部保留散度引入 MCVSVM, 提出一种最小类内局部保留方差支持向量机 (MCLPVSVM), 它在分类中考虑了数据的局部流形. 文献 [4,13] 指出, 现有局部保留类方法在保持模式之间的局部信息时, 忽略了模式之间的非局部

收稿日期: 2011-05-05；**修回日期:** 2011-07-26.
基金项目: 国家自然科学基金项目 (60975027, 60903100)；宁波市自然科学基金项目 (2009A610080).
作者简介: 陶剑文(1973−), 男, 副教授, 博士生, 从事模式识别与数据挖掘等研究；王士同(1964−), 男, 教授, 博士生导师, 从事人工智能、机器学习等研究.

信息, 导致识别性能下降. 另外, 该类方法通过惩罚因子来最小化局部离散度, 使得邻域内的样本投影后比较接近, 当邻域内的样本过于接近时, 容易造成邻域内样本之间的差异信息丢失(即过学习问题), 降低了模式分类性能[13].

作为大间隔学习机, MCVSVM 和 MCLPVSVM 等方法均考虑了数据的局部信息, 具有比经典 SVM 类方法更优的分类性能[3]. 但是, MCVSVM 仅考虑了类内数据的散度信息, 而忽视了数据空间的局部流形信息; MCLPVSVM 在考虑样本空间的局部流形结构时忽略了样本的非局部信息, 使得对于模式分类问题同样不能取得最优的学习性能. 对此, 本文提出一种新的局部学习支持向量机 v-LPMIVSVM. 对于模式分类问题, v-LPMIVSVM 充分利用数据空间的局部内和局部间信息, 通过最小局部局部内散度和最大局部间散度寻求一个最优的模式分割超平面, 其中散度信息通过一个代表数据点间邻居关系的邻接图来建模.

2 v-LPMIVSVM

2.1 问题描述

为了简单起见, 本文主要考虑二元分类任务, 对于多类分类问题, 可采用一对一方法将其转化为多个二元分类问题. 对于一个包含 N 个模式的二元分类问题, 设给定训练集 $T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$. 其中: $x_i\in X\subset R^d$ 为输入数据, $y_i\in\{+1,-1\}$ 为类标签, $1\leqslant i\leqslant N$.

设 $N_k(x_i)$ 为数据点 x_i 的 k 近邻集, G 代表数据集 X 的加权邻接图, 其中第 i 个顶点代表数据点 x_i. 如果 $x_i\in N_k(x_j)$ 或 $x_j\in N_k(x_i)$, 则 G 中顶点 i 与 j 相连. 有多种计算图 G 的权值矩阵 W 的方法, 其中较常用的方法是采用如下高斯型热核函数[3]:

$$W_{ij}=\exp(-d(x_i,x_j)^2/t). \tag{1}$$

其中: $t>0$ 为热核参数, 可通过交叉验证确定; $d(x_i,x_j)$ 为 x_i 与 x_j 间距离. 为了有效反应数据点对间的局部流形结构, 本文采用测地线距离来度量点对间的距离[12].

为了更好地描述 v-LPMIVSVM 问题, 首先给出如下相关定义:

定义 1(局部内散度矩阵) 设 L 为数据集 X 的 Laplacian 矩阵, 矩阵

$$H_L=XLX^{\mathrm T}=X(D-A)X^{\mathrm T}$$

称为局部内散度矩阵. 其中: D 为一对角矩阵, 且 $D_{ij}=\sum_j A_{ij}$; A 为权重矩阵, 定义为

$$\overline{A}_{ij}=\begin{cases}W_{ij},\ x_i\in N_k(x_j)\text{ or }x_j\in N_k(x_i)\text{ and}\\ \qquad\text{both have the same labels;}\\ 0,\ \text{other.}\end{cases} \tag{2}$$

局部内散度矩阵 H_L 为对称且半正定型, 其形式上与类内散度矩阵相似, 不同的是 H_L 反映的是数据空间的局部几何流形结构.

定义 2(局部间散度矩阵) 设 $\overline{L}$ 为数据集 X 的 Laplacian 矩阵, 矩阵

$$H_B=X\overline{L}X^{\mathrm T}=X(\overline{D}-\overline{A})X^{\mathrm T}$$

称为局部间散度矩阵. 其中: $\overline{D}$ 为一对角矩阵, $\overline{D}_{ij}=\sum_j\overline{A}_{ij}$; $\overline{A}$ 为权重矩阵, 定义为

$$\overline{A}_{ij}=\begin{cases}W_{ij},\ x_i\notin N_k(x_j)\text{ and }x_j\notin N_k(x_i);\\ 0,\ \text{other.}\end{cases} \tag{3}$$

局部间散度矩阵 H_B 为对称且半正定型.

定义 3(局部学习信息矩阵) 由定义 1 和定义 2 可知, 矩阵

$$M=\lambda H_L-(1-\lambda)H_B,\ 0\leqslant\lambda\leqslant 1$$

称为局部学习信息矩阵. 根据以上定义可知, M 为一对称且半正定矩阵.

2.2 v-LPMIVSVM 算法

对于一个二元分类问题, v-LPMIVSVM 的原始优化问题描述为

$$\min_{w,\rho,\xi,b} f=\frac{1}{2}w^{\mathrm T}Mw-\mu\rho+C\sum_{i=1}^N\xi_i. \tag{4}$$

$$\text{s.t.}\ \ y_i(w^{\mathrm T}x_i+b)\geqslant\rho-\xi_i,\ i=1,2,\cdots,N; \tag{5}$$

$$\xi\geqslant 0,\ \rho\geqslant 0. \tag{6}$$

其中: ρ 为类间最小间隔, $\xi=[\xi_1,\xi_2,\cdots,\xi_N]\in R^d$ 为松弛向量, C 为一正则化常量.

v-LPMIVSVM 具有和 v-SVM[14] 相似的原始优化问题形式, 按照文献 [14] 中方法的对偶推导原理, 有如下结论:

定理 1 线性 v-LPMIVSVM 方法原始优化问题 (4)~(6) 的对偶问题如下:

$$\min_\alpha \frac{1}{2}\alpha^{\mathrm T}H\alpha. \tag{7}$$

$$\text{s.t.}\ \ 0\leqslant\alpha_i\leqslant C,\ i=1,2,\cdots,N; \tag{8}$$

$$\sum_{i=1}^N\alpha_iy_i=0; \tag{9}$$

$$\sum_{i=1}^N\alpha_i\geqslant\mu. \tag{10}$$

其中: $\alpha_i\geqslant 0$ 为 Lagrangian 乘子; $[H]_{i,j}=y_iy_jx_i^{\mathrm T}M^{-1}x_j$, M^{-1} 为矩阵 M 的逆运算, 且 v-LPMIVSVM 原始问题中投影向量 w 和偏置变量 b 分别为

$$w^*=M^{-1}\sum_{i=1}^N\alpha_i^*y_ix_i, \tag{11}$$

$$b^* = -\frac{1}{2}\Big(\frac{1}{|S_+|}\sum_{x\in S_+}\sum_{j=1}^{N}\alpha_j^* y_j x_j^{\mathrm{T}} M^{-1} x + \frac{1}{|S_-|}\sum_{x\in S_-}\sum_{j=1}^{N}\alpha_j^* y_j x_j^{\mathrm{T}} M^{-1} x\Big). \tag{12}$$

其中

$S_{\pm} = \{x_i|0 \leqslant \alpha_i \leqslant C,\ y_i = \pm 1\}$, $N_{sv} = |S_+| + |S_-|$, $|\cdot|$ 表示集合基数.

从而, 为了测试一个新模式 $x \in X$ 的类别, v-LPMIVSVM 的决策函数为

$$g(x) = \mathrm{sign}\Big(\sum_{i=1}^{N}\alpha_i^* y_i x_i^{\mathrm{T}} M^{-1} x + b^*\Big). \tag{13}$$

为了处理非线性分类情况, 一般采用核映射技术[15], 即引入一个非线性映射 ϕ, 将输入空间映射到高维甚至无限维特征空间 H 中, 实现模式线性可分, 高维特征空间的线性超平面对应原始输入空间的非线性超平面. 在 H 空间中, 两个向量 $\phi(x_i)$, $\phi(x_j)$ 的内积可利用一个满足 Mercer 条件的核函数然 $K(x_i, x_j) = \phi(x_i)\cdot\phi(x_j)$ 来表示计算.

需要说明的是, 局部学习信息矩阵 M 可能为奇异矩阵, 即所谓的小样本问题, 因此 M 不可逆. 对于这种情况处理方法较多, 本文按照文献 [3] 中的方法, 利用主成分分析 (PCA 或 KPCA)[8]方法将输入空间数据进行降维, 从而使原始问题转化为一个在低维空间的等价优化问题, 避免了局部保留最大信息矩阵 M 的奇异问题, 从而使矩阵 M 可逆.

v-LPMIVSVM 算法由于要计算局部学习信息矩阵 M, 使 v-LPMIVSVM 与传统的 v-SVM 方法[14]相比具有较高的空间复杂度 ($O(d^2)$) 和时间复杂度 ($O(d^3)$), 特别是在处理高维小样本数据时尤为明显. 为了在一定程度上提高本文方法的执行效率, 在训练高维数据时, 首先采用 PCA(或 KPCA) 方法对数据进行相应的预处理, 以提高所提方法的执行效率.

3　与相关方法比较分析

v-LPMIVSVM 具有与 MCVSVM 等方法相似的优化形式, 且具有与 MCVSVM 和 MCLPVSVM 等方法相同的某些属性. 但与 MCVSVM 相比, v-LPMIVSVM 侧重考虑样本空间的局部流形信息, 而本质流形信息有利于增强模式分类性能; 与 MCLPVSVM 相比, v-LPMIVSVM 同时考虑了数据的局部内流形信息和局部间流形信息. 另外, v-LPMIVSVM 具有和 v-SVM 相同的优化问题, 即 v-LPMIVSVM 问题可通过现有的 SVM 软件包来求解.

定理 2　设有 N 个训练样本 $x_1, x_2, \cdots, x_N$, 分别属于 C 个类, 且每个类包含相同的样本数 n, 令 $W = 1_N 1_N^{\mathrm{T}}$(即式 (1) 中热核函数参数 $t \to +\infty$), 在理想的聚类情况下, v-LPMIVSVM 中各样本的局部 k 近邻数正好等于属于同一类的样本数, 即 $k = n$. 在此种情况下, v-LPMIVSVM 等同于 MCVSVM 问题.

从定理 2 可以推断, MCVSVM 是 v-LPMIVSVM 的一种特例, 即 v-LPMIVSVM 方法是 MCVSVM 方法的一个推广. 有关 MCVSVM 的详细信息可参考文献 [2].

v-LPMIVSVM 和 MCLPVSVM 一样均是考虑了数据空间的局部 (或类内) 本质流形结构信息, 最大区别在于前者同时考虑了数据空间的局部内信息和局部间 (或类间) 信息, 且在度量数据间相似性时, v-LPMIVSVM 采取测地线距离度量, 在高维流形空间测地线距离度量比欧氏距离度量更能反映数据的本质信息[12]. 同时, 针对 MCLPVSVM 可能会产生大量支持向量从而降低泛化性能的问题[14], v-LPMIVSVM 引入了一个能同时控制间隔误差和支持向量的参数 μ[14,16], 从而能进一步提升分类泛化能力. 当 v-LPMIVSVM 中参数 $\lambda = 1$ 且当 v-LPMIVSVM 中最近邻数 k 等于属于同一类的模式数时, v-LPMIVSVM 等同于 MCLPVSVM.

4　实验结论

为了比较分析所提出方法 v-LPMIVSVM 与相关方法 v-SVM, MCVSVM 和 MCLPVSVM 的性能差异, 首先, 通过一簇 UCI 数据集来评价所提出方法在线性和非线性情况下的性能优势; 然后, 通过两个人脸识别实验进一步分析该方法在非线性小样本情况下的分类性能.

4.1　实验设置

根据算法 1, v-LPMIVSVM 的应用需要事先确定 4 个参数 $\{t, k, \lambda, \mu, C\}$. 其中: t 为热核函数参数, λ 为平衡局部散度与局部间散度信息, k 为各样本最近邻数, C 为分类器正则因子. 实验中, 热核参数 t 优化搜索区域为 $\{0.5, 1, 1.5, 2, 2.5, 5, 10, 15, 20\}$. 根据文献 [4], 平衡参数 λ 设置为

$$\lambda = 2^a/4.5\frac{\lambda_{\max}(H_L)}{\lambda_{\max}(H_T)}.$$

其中: $\lambda_{\max}(A)$ 为矩阵 A 的最大特征值, a 的优化搜索范围为 $\{-20, -19, \cdots, 0, \cdots, 19, 20\}$. 参数 μ 在区域 $\{0.01{*}n,\ 0.1{*}n\}$ 中搜索选取, 其中 n 为 1~10 间整数. 由文献 [14] 可知, 在 SVM 中引入 v 参数后, SVM 中正则参数 C 可定为一个固定常数, 由此本文确定 $C = 1/N$, 其中 N 为总体样本数. 数据集最近邻数 k 取值搜索范围为 $\{5, 10, 15, \cdots, 100\}$.

对于其他相关比较方法的参数确定将采取文献 [3] 中相同的策略, 即根据最好的交叉验证参数集的平均值来确定优化的实验参数. 对于 v-SVM,

参数 v 从区域 $\{0.01*n, 0.1*n\}$ 中搜索选取, 其中 n 为 1～10 间整数; 对于 MCVSVM 和 MCLPVSVM, 正则参数 C 从 {0.001, 0.01, 0.1, 1, 10, 100} 中搜索选取; MCLPVSVM 方法中热核参数 t 从区域 {0.5, 1, 1.5, 2, 2.5, 5, 10, 20} 中搜索选取.

4.2 UCI 数据集实验

为了评价本文方法在线性模式分类情况下的泛化性能, 下面实验将选取 14 个 UCI 实际数据集作为评测数据[16], 数据集详细信息如表 1 所示.

表 1 实验中所采用数据集

数据集	模式数	特征数	类数
Iris-setosa	150	4	3
Breast	699	15	2
Heart Healthy	303	13	2
vehicle	846	8	4
Glass	214	9	6
Ionosphere	351	34	2
Wine	178	13	4
Waveform 0	900	21	3
Balance-scale left	625	4	3
Sonar mines	208	60	2
Hepatitis	155	19	2
Biomed	194	5	2
Diabetes	768	8	2
Liver	345	6	2

评价所有分类机的泛化性能的标准是基于数据集总体样本的 5 重交叉验证精度, 在 5 重交叉验证测试中, 数据集被随机划分为 5 个子集, 每次验证取这 5 个子集中的1份作为测试集, 其他作为训练集. 该过程重复 5 次, 取其平均值作为最后的实验结果.

本文记录 5 重交叉验证的平均精度和标准差, 表 2 为本文方法 v-LPMIVSVM 与 v-SVM, MCVSVM 及 MCLPVSVM 在给定数据集上的模式分类精度和标准差. 需要指出的是, 这里仅记录了在最优参数下的实验结果值. 从表 2 可知, 与相关方法相比, 本文方法对于所有数据集均具有优于或可比较的模式分类性能. 同时也可看出, MCVSVM 和 MCLPVSVM 比 v-SVM 具有更好或可比较的泛化性能. 另外, 在考虑了数据空间的本质流形结构的情况下, MCLPVSVM 和 v-LPMIVSVM 表现出明显的泛化能力优势, 尤其是在同时考虑类内流形结构和类间流形结构的情况下, 大间隔分类器 v-LPMIVSVM 可表现出最优的泛化性能.

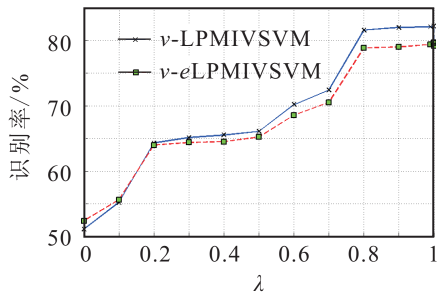

图 1 不同 λ 值下的 Yale 识别率

4.3 脸识别实验

本实验将评价本文方法在小样本非线性模式分类情况下的泛化性能, 分别采用 Yale 人脸数据库和 PIE 人脸数据库[10-11]作为实验数据. Yale 人脸数据库包括 15 张人脸的 165 个灰度级图像, 这些图像分别显示了不同光照条件和脸部表情. 实验中, 对图像定位进行了预处理, 使其缩放到 32×32 像素大小, 且每个像素为 256 灰度级, 则在图像空间, 每张图像由一个 1 024 维向量表示, 更多信息可参考文献 [9]. PIE 人脸数据库包含 68 位志愿者的 41,368 张多姿态、光照和表情的面部图像. 实验中, 对数据库中图像进行预处理, 使其缩放到 32×32 像素大小, 且每个像素为 256 灰度级, 则在图像空间, 每张图像由一个 1 024 维向量表示, 更多信息可参考文献 [11].

表 2 UCI 数据集平均分类精度 %

序号	数据集	v-SVM Acc. Rate (v)	MCVSVM Acc. Rate (C)	MCLPVSVM Acc. Rate (C, t)	v-LPMIVSVM Acc. Rate (μ, k, t, λ)
1	Iris-setosa	97.3±0.12 (0.4)	98.7±0.02 (0.9)	98.6±0.03 (1, 5)	**100±1.17 (0.4,12,5,0.999)**
2	Breast	97.12±1.12 (0.6)	96.94±0.04 (100)	**97.2±0.02 (100,0.2)**	**97.2±0.12 (0.7,10,0.9999)**
3	Heart healthy	79.9±0.045 (0.4)	83.4±0.037 (10)	84.3±0.032 (1, 2)	**87.4±0.017 (0.4,15,20,0.999)**
4	vehicle	76.7±0.014 (0.5)	80.5±0.015 (1)	82.1±0.015 (100, 5)	**83.3±0.13 (0.7,10,0.89)**
5	glass	63.2±0.04 (0.6)	60.5±0.075 (10)	63.57±0.065 (100, 2)	**72.9±1.23 (0.3,20,20,0.999)**
6	Ionosphere	87.67±0.035 (0.5)	83.6±0.045 (100)	89.8±0.44 (100, 0.8)	**91.42±0.046 (0.35,15,20,0.999)**
7	Wine	93.2±0..044 （0.4）	93.48±0.0.34 （100）	95.51±0.0.36 （100,5）	**98.9±0.034 （0.4,12,10,0.999）**
8	Wave form 0	88.7±0.052 (0.4)	86.44±0.421 （80）	85.17±0.047 (10,10)	**90.9±0.17 (0.4,5,10,0.999)**
9	Balance scale left	**94.2±2.232 (0.4)**	93.9±0.023 （100）	93.67±0.033 (100,10)	**94.2±1.37 (0.4,7,3,0.9999)**
10	Sonar mines	69.9±0.012 (0.5)	75.45±0.412 (100)	76.26±1.131 (10,3)	79.6 (0.3,15,20,0.999)
11	Hepatitis	84.41±0.042 (0.4)	**87.01±0.013 (10)**	85.46±0.054 (100,10)	**87.01±0.073 (0.4,15,10,0.999)**
12	Biomed	86.9±0.072 (0.4)	87.5±0.014 (10)	90.82±0.074 (100,5)	**91.67±0.062 (0.3,15,8,0.999)**
13	Diabetes	**76.2±0.064 (0.4)**	75.26 (100)	75.67±2.054 (100,5)	75.8±0.420 (0.3,20,20,0.9999)
14	Liver	65.7±3.21 (0.3)	71.5±0.016 (100)	71.84±0.018 (100,3.5)	**72.67±0.172 (0.3,7,6,0.999)**

表 3　在 Yale 数据库上 5 重交叉验证实验平均精度和标准偏差　　%

维数	v-SVM	MCVSVM	MCLPVSVM	v-eLPMIVSVM	v-LPMIVSVM
	Acc. Rate (v)	Acc. Rate (C)	Acc. Rate (C, t)	Acc. Rate (μ, t, λ)	Acc. Rate (μ, t, λ)
3	61.78±0.02 (0.5)	62.45±0.055 (100)	**63.78±0.043 (100, 5)**	62.67±0.012 (0.4,5,0.99)	62.47±0.032 (0.4,5,0.99)
6	73.36±0.024 (0.4)	**75.44±0.13 (1)**	74.45±0.11 (10, 20)	74.32±1.12 (0.5,15,0.999)	74.26±0.046 (0.6,15,0.999)
9	76.3±0.12 (0.6)	75.94±0.089 (100)	76.9±0.145 (1,5)	76.9±0.062 (0.7,1,0.899)	**77.4±1.15 (0.7,5,0.9)**
12	75.62±0.145 (0.4)	77.64±0.131 (100)	79.65±0.112 (10, 5)	79.68±0.17 (0.4,15,0.9)	**81.24±1.07 (0.4,15,0.99)**
15	77.23±0.014 (0.3)	76.23±0.115 (100)	77.211±0.13 (10, 5)	78.11±0.031 (0.6,10,1)	**80.62±0.23 (0.6,10,0.999)**
$N-2$	76.2±0.14 (0.2)	77.43±0.143 (100)	78.67±0.132 (10, 5)	80.23±0.012 (0.3,5,0.999)	**82.23±0.13 (0.3,10,0.999)**
All	**76.72±0.142 (0.3)**	–	–	–	–

为了处理高维数据的非线性分类问题, 实验中所采用的典型核函数为高斯核函数 $\exp(-(u-v)^{\rm T}(u-v)/2\sigma^2)$, 其中 σ 为高斯核函数的幅度, 实验中令 $\sigma=10$. 此时经该核函数高维映射后, v-LPMIVSVM 中局部内散度矩阵 H_L, MCLPVSVM 中局部保留类内散度矩阵 Z_w 和 MCVSVM 中类内散度矩阵 S_w 均为奇异型. 因此, 先利用 PCA 或 KPCA 将原始空间数据投影到某个子空间, 使上述奇异型矩阵变成非奇异型. 为了对各方法进行有效比较, 实验前将数据降维到同样维度. 实验结果取自 5 重交叉验证实验的平均精度和标准差值. 实验中为了比较两种距离度量方式 (欧氏距离和测地线距离) 对本文方法的泛化性能影响, 并且为了区别起见, 本文称欧氏距离度量的 v-LPMIVSVM 为 v-eLPMIVSVM.

表 3 显示了不同维度下几种分类方法与本文方法在 Yale 数据集上的识别精度比较. 由于训练样本间线性无关, 可将样本投射到 N-2 维空间, 其中 N 为训练样本数. v-SVM 在未经降维的样本上的实验结果记录于表 3 底部.

从表 3 可看出, 相比其他方法, 基于测地线距离度量的 v-LPMIVSVM 在所有维度下均具有优于或可比较的识别性能. 随着维度的增大, 所有方法的识别率均上升, 直到达到某个优化维度值 (如 $d=12$). 在低维段, MCLPVSVM 方法略优于其他方法, 在高维段, 本文方法 v-LPMIVSVM 的识别率明显高于其他方法. 另外也可以看出, 基于测地线距离度量的 v-LPMIVSVM 方法优于基于欧氏距离度量的 v-eLPMIVSVM 方法. 图 1 显示了不同 λ 参数值下, v-eLPMIVSVM 和 v-LPMIVSVM 两种方法的 Yale 人脸识别率比较, 其中数据维度为 12. 由图 1 可看出, v-LPMIVSVM 的识别性能优于 v-eLPMIVSVM. 当 $\lambda=0$ 时, 即为忽略局部流形内散度信息时, 识别率最低, 随着 λ 逐渐增加, 两种方法识别率均上升; 但当 $\lambda=1$ 时, 即忽略数据流形局部间散度信息时, v-LPMIVSVM 和 v-eLPMIVSVM 的识别率均下降. 从图 1 又可看出, 优化的 λ 取值区域为 (0.8, 1).

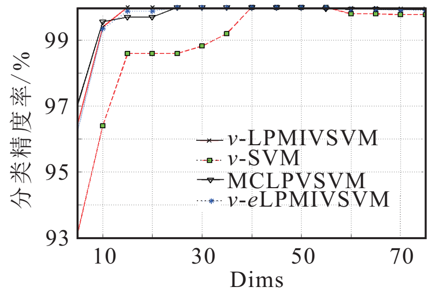

图 2　不同数据维数下的 PIE 识别率

图 2 显示了 4 种方法在不同数据维数下的 PIE 人脸识别精度比较. 不论在低维段还是在高维段, 本文方法 v-LPMIVSVM 均表现出优于或可比较的识别精度. 图 3 显示了在最近邻数 k 不同的情况下, 3 种方法的 PIE 人脸识别率, 其中数据维度固定为 30. 从图 3 可看出, k 值的变化对 MCLPVSVM 方法的识别率无影响, 与文献 [3] 中的结论一致. 在 k 值相对较小时, 本文方法 v-LPMIVSVM 能达到最优识别率, 而 v-eLPMIVSVM 方法在 k 值较大时才能达到优化识别率. 随着 k 值进一步增大, v-LPMIVSVM 对 PIE 人脸识别率呈下降趋势. 这是因为随着最近邻数 k 值的增大, 最短路径已不能很好地近似计算测地线距离, 从而导致对数据的本质几何流形结构的估计误差增大.

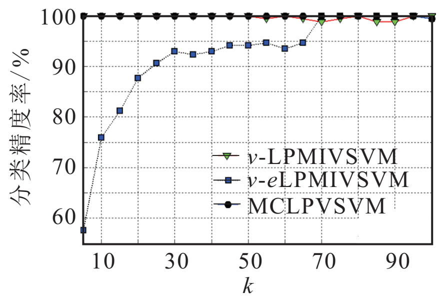

图 3　不同最近邻数下的 PIE 识别率

5　结　论

本文提出一种新的局部学习支持向量机 v-LPMIVSVM, 它充分考虑了数据空间局部内和局部间本质几何信息. 在线性和非线性两种实际数据集上的实验验证了所提出方法具有优于或等同于相关方法的泛化性能. 在对待小样本问题上, 尽管 v-LPMIVSVM 采用 PCA 或 KPCA 对原始输入数据进行

降维处理来达到解决问题的目的, 但许多随之而产生的问题 (如维数的优化、信息缺失等) 有待进一步解决.

参考文献(References)

[1] 文传军, 詹永照, 等. 最大间隔最小体积球形支持向量机[J]. 控制与决策, 2010, 25(1): 79-83.
(Wen C J, Zhan Y Z, et al. Maximal-margin minimal-volume hypersphere support vector machine[J]. Control and Decision, 2010, 25(1): 79-83.)

[2] Vapnik V. The nature of statistical learning theory[M]. Springer Verlag, 1995.

[3] Xiaoming Wang, Fu-lai Chung, Shitong Wang. On minimum class locality preserving variance support vector machine[J]. Pattern Recognition, 43 (2010): 2753-2762.

[4] Haixian Wang, Sibao Chen, et al. Locality-preserved maximum information projection[J]. IEEE Trans on Neural Networks, 2008, 19(4): .

[5] Liu Q, Tang X, Lu H, et al. Face recognition using kernel scatter-difference-based discriminant analysis[J]. IEEE Trans on Neural Network, 2006, 17(4): 1081-1085.

[6] Shivaswamy P, Jebara T. Maximum relative margin and data-dependent regularization[J]. J of Machine Learning Research, 2010, 11: 747-788.

[7] Zafeiriou S, Tefas A, Pitas I. Minimum class variance support vector machines[J]. IEEE Trans on Image Processing, 2007, 16(10): 2551-2564.

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

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

[10] He X, Niyogi P. Locality preserving projections[C]. Proc of the Conf on Advances in Neural Information Processing Systems. 2003: 585-591.

[11] Cai D, He X, Han J, et al. Orthogonal Laplacianfaces for face recognition[J]. IEEE Trans on Image Processing, 2006, 15(11): 3608-3614.

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

[13] 高全学, 谢德燕, 等. 融合局部结构和差异信息的监督特征提取算法[J]. 自动化学报, 2010, 36(8): 1107-1114.
(Gao Q X, Xie D Y, Xu H, et al. Supervised feature extraction based on information fusion of local structure and diversity information[J]. Acta Automatica Sinica, 2010, 36(8): 1107-1114.)

[14] Schölkopf B, Smola A J, Williamson R, et al. New support vector algorithms[J]. Neural Computation, 2000, 12(1): 1207-1245.

[15] Müller K R, Mika S, Ratsch G, et al. An introduction to kernel-based learning algorithms[J]. IEEE Trans on Neural Networks, 2001, 12(2): 181-201.

[16] Mingrui Wu, Jieping Ye. A small sphere and large margin approach for novelty detection using training data with outliers[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2009, 31(11): 2088-2092.