

边界监督多重集典型相关分析

杨 静¹ 高希占²

(1.聊城大学 数学科学学院,山东 聊城 252059;2.南京理工大学 计算机科学与工程学院,江苏 南京 210094)

摘 要 多重集典型相关分析(multiset canonical correlation analysis, MCCA)仅仅考虑了多组数据间的相关性信息,不能有效地反映样本数据的几何结构与鉴别信息,因此为了解决这个问题,首先在 LDA 思想的启发下,构建了监督多重集典型相关分析(supervised multiset canonical correlation analysis, SMCC)的理论框架,并以此为基础,结合边界 Fisher 分析(marginal fisher analysis, MFA),提出了边界监督多重集典型相关分析(marginal SMCC, MSMCC).该算法的基本思想是在最大化数据相关性的同时,还要最大化组内数据的类间离散度以及最小化组内数据的类内离散度.在人脸图像与目标数据库上的实验结果验证了所提算法的有效性.

关键词 典型相关分析;多重集典型相关分析;特征抽取;降维;监督学习

中图分类号 TP391

文献标识码 A

0 引言

典型相关分析(Canonical Correlation Analysis, CCA)^[1]由 H. Hotelling 于 1936 年首次提出,是一种重要的多元统计分析方法. CCA 主要研究两组随机变量之间的相关性. 目前, CCA 已在信号处理、数据可视化、医学数据分析、模式识别及计算机视觉领域得到了广泛的应用,成为研究的热点.

从模式识别的角度, Sun 等^[2,3]提出了基于 CCA 理论的特征融合方法,即首先抽取相同模式的两组特征,并在此基础上建立了两组特征间的相关性判别准则,然后依此准则对其进行融合,抽取出组合的典型相关特征并应用于分类. 同时,在小样本问题中,他们首先使用 PCA 将高维数据降维到一个低维子空间,在不损失信息的情况下解决协方差矩阵奇异的问题,接着在该子空间中使用 CCA 进一步进行特征抽取. 随后, Carlos 等^[4]提出了正则化核 CCA,该方法将支持向量机(Support vector machine, SVM)^[5,6]引入到了核 CCA 中. 由于 CCA 本质上为无监督的学习方法,为了引入监督信息, Sun 等^[7]利用类内信息给出了广义典型相关分析(Generalized CCA, GCCA)方法. GCCA 将最小化类内散布矩阵作为目标之一,极大地减少了样本数据的类内离散度,进而提高了低维特征的鉴别能力. 随后,判别型 CCA(Discriminative CCA, DCCA)^[8]将样本间相关作为相似性度量,引入了样本的鉴别(类别)信息,充分考虑了同类样本之间的最大相关与不同类样本之间的最小相关及其对分类的影响. 尽管 GCCA 获得了比 CCA 好的识别结果,但在特征抽取的过程中,训练样本往往受到噪声等条件的影响,这些影响可能会使各类之间的界限变得模糊,并且使得一些样本位于或接近类之间的重叠区域,样本分布的这一特点将会损失对分类有用的信息,因此为了解决这一问题,文献^[9]提出了基于隶属度的广义典型相关分析(Membership-degree GCCA, MGCCA). 此外,由于 DCCA 抽取的最大特征维数受到类别数的限制,可能出现投影轴数不足的问题,因而在一定程度上限制了算法的应用. 在线性判别分析(Linear discriminant analysis, LDA)^[10]思想的启发下, Kim 等^[11]在图像集分类中提出了鉴别典型相关(Discriminant-analysis of Canonical Correlation, DCC),该方法从线性特征抽取的角度对图像集子空间进行了维数压缩.

收稿日期:2018-12-10

基金项目:国家自然科学基金项目(11801248);山东省自然科学基金项目(ZR2018BF010)资助

通讯作者:高希占,男,汉族,博士,研究方向:模式识别,机器学习理论与应用, E-mail: gaoxizhan123@126.com.

由于 CCA 是研究两组变量的多元统计方法,因此上述基于 CCA 的方法在多表示数据(多于两组数据)的特征抽取与融合中很难获得满意的效果,多重集典型相关分析(Multiset canonical correlation analysis, MC-CA)^[12]作为研究多组随机矢量间相关性的方法,非常适于多表示数据的特征抽取与融合. MCCA 是对 CCA 方法的一种广义化扩展,其本质上是寻找一组线性变换,在尽量保持原始数据蕴含的有效信息下,将高维空间中的数据压缩到一个低维子空间中,然后在多组高维数据的低维表示中抽取相关信息. 由于其能够较好地反映多个数据集间的相关结构与内在联系,因此, MCCA 方法在多组高维数据的相关分析中具有重要的地位. 近年来, MCCA 已被成功地应用于盲源分离^[13]、目标识别^[14-16]、遥感图像分析^[12]以及 fMRI 数据分析^[17-19]等许多科学领域.

然而,由于 MCCA 是一种无监督的子空间学习方法,在其模型中并没有包含数据空间中的判别信息,这将不利于数据的分类. 为解决这个问题, Gao 等^[20]在 DCCA 思想的启发下,提出了判别型 MCCA,该方法将样本间相关作为相似性度量,引入了样本的类别信息,充分考虑了同类样本之间的最大相关与不同类样本之间的最小相关关系及其对分类的影响. 随后, Lee 等^[21]提出了监督多视图典型相关分析(Supervised multi-view canonical correlation analysis, SMVCCA),该算法引入了样本的标签信息,并能够同时将模式之间的关系以及标签信息关联起来,进而生成一个更能区分对象类的低维表示. 尽管上述基于 MCCA 的方法引入了标签信息并通过对高维多表示数据的联合特征抽取,提高了特征表示的鉴别力与鲁棒性,增强了算法的稳定性. 然而,由于它们仅仅考虑了多组高维数据间的相关结构,因此在实际应用中并不能反映数据的本质几何结构与分布信息. 基于该问题及在 LDA、边界 Fisher 分析(Marginal Fisher Analysis, MFA)^[22]思想的启发,本文提出了一种新的监督的联合特征抽取方法,即边界监督多重集典型相关分析(Marginal supervised MC-CA, MSMCC),该方法在极大化数据相关性的同时,还最大化了组内数据的类间散布与最小化了组内数据的类内散布,并在保持样本数据相关性的同时还反映了样本数据的本质几何结构与分布信息. 在 COIL 数据库以及人脸图像数据库上的实验表明,该方法具有良好的识别性能,并揭示出其具有较强的鉴别能力.

本文第 1 节对 MCCA 作简单介绍. 第 2 节提出监督 MCCA. 第 3 节具体介绍 MSMCC 算法. 第 4 节通过实验对所提算法进行性能测试,并对实验结果进行分析. 最后对本文的工作进行总结.

1 多重集典型相关分析

多重集典型相关分析主要研究多组(两组以上)随机变量间的相关性,其本质上是 CCA 的一种泛化扩展. MCCA 有多种不同的形式,如下的一种形式在信号处理和目标识别等领域内获得了广泛的应用.

给定 m 组随机变量 $\mathbf{x}_i \in \mathbf{R}^{q_i}$, $i = 1, 2, \dots, m$, 且 $E(\mathbf{x}_i) = 0$, 其投影方向集为 $\{\boldsymbol{\alpha}_i \in \mathbf{R}^{q_i}\}_{i=1}^m$, 则 MCCA 的最优化模型为

$$\begin{aligned} \max \rho(\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_m) &= \sum_{i=1}^m \sum_{j=1}^m \boldsymbol{\alpha}_i^T \mathbf{S}_{ij} \boldsymbol{\alpha}_j, \\ \text{s. t. } \sum_{j=1}^m \boldsymbol{\alpha}_i^T \mathbf{S}_{ij} \boldsymbol{\alpha}_i &= 1, \end{aligned} \quad (1)$$

其中 \mathbf{S}_{ii} 表示随机向量 \mathbf{x}_i 的组内协方差矩阵, \mathbf{S}_{ij} ($i \neq j$) 表示随机向量 \mathbf{x}_i 和 \mathbf{x}_j 的组间协方差矩阵, 并且 $\boldsymbol{\alpha}^T = (\boldsymbol{\alpha}_1^T, \boldsymbol{\alpha}_2^T, \dots, \boldsymbol{\alpha}_m^T)$.

通过引入 Lagrange 乘子,模型(1)可以转化为如下的问题

$$\begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} & \cdots & \mathbf{S}_{1m} \\ \mathbf{S}_{21} & \mathbf{S}_{22} & \cdots & \mathbf{S}_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}_{m1} & \mathbf{S}_{m2} & \cdots & \mathbf{S}_{mm} \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \vdots \\ \boldsymbol{\alpha}_m \end{pmatrix} = \lambda \begin{pmatrix} \mathbf{S}_{11} & & & \\ & \mathbf{S}_{22} & & \\ & & \ddots & \\ & & & \mathbf{S}_{mm} \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \vdots \\ \boldsymbol{\alpha}_m \end{pmatrix}, \quad (2)$$

其中 λ 为 Lagrange 乘子,等式(2)为广义特征值问题. 为了使优化模型(1)中的目标函数能够取得最大值,选取前 d 个最大特征值对应的特征向量作为等式(2)的解向量.

2 监督多重集典型相关分析

MCCA 在本质上是 CCA 的一种广义化扩展,它能够较好的反映多个特征数据集间的相关结构与内在联系,但由于其在特征抽取的过程中,并没有使用数据中的鉴别信息(如类别信息),因此在模式分类中并不能获得很满意的识别性能.为了充分利用类标签信息,本节提出了监督多重集典型相关分析(SMCC).

具体来说,假设 $\omega_1, \omega_2, \dots, \omega_c$ 为 c 个已知的模式类,给定来自于这 c 个模式类的 m 组高维数据 $\{\mathbf{X}^{(i)} \in \mathbf{R}^{p_i \times n}\}_{i=1}^m$,其中 $\mathbf{X}^{(i)} = (\mathbf{x}_1^{(i)}, \mathbf{x}_2^{(i)}, \dots, \mathbf{x}_n^{(i)})$, n 为样本数, p_i 为 $\mathbf{X}^{(i)}$ 的维数, $i = 1, 2, \dots, m$. 假设样本集 $\{\mathbf{X}^{(i)} \in \mathbf{R}^{p_i \times n}\}_{i=1}^m$ 均已被中心化,设随机矢量 $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \dots, \mathbf{X}^{(m)}$ 通过投影方向 $\mathbf{V} = \{\alpha_i \in \mathbf{R}^{p_i}\}_{i=1}^m$ 作线性投影后得到的随机矢量为 $\mathbf{Y}^{(i)} = \alpha_i^T \mathbf{X}^{(i)}, i = 1, 2, \dots, m$.

LDA 的目标是寻找一投影方向,使得原始样本投影后,同类样本之间相互靠近而不同类样本之间相互分离.一般情况下,最优的投影方向 \mathbf{W} 是使样本类间散布矩阵与类内散布矩阵的迹比值最大的那些特征向量.因此, LDA 的判别准则函数可定义为

$$J_{LDA} = \operatorname{argmax} \frac{\operatorname{tr}(\mathbf{W}^T \mathbf{S}_b \mathbf{W})}{\operatorname{tr}(\mathbf{W}^T \mathbf{S}_w \mathbf{W})}, \quad (3)$$

其中 $\operatorname{tr}(\cdot)$ 表示矩阵的迹,通过求解上述准则函数可获得最优投影方向 \mathbf{W} .

结合 MCCA 与 LDA 算法的基本思想,提出了监督的多重集典型相关分析算法. SMCC 方法在极大化数据相关性的同时,还最大化了组内数据的类间散布与最小化了组内数据的类内散布.其目标函数如下所示

$$J_{SMCC}(\alpha_1, \alpha_2, \dots, \alpha_m) = \frac{\sum_{i=1}^m \sum_{j=1}^m \alpha_i^T \mathbf{S}_{ij} \alpha_j}{\sqrt{\sum_{i=1}^m \alpha_i^T \mathbf{S}_i^w \alpha_i}}, \quad (4)$$

其中 $\{\alpha_i \in \mathbf{R}^{p_i}\}_{i=1}^m$ 分别表示随机矢量 $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \dots, \mathbf{X}^{(m)}$ 的投影方向, \mathbf{S}_i^w 表示 $\mathbf{X}^{(i)}$ 的类内散布矩阵, $\mathbf{S}_{ij} = \begin{cases} \operatorname{cov}(\mathbf{S}_i^b, \mathbf{S}_j^b), & i \neq j \\ \mathbf{S}_i^b, & i = j \end{cases}$, $i, j \in m$, $\operatorname{cov}(\cdot, \cdot)$ 表示协方差, \mathbf{S}_i^b 为 $\mathbf{X}^{(i)}$ 的类间散布矩阵.

从 SMCC 的目标函数可以看出,其投影方向不仅可以最大化数据间的相关性,而且还可以使组内数据的类内散布尽可能地小同时使其类间散布尽可能地大.

式(4)对应的准则函数的求解可以表述为如下的最优化模型

$$\begin{aligned} \max J_{SMCC}(\alpha_1, \alpha_2, \dots, \alpha_m) &= \sum_{i=1}^m \sum_{j=1}^m \alpha_i^T \mathbf{S}_{ij} \alpha_j, \\ \text{s. t. } \sum_{i=1}^m \alpha_i^T \mathbf{S}_i^w \alpha_i &= 1. \end{aligned} \quad (5)$$

对模型(5)利用 Lagrange 乘子法进行优化求解,可得到如下 Lagrange 函数

$$L(\alpha_1, \alpha_2, \dots, \alpha_m) = \sum_{i=1}^m \sum_{j=1}^m \alpha_i^T \mathbf{S}_{ij} \alpha_j - \lambda \left(\sum_{i=1}^m \alpha_i^T \mathbf{S}_i^w \alpha_i - 1 \right). \quad (6)$$

再对 $L(\alpha_1, \alpha_2, \dots, \alpha_m)$ 分别求取 $\alpha_1, \alpha_2, \dots, \alpha_m$ 的偏导数并令其为零,得到

$$\sum_{j=1}^m \mathbf{S}_{ij} \alpha_j = \lambda \mathbf{S}_i^w \alpha_i, \quad i = 1, 2, \dots, m. \quad (7)$$

进一步对式(7)作简化处理,则可获得如下的广义特征值问题

$$\begin{pmatrix} \mathbf{S}_1^b & \mathbf{S}_{12} & \cdots & \mathbf{S}_{1m} \\ \mathbf{S}_{21} & \mathbf{S}_2^b & \cdots & \mathbf{S}_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}_{m1} & \mathbf{S}_{m2} & \cdots & \mathbf{S}_m^b \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{pmatrix} = \lambda \begin{pmatrix} \mathbf{S}_1^w & & & \\ & \mathbf{S}_2^w & & \\ & & \ddots & \\ & & & \mathbf{S}_m^w \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{pmatrix}. \quad (8)$$

通过求解式(8)中的广义特征值问题,选取前 d 个最大特征值 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$ 对应的特征向量 $\alpha_1, \alpha_2, \dots, \alpha_d$ 作为 SMCC 的投影方向,记为 $\{\mathbf{W}_i = (\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{id})\}_{i=1}^m$. 对于 m 组高维数据 $\{\mathbf{X}^{(i)}\}_{i=1}^m$, 利用 $\{\mathbf{W}_i^T \mathbf{X}^{(i)}\}_{i=1}^m$ 实现联合特征抽取.

3 基于边界的监督多重集典型相关分析

3.1 局部类内散布矩阵的构造

对于 $\mathbf{X}^{(i)}$ 中的 n 个样本点 $\mathbf{x}_1^{(i)}, \mathbf{x}_2^{(i)}, \dots, \mathbf{x}_n^{(i)}$, 构建其类内图 $\mathbf{G}_w^{(i)} = [\mathbf{X}^{(i)}, \mathbf{W}_w^{(i)}]$, 其中 $\mathbf{G}_w^{(i)}$ 的每一个顶点对应于一个样本点 $\mathbf{x}_j^{(i)}$, 并且对于每个样本点 $\mathbf{x}_j^{(i)}$, 在相同的类中找到它的 k_1 最近邻点并分别将 $\mathbf{x}_j^{(i)}$ 与其近邻点相连, 同时分配相应的权重, 进而构建出近邻图 $\mathbf{G}_w^{(i)}$. 通过图的构建过程可以看出, $\mathbf{G}_w^{(i)}$ 能够有效地反映出数据的局部几何结构与分布信息. 在获得 n 个样本点 $\mathbf{x}_1^{(i)}, \mathbf{x}_2^{(i)}, \dots, \mathbf{x}_n^{(i)}$ 在投影轴 $\boldsymbol{\alpha}_i \in \mathbf{R}^{p_i}$ 上的投影 $\mathbf{y}_1^{(i)}, \mathbf{y}_2^{(i)}, \dots, \mathbf{y}_n^{(i)}$ 后, 根据 MFA 的思想, 数据的局部类内离散度能够被定义为

$$J_w^{(i)} = \frac{1}{2} \sum_{j=1}^n \sum_{k=1}^n W_{w,jk}^{(i)} \|\mathbf{y}_j^{(i)} - \mathbf{y}_k^{(i)}\|^2, \quad (9)$$

其中 $W_{w,jk}^{(i)} \in \mathbf{R}$ 表示近邻图 $\mathbf{G}_w^{(i)}$ 的加权矩阵 $\mathbf{W}_w^{(i)} \in \mathbf{R}^{n \times n}$ 中的第 (j, k) 个元素, 并且被定义为

$$W_{w,jk}^{(i)} = \begin{cases} 1, & \mathbf{x}_k^{(i)} \in N_{k_1}^{(i)}(\mathbf{x}_j^{(i)}) \text{ or } \mathbf{x}_j^{(i)} \in N_{k_1}^{(i)}(\mathbf{x}_k^{(i)}), \\ 0, & \text{else,} \end{cases} \quad (10)$$

其中 $N_{k_1}^{(i)}(\mathbf{x}_j^{(i)})$ 表示在相同类中样本点 $\mathbf{x}_j^{(i)}$ 的 k_1 最近邻, 并且加权矩阵 $\mathbf{W}_w^{(i)}$ 是对称的, 即 $\mathbf{W}_w^{(i)} = \mathbf{W}_w^{(i)\top}$.

利用数学推导, 可将式(9)简化为

$$\begin{aligned} J_w^{(i)} &= \frac{1}{2} \sum_{j=1}^n \sum_{k=1}^n W_{w,jk}^{(i)} (\boldsymbol{\alpha}_i^\top \mathbf{x}_j^{(i)} - \boldsymbol{\alpha}_i^\top \mathbf{x}_k^{(i)}) (\boldsymbol{\alpha}_i^\top \mathbf{x}_j^{(i)} - \boldsymbol{\alpha}_i^\top \mathbf{x}_k^{(i)})^\top \\ &= \boldsymbol{\alpha}_i^\top \left[\frac{1}{2} \sum_{j=1}^n \sum_{k=1}^n W_{w,jk}^{(i)} (\mathbf{x}_j^{(i)} - \mathbf{x}_k^{(i)}) (\mathbf{x}_j^{(i)} - \mathbf{x}_k^{(i)})^\top \right] \boldsymbol{\alpha}_i \\ &= \boldsymbol{\alpha}_i^\top \tilde{\mathbf{S}}_w^{(i)} \boldsymbol{\alpha}_i, \end{aligned} \quad (11)$$

其中 $\tilde{\mathbf{S}}_w^{(i)} \in \mathbf{R}^{p_i \times p_i}$ 被称为组内数据的局部类内散布矩阵, 并满足

$$\tilde{\mathbf{S}}_w^{(i)} = \frac{1}{2} \sum_{j=1}^n \sum_{k=1}^n W_{w,jk}^{(i)} (\mathbf{x}_j^{(i)} - \mathbf{x}_k^{(i)}) (\mathbf{x}_j^{(i)} - \mathbf{x}_k^{(i)})^\top, \quad (12)$$

由于矩阵 $\mathbf{W}_w^{(i)}$ 的对称性, 故有

$$\begin{aligned} \tilde{\mathbf{S}}_w^{(i)} &= \sum_{j=1}^n \mathbf{D}_{jj}^{(i)} \mathbf{x}_j^{(i)} \mathbf{x}_j^{(i)\top} - \sum_{j=1}^n \sum_{k=1}^n W_{w,jk}^{(i)} \mathbf{x}_j^{(i)} \mathbf{x}_k^{(i)\top} \\ &= \mathbf{X}^{(i)} \mathbf{D}^{(i)} \mathbf{X}^{(i)\top} - \mathbf{X}^{(i)} \mathbf{W}_w^{(i)} \mathbf{X}^{(i)\top} \\ &= \mathbf{X}^{(i)} \mathbf{L}^{(i)} \mathbf{X}^{(i)\top}, \end{aligned} \quad (13)$$

其中 $\mathbf{D}^{(i)} = \text{diag}(D_{11}^{(i)}, D_{22}^{(i)}, \dots, D_{mm}^{(i)})$, 其第 j 个对角元素是 $\mathbf{W}_w^{(i)}$ 中第 j 行所有元素的和(或是第 j 列所有元素的和), 即 $D_{jj}^{(i)} = \sum_{k=1}^n W_{w,jk}^{(i)}$, $\mathbf{L}^{(i)} = \mathbf{D}^{(i)} - \mathbf{W}_w^{(i)}$ 被称为 Laplacian 矩阵^[23].

3.2 局部类间散布矩阵的构造

对于给定的 n 个样本点 $\mathbf{x}_1^{(i)}, \mathbf{x}_2^{(i)}, \dots, \mathbf{x}_n^{(i)}$, 相似于其类内图 $\mathbf{G}_w^{(i)}$ 的构建, 类间图 $\mathbf{G}_b^{(i)} = [\mathbf{X}^{(i)}, \mathbf{W}_b^{(i)}]$ 的构建如: 对于样本点 $\mathbf{x}_j^{(i)}$, 在其不同的类中找到它的 k_2 最近邻点并分别将 $\mathbf{x}_j^{(i)}$ 与其近邻点相连, 同时分配相应的权重. 通过类间图 $\mathbf{G}_b^{(i)}$ 的构建过程可以看出, $\mathbf{G}_b^{(i)}$ 能够有效地反映出数据的局部几何结构与分布信息. 基于类间图 $\mathbf{G}_b^{(i)}$, 数据的局部类间离散度能够被定义为

$$\begin{aligned} J_b^{(i)} &= \frac{1}{2} \sum_{j=1}^n \sum_{k=1}^n W_{b,jk}^{(i)} \|\mathbf{y}_j^{(i)} - \mathbf{y}_k^{(i)}\|^2 \\ &= \frac{1}{2} \sum_{j=1}^n \sum_{k=1}^n W_{b,jk}^{(i)} (\boldsymbol{\alpha}_i^\top \mathbf{x}_j^{(i)} - \boldsymbol{\alpha}_i^\top \mathbf{x}_k^{(i)}) (\boldsymbol{\alpha}_i^\top \mathbf{x}_j^{(i)} - \boldsymbol{\alpha}_i^\top \mathbf{x}_k^{(i)})^\top \\ &= \boldsymbol{\alpha}_i^\top \left[\frac{1}{2} \sum_{j=1}^n \sum_{k=1}^n W_{b,jk}^{(i)} (\mathbf{x}_j^{(i)} - \mathbf{x}_k^{(i)}) (\mathbf{x}_j^{(i)} - \mathbf{x}_k^{(i)})^\top \right] \boldsymbol{\alpha}_i \\ &= \boldsymbol{\alpha}_i^\top \tilde{\mathbf{S}}_b^{(i)} \boldsymbol{\alpha}_i, \end{aligned} \quad (14)$$

其中 $\tilde{\mathbf{S}}_i^b \in \mathbf{R}^{p_i \times p_i}$ 被称为组内数据的局部类间散布矩阵. 由于矩阵 $\mathbf{W}_b^{(i)}$ 的对称性,有

$$\begin{aligned}\tilde{\mathbf{S}}_i^b &= \frac{1}{2} \sum_{j=1}^n \sum_{k=1}^n \mathbf{W}_{b,jk}^{(i)} (\mathbf{x}_j^{(i)} - \mathbf{x}_k^{(i)}) (\mathbf{x}_j^{(i)} - \mathbf{x}_k^{(i)})^T \\ &= \sum_{j=1}^n \tilde{\mathbf{D}}_{jj}^{(i)} \mathbf{x}_j^{(i)} \mathbf{x}_j^{(i)T} - \sum_{j=1}^n \sum_{k=1}^n \mathbf{W}_{\omega,jk}^{(i)} \mathbf{x}_j^{(i)} \mathbf{x}_k^{(i)T} \\ &= \mathbf{X}^{(i)} \tilde{\mathbf{D}}^{(i)} \mathbf{X}^{(i)T} - \mathbf{X}^{(i)} \mathbf{W}_w^{(i)} \mathbf{X}^{(i)T} \\ &= \mathbf{X}^{(i)} \tilde{\mathbf{L}}^{(i)} \mathbf{X}^{(i)T},\end{aligned}\quad (15)$$

其中 $\tilde{\mathbf{D}}^{(i)} = \text{diag}(\tilde{\mathbf{D}}_{11}^{(i)}, \tilde{\mathbf{D}}_{22}^{(i)}, \dots, \tilde{\mathbf{D}}_m^{(i)})$, 它的第 j 个对角元素是矩阵 $\mathbf{W}_b^{(i)}$ 中第 j 行所有元素的和(或是第 j 列所有元素的和), 即 $\tilde{\mathbf{D}}_{jj}^{(i)} = \sum_{k=1}^n \mathbf{W}_{b,jk}^{(i)}$, $\tilde{\mathbf{L}}^{(i)} = \tilde{\mathbf{D}}^{(i)} - \tilde{\mathbf{W}}_w^{(i)}$, 并且 $\mathbf{W}_{b,jk}^{(i)} \in \mathbf{R}$ 表示近邻图 $\mathbf{G}_b^{(i)}$ 的加权矩阵 $\mathbf{W}_b^{(i)} \in \mathbf{R}^{n \times n}$ 中的第 (j, k) 个元素, 并且被定义为

$$\mathbf{W}_{b,jk}^{(i)} = \begin{cases} 1, & \mathbf{x}_k^{(i)} \in P_{k_2}^{(i)}(\mathbf{x}_j^{(i)}) \text{ or } \mathbf{x}_j^{(i)} \in P_{k_2}^{(i)}(\mathbf{x}_k^{(i)}), \\ 0, & \text{else,} \end{cases}\quad (16)$$

其中 $P_{k_2}^{(i)}(\mathbf{x}_j^{(i)})$ 表示在不同类中样本点 $\mathbf{x}_j^{(i)}$ 的 k_2 最近邻.

3.3 模型构建与求解

尽管 SMCC 在特征抽取过程中成功地利用了样本的类别信息, 但它不能够有效地发现数据空间中的本质几何结构. 因此, 为了能够更好地发掘数据内在的几何结构, 为了刻画多表示数据中的几何结构, 以 MFA 思想为基础, 提出了边界监督多重集典型相关分析 (MSMCC), 其对应的最优化模型如

$$\begin{aligned}\max \rho(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_m) &= \sum_{i=1}^m \sum_{j=1}^m \boldsymbol{\alpha}_i^T \tilde{\mathbf{S}}_{ij} \boldsymbol{\alpha}_j, \\ \text{s. t. } \sum_{i=1}^m \boldsymbol{\alpha}_i^T \tilde{\mathbf{S}}_i^w \boldsymbol{\alpha}_i &= 1,\end{aligned}\quad (17)$$

其中 $\boldsymbol{\alpha}^T = (\boldsymbol{\alpha}_1^T, \boldsymbol{\alpha}_2^T, \dots, \boldsymbol{\alpha}_m^T)$, $\tilde{\mathbf{S}}_{ij} = \begin{cases} \text{cov}(\tilde{\mathbf{S}}_i^b, \tilde{\mathbf{S}}_j^b), & i \neq j, i, j \in m, \\ \tilde{\mathbf{S}}_i^b, & i = j, i, j \in m. \end{cases}$

利用 Lagrange 乘子理论, 最优化模型(17)可以转化为如下的 Lagrange 函数

$$L(\boldsymbol{\alpha}, \lambda) = \sum_{i=1}^m \sum_{j=1}^m \boldsymbol{\alpha}_i^T \tilde{\mathbf{S}}_{ij} \boldsymbol{\alpha}_j - \lambda \left(\sum_{i=1}^m \boldsymbol{\alpha}_i^T \tilde{\mathbf{S}}_i^w \boldsymbol{\alpha}_i - 1 \right),\quad (18)$$

其中 λ 为 Lagrange 乘子. 对函数 $L(\boldsymbol{\alpha}, \lambda)$ 分别求取 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_m$ 的偏导数并令其为零, 则有

$$\frac{\partial L}{\partial \boldsymbol{\alpha}_i} = 2 \sum_{j=1}^m \tilde{\mathbf{S}}_{ij} \boldsymbol{\alpha}_j - 2\lambda \tilde{\mathbf{S}}_i^w \boldsymbol{\alpha}_i = 0, \quad i = 1, 2, \dots, m,\quad (19)$$

经过代数运算, 得到

$$\tilde{\mathbf{S}} \boldsymbol{\alpha} = \lambda \tilde{\mathbf{S}}_d^w \boldsymbol{\alpha},\quad (20)$$

其中 $\tilde{\mathbf{S}} \in \mathbf{R}^{p \times p}$ 的第 (i, j) 个块元素为 $\tilde{\mathbf{S}}_{ij}$, 并且 $\tilde{\mathbf{S}}_d^w = \text{diag}(\tilde{\mathbf{S}}_1^w, \tilde{\mathbf{S}}_2^w, \dots, \tilde{\mathbf{S}}_m^w) \in \mathbf{R}^{p \times p}$, $p = \sum_{i=1}^m p_i$. 进一步可获得如下的广义特征值问题

$$\begin{pmatrix} \tilde{\mathbf{S}}_1^b & \tilde{\mathbf{S}}_{12} & \cdots & \tilde{\mathbf{S}}_{1m} \\ \tilde{\mathbf{S}}_{21} & \tilde{\mathbf{S}}_2^b & \cdots & \tilde{\mathbf{S}}_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\mathbf{S}}_{m1} & \tilde{\mathbf{S}}_{m2} & \cdots & \tilde{\mathbf{S}}_m^b \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \vdots \\ \boldsymbol{\alpha}_m \end{pmatrix} = \lambda \begin{pmatrix} \tilde{\mathbf{S}}_1^w & & & \\ & \tilde{\mathbf{S}}_2^w & & \\ & & \ddots & \\ & & & \tilde{\mathbf{S}}_m^w \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \vdots \\ \boldsymbol{\alpha}_m \end{pmatrix}.\quad (21)$$

通过求解式(21)中的广义特征值问题, 选取前 d 个最大特征值 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$ 对应的特征向量 $\{\boldsymbol{\alpha}_i^{(1)}\}_{i=1}^m, \{\boldsymbol{\alpha}_i^{(2)}\}_{i=1}^m, \dots, \{\boldsymbol{\alpha}_i^{(d)}\}_{i=1}^m$ 作为 MSMCC 的投影方向, 记为 $\mathbf{W}_i = (\boldsymbol{\alpha}_i^{(1)}, \boldsymbol{\alpha}_i^{(2)}, \dots, \boldsymbol{\alpha}_i^{(d)}) \in \mathbf{R}^{p_i \times d}$, $i = 1, 2, \dots, m$. 对于 m 组高维数据 $\{\mathbf{X}^{(i)}\}_{i=1}^m$, 利用 $\{\mathbf{W}_i^T \mathbf{X}^{(i)}\}_{i=1}^m$ 实现联合特征抽取.

3.4 算法描述

根据上文的描述, MSMCC 算法的具体过程如下.

Step 1 PCA 降维. 由于原始样本数据的维数较高, 为克服小样本问题以及消除部分噪声的影响, 需利用 PCA 对原始样本数据进行降维处理;

Step 2 构建类内近邻图与类间惩罚图. 在每组数据的近邻图中, 对于类内的两个样本点 $\mathbf{x}_j^{(i)}$ 与 $\mathbf{x}_k^{(i)}$, 根据式(10)计算它们之间的权重 $\mathbf{W}_{w,jk}^{(i)}$ 并在加权矩阵 $\mathbf{W}_w^{(i)}$ 中令 $\mathbf{W}_{w,jk}^{(i)} = \mathbf{W}_{w,kj}^{(i)}$. 在每组数据的惩罚图中, 对于每个样本点 $\mathbf{x}_j^{(i)}$ 与它的不同类样本点 $\mathbf{x}_k^{(i)}$, 根据式(16)计算它们之间的权重 $\mathbf{W}_{b,jk}^{(i)}$ 并同时令 $\mathbf{W}_{b,jk}^{(i)} = \mathbf{W}_{b,kj}^{(i)}$;

Step 3 计算 Laplacian 矩阵. 在每组数据中, 利用加权矩阵 $\mathbf{W}_w^{(i)}$ 和 $\mathbf{W}_b^{(i)}$ 分别计算 Laplacian 矩阵 $\mathbf{L}^{(i)}$ 与 $\bar{\mathbf{L}}^{(i)}$;

Step 4 求解广义特征方程. 通过求解式(21)中的广义特征值问题, 选取前 d 个最大特征值 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$ 对应的特征向量 $\{\boldsymbol{\alpha}_i^{(1)}\}_{i=1}^m, \{\boldsymbol{\alpha}_i^{(2)}\}_{i=1}^m, \dots, \{\boldsymbol{\alpha}_i^{(d)}\}_{i=1}^m$ 作为最优化模型(17)的解向量, 并最终形成 m 组数据对应的投影矩阵 $\{\mathbf{W}_i = (\boldsymbol{\alpha}_i^{(1)}, \boldsymbol{\alpha}_i^{(2)}, \dots, \boldsymbol{\alpha}_i^{(d)})\}_{i=1}^m$;

Step 5 投影样本. 对于任意给定的多表示样本 $\mathbf{x}^T = (\mathbf{x}^{(1)T}, \mathbf{x}^{(2)T}, \dots, \mathbf{x}^{(m)T})$, 其中 $\mathbf{x}^{(i)} \in \mathbf{R}^{p_i}$, 使用 $\{\mathbf{W}_i^T \mathbf{x}^{(i)}\}_{i=1}^m$ 执行特征抽取, 并利用融合策略 $\mathbf{y} = \mathbf{W}^T \mathbf{x}$ 组合抽取出的特征, 以形成有效的鉴别矢量, 其中 $\mathbf{W}^T = (\mathbf{W}_1^T, \mathbf{W}_2^T, \dots, \mathbf{W}_m^T)$.

4 实验结果与分析

为了检验所提算法性能, 分别在 COIL20、COIL100 和 AT&T 三个数据库中进行了识别实验, 与 MCA^[24]、MCCA、MICCA^[25]、SMCC 以及基准算法 PCA 进行对比. 本实验中对比方法的详细信息如下.

MCA^[24]: 将每组特征视为张量组成中的一阶, 然后构建出协方差张量, 并对其采用高阶奇异值分解来获取每一阶数据的投影矩阵, 最后利用对应的投影矩阵抽取和融合多组特征.

MICCA^[25]: 先抽取同一模式的多组特征向量, 然后依据多组向量的判别准则函数, 求取多重集整体典型相关特征, 最后融合抽取到的特征.

在 MSMCC 算法中有两个参数, 即近邻参数 k_1 与 k_2 . 但如何理论上确定最优的参数仍然是一个未被解决的问题. 因此, 在接下来的识别实验中, 采用经验的方式设置参数值, 即近邻参数 k_1 选自集合 $\{1, 2, \dots, l-1\}$, 其中 l 为每类训练样本数, 最后采用交叉验证法来确定最优的参数 k_1 . 近邻参数 k_2 经验地设置为 $k_2 = k_1$. 在识别阶段, 采用了余弦距离下的最近邻分类器.

4.1 在 COIL-20 数据库上的实验

COIL-20^[26] 数据库包含了 20 个对象, 每个对象有 1440 幅灰度图像. 对每个对象从 0° 至 360° 进行水平方向的旋转, 并且每隔 5° 采样一幅图像, 这样每个对象总共有 72 幅图像. 其中每个对象都有复杂的几何结构, 并且每幅图像的大小为 128×128 . 图 1 显示了所有对象的某一幅图像.



图 1 COIL-20 数据库中所有对象的某一幅图像

在实验中将每张图像的尺寸归一化到 32×32 . 根据文献[12]中的思想, 分别采用 Symlets、Daubechies 以及 Coiflets 正交小波变换对原始图像进行分解从而获得三组低频子图像, 然后使用 K-L 变换对三组低频特征向量分别降至 150 维, 以增强算法的稳定性以及避免小样本问题的出现. 然后随机地选取 N ($N = 36, 42, 48$) 幅图像进行训练, 而剩下的 $72 - N$ 幅图像用于测试. 对于每个给定的 N , 随机地进行 10 次实验, 然后计算其平均识别率. 表 1 给出了最终的识别结果.

表 1 COIL-20 数据库中不同方法在不同训练样本个数下的错分结果

方 法	N=36	N=42	N=48
PCA	0.1132(147)	0.0980(150)	0.0879(150)
MCA	0.1744(137)	0.1247(150)	0.1073(148)
MCCA	0.0681(36)	0.0450(61)	0.0479(75)
MICCA	0.0794(43)	0.0638(40)	0.0508(46)
SMCC	0.0694(35)	0.0603(66)	0.0540(42)
MSMCC	0.0403(115)	0.0350(130)	0.0167(131)

表 1 显示出 MSMCC 方法在所有特征组合上均优于 PCA、MCA、MCCA、MICCA 以及 SMCC 方法. 特别是在训练样本数为 48 的情况下, MSMCC 的识别率是 98.33%, 分别超出 PCA 方法 7.12%, MCA 方法 9.06%, MCCA 方法 3.12%, MICCA 方法 3.41% 以及 SMCC 方法 3.73%. 这揭示出 MSMCC 方法抽取出的特征相比于其它方法具有更好的鉴别能力.

4.2 在 COIL-100 数据库上的实验

COIL-100^[27] 数据库包含了 100 个对象的 7200 幅灰度图像. 对每一个对象从 0° 至 360° 进行水平方向的旋转, 每隔 5° 采样一幅图像, 这样每个对象共有 72 幅图像. 每个对象都有复杂的几何结构, 并且图像的大小为 128×128. 图 2 显示了每个对象的某一幅图像.

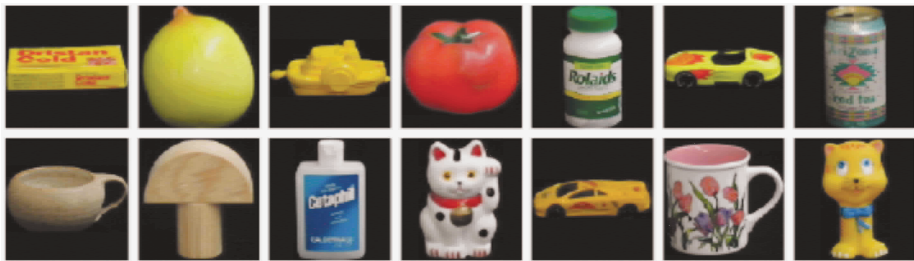


图 2 COIL-100 数据库中每个对象的某一幅图像

在实验中, 将每张图像的尺寸归一化到 32×32, 并随机地选取 N 幅 (N = 30, 36, 42, 48) 图像进行训练, 而剩下的 72 - N 图像被用于测试. 对于每个给定的 N, 随机地进行 10 次实验, 然后计算其平均识别率. 最终识别结果如表 2 所示. 图 3 给出在 36 个训练样本下各方法的识别结果随着特征维数增加变化情况.

表 2 COIL-100 数据库中不同方法在不同训练样本个数下的错分结果

方 法	N=30	N=36	N=42	N=48
PCA	0.2267(195)	0.1881(195)	0.1579(197)	0.1304(197)
MCA	0.2512(111)	0.2283(136)	0.1960(149)	0.1625(149)
MCCA	0.1414(111)	0.1222(93)	0.1060(113)	0.0913(127)
MICCA	0.1291(64)	0.1093(69)	0.0924(66)	0.0802(118)
SMCC	0.1477(143)	0.1189(144)	0.0989(146)	0.0831(62)
MSMCC	0.0976(140)	0.0744(139)	0.0617(137)	0.0450(125)

通过表 2 可以看出, 无论训练样本个数是多少, MSMCC 获得了比其他五个算法都好的识别结果; 在所有方法中, PCA 方法获得了最差的识别结果, 并且它获得的最优识别结果的维数是最大的; 对于所有算法来说, 随着训练样本个数的增加, 它们的识别率也在增加. 图 3 显示出当特征维数大于 30 时, MSMCC 方法随着维数的增加始终优于其它五个算法. 这些结果再一次表明 MSMCC 是强有力的, 而且还进一步显示出它在对象种类识别中也是一个很好的方法.

4.3 在 AT&T 数据库上的实验

AT&T 人脸数据库^[10] 共有 400 幅图像, 包含 40 个人, 每人 10 幅, 图像分辨率为 92×112. 这些图像分别拍

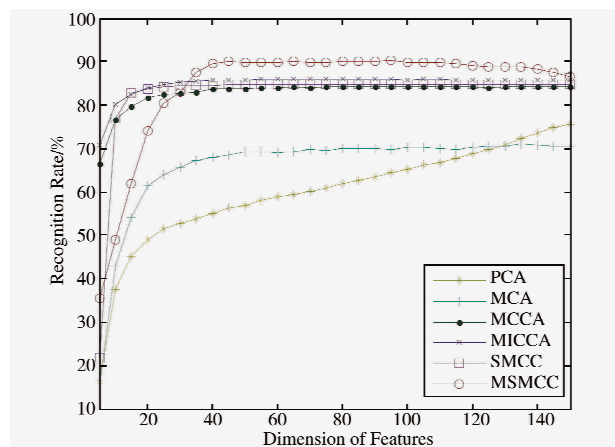


图 3 PCA、MCA、MCCA、MICCA、SMCC 以及 MSMCC 在 COIL-100 上的识别结果随特征维数变化的趋势

摄于不同的时期,其中脸部的表情、脸部的细节、人脸姿态和人脸的尺度都有着不同程度的变化,比如眼睛或睁或闭,笑或不笑,戴或不戴眼镜.图4给出了AT&T数据库中某个人的10幅图像.

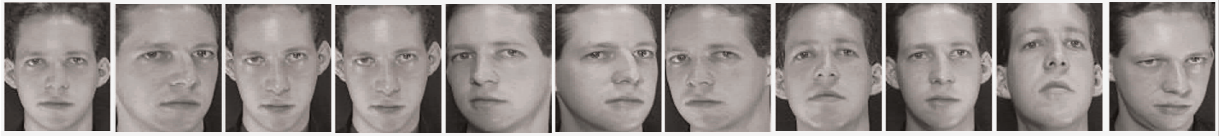


图4 在AT&T数据库中某个人的10幅图像

表3 AT&T数据库中不同方法在不同训练样本个数下的错分结果

方法	$N=5$	$N=6$	$N=7$
PCA	0.1070(114)	0.0769(114)	0.0433(77)
MCA	0.4750(130)	0.3638(142)	0.2625(145)
MCCA	0.2650(103)	0.2125(71)	0.1750(48)
MICCA	0.2070(37)	0.1563(60)	0.1075(114)
SMCC	0.2350(85)	0.1419(97)	0.0892(115)
MSMCC	0.0700(76)	0.0250(142)	0.0083(104)

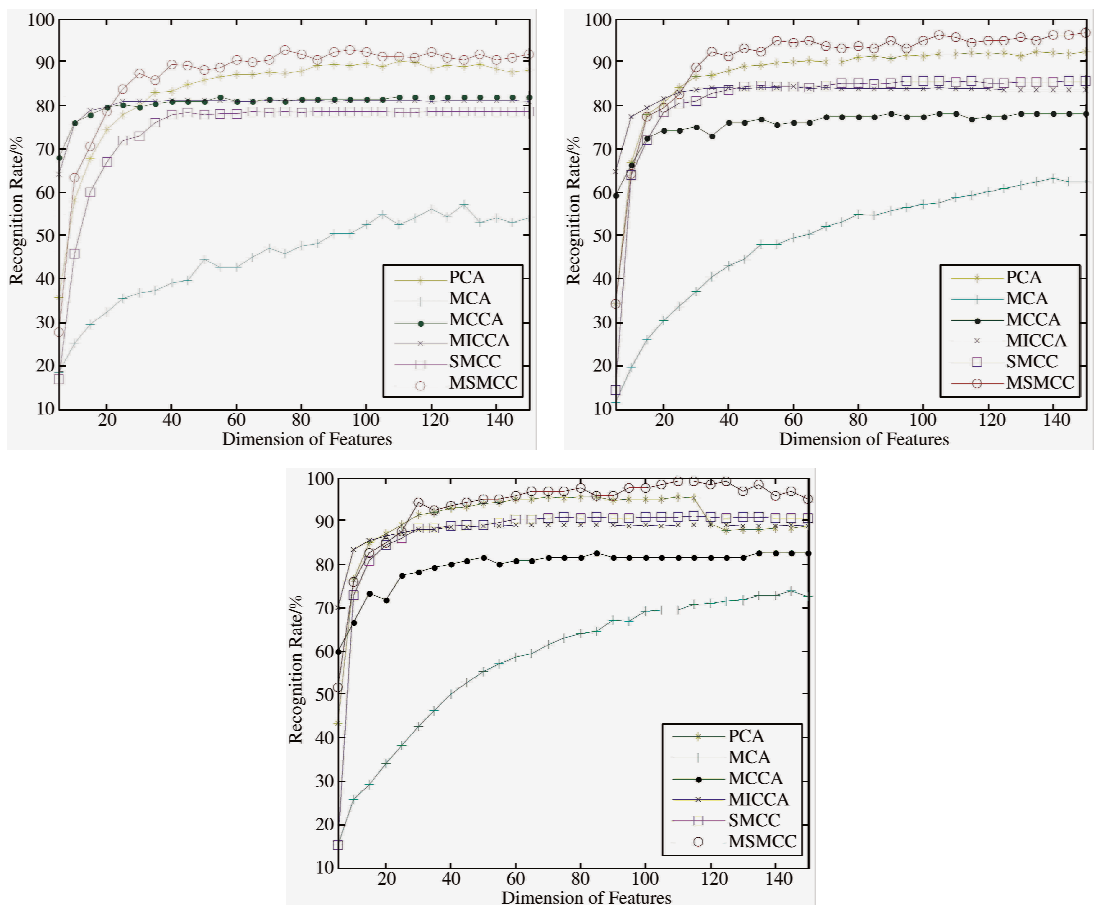


图5 PCA、MCA、MCCA、MICCA、SMCC以及MSMCC在AT&T上的识别结果

随特征维数变化的趋势 (a) $N=5$; (b) $N=6$; (c) $N=7$

在实验中,随机地选取 N 幅($N=5,6,7$)图像进行训练,而剩下的 $10-N$ 图像被用于测试.对于每个给定的 N ,随机地进行10次实验,然后计算其平均识别率.最终识别结果如表3所示.图5给出了各方法的识别结果随着特征维数增加的变化情况.

从表3与图5可以看出,MSMCC在所有情况下均优于PCA、MCA、MCCA、MICCA与SMCC算法.尽管与基准算法PCA相比,MSMCC的改善并不明显,但它仍是执行的最好的.这再次表明MSMCC抽取出的特征具有较好的鉴别力.在这个数据集上,MCA获得了非常低的识别结果.

5 结论

为了同时刻画样本数据的几何结构与分布信息,在 MFA 的思想启发下,提出一种新的特征抽取方法,即边界监督多重集典型相关分析(MSMCC).该算法的核心思想是在多重集相关分析的框架下加入多特征表示数据的鉴别信息与本质几何结构.基于这个目标,首先考虑样本的类别信息,提出了监督多重集典型相关分析(SMCC),然后在此基础上构建了 MSMCC 的最优模型,最后给出了求解算法.在人脸图像数据库以及对对象数据库上的实验结果表明,所提出的 MSMCC 算法获得了令人满意的识别性能,其抽取出的特征比 PCA、MCA、MCCA、MICCA 以及所提 SMCC 方法更具有鉴别能力.

参 考 文 献

- [1] Hotelling H. Relations between two sets of variates[J]. *Biometrika*, 1936, 28(3/4): 321-377.
- [2] Sun Q S, Zeng S G, Liu Y, et al. A new method of feature fusion and its application in image recognition[J]. *Pattern Recognition*, 2005, 38(12): 2437-2448.
- [3] 孙权森, 曾生根, 王平安. 典型相关分析的理论及其在特征融合中的应用[J]. *计算机学报*, 2005, 28(9): 1524-1533.
- [4] Carlos A, Johan A K S. A regularized kernel CCA contrast function for ICA[J]. *Neural Networks*, 2008, 21(2/3): 170-181.
- [5] 许颖春, 范丽亚. 基于 L_1 范数的非线性 TSVR[J]. *聊城大学学报(自然科学版)*, 2017, 30(3): 6-11.
- [6] 夏文静, 陈耿, 范丽亚. 八种最小二乘 SVM 型学习算法的优势比较[J]. *聊城大学学报(自然科学版)*, 2016, 29(2): 33-41.
- [7] Sun Q S, Liu Z D, Heng P A, et al. A theorem on the generalized canonical projective vectors[J]. *Pattern Recognition*, 2005, 38(3): 449-452.
- [8] Sun T K, Chen S C, Yang J Y, et al. A supervised combined feature extraction method for recognition[C]. // *Proceedings of IEEE International Conference on Data Mining, Pisa, Italy, 2008*.
- [9] Yang J, Sun Q S. A novel generalized fuzzy canonical correlation analysis framework for feature fusion and recognition[J]. *Neural Processing Letters*, 2017, 46: 521-536.
- [10] Yang J, Sun Q S, Yuan Y H. Feature extraction using fractional-order embedding direct linear discriminant analysis[J]. *Neural Processing Letters*, 2018, 48(3): 1583-1595.
- [11] Kim T K, Kittler J, Cipolla R. Discriminative learning and recognition of image set classes using canonical correlations[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, 29(6): 1005-1018.
- [12] Li Y O, Adali T, Wang W, et al. Joint blind source separation by multiset canonical correlation analysis[J]. *IEEE Transactions on Signal Processing*, 2009, 57(10): 3918-3929.
- [13] Thompson B, Cartmill J, Azimi-Sadjadi M R, et al. A multichannel canonical correlation analysis feature extraction with application to buried underwater target classification[C]. // *International Joint Conference on Neural Networks*, 2006.
- [14] Ji H K, Shen X B, Sun Q S, et al. Sparse Discrimination based Multiset Canonical Correlation Analysis for Multi-Feature Fusion and Recognition[C]. // *British Machine Vision Conference*, 2015.
- [15] Ji H K, Sun Q S, Yuan Y H, et al. C2DMCP: View-consistent collaborative discriminative multiset correlation projection for data representation[J]. *Journal of Visual Communication and Image Representation*, 2016, 40: 393-405.
- [16] Nielsen A A. Multiset canonical correlations analysis and multispectral, truly multitemporal remote sensing data[J]. *IEEE Transactions on Image Processing*, 2002, 11(3): 293-305.
- [17] Correa N M, Eichele T, Adal T, et al. Multi-set canonical correlation analysis for the fusion of concurrent single trial ERP and functional MR[J]. *NeuroImage*, 2010, 50(4): 1438-1445.
- [18] LiEmail Y O, Eichele T, Calhoun V D, et al. Group Study of Simulated Driving fMRI Data by Multiset Canonical Correlation Analysis[J]. *Journal of Signal Processing Systems*, 2012, 68(1): 31-48.
- [19] Deleus F, Hulle M M V. Functional connectivity analysis of fMRI data based on regularized multiset canonical correlation analysis[J]. *Journal of Neuroscience Methods*, 2011, 197(1): 143-157.
- [20] Gao L, Qi L, Chen E Q, et al. Discriminative Multiple Canonical Correlation Analysis for Multi-Feature Information Fusion[C]. // *2012*

IEEE Internation Symposium on Multimedia,2012.

- [21] Lee G, Singanamalli A. Supervised Multi-View Canonical Correlation Analysis (sMVCCA): Integrating histologic and proteomic features for predicting recurrent prostate cancer[J]. IEEE Transactions on Medical Imaging, 2015, 34(1): 284-297.
- [22] Yan S, Xu D, Zhang B, et al. Graph embedding and extensions: a general framework for dimensionality reduction[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(1): 40-51.
- [23] Chung F R K. Spectral graph theory [D]. American: Mathematical Society, 1997.
- [24] Hou S, Sun Q, Xia D. Feature fusion using multiple component analysis[J]. Neural Processing Letters, 2011, 34(3): 259-275.
- [25] Yuan Y H, Sun Q S, Zhou Q, et al. A novel multiset integrated canonical correlation analysis framework and its application in feature fusion[J]. Pattern Recognition, 2011, 44(5): 1031-1040.
- [26] Nene S A, Nayar S K, Murase H. Columbia object image library (COIL-20)[R]. Technical Report CUCS-005-96, 1996.
- [27] Nene S A, Nayar S K, Murase H. Columbia object image library (COIL-100)[R]. Technical Report CUCS-006-96, 1996.

Marginal Supervised Multiset Canonical Correlation Analysis

YANG Jing¹ GAO Xi-zhan²

(1. School of Mathematical Sciences, Liaocheng University, Liaocheng 252059, China;

2. School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China)

Abstract Due to multiset canonical correlation analysis (MCCA) fails to discover the intrinsic geometrical and discriminating structure of multiple data spaces in real world applications. In order to solve this problem, we first construct supervised multiset canonical correlation analysis (SMCC) by using the ideas from LDA. Based on marginal fisher analysis (MFA), we then propose marginal supervised multiset canonical correlation analysis (MSMCC). It can not only express the correlation among multiple feature vectors, but also effectively depict the data between the geometry and discriminative structure. Extensive experiments on both face image databases and object databases demonstrate the effectiveness of the proposed method.

Key words canonical correlation analysis; multiset canonical correlation analysis; feature extraction; dimensionality reduction; supervised learning