

# 基于核稀疏保持投影的典型相关分析算法

张 荣 孙权森

(南京理工大学计算机科学与工程学院, 南京, 210094)

**摘 要:** 模式识别的技术核心就是特征提取, 而特征融合则是对特征提取方法的强力补充, 对于提高特征的识别效率具有重要作用。本文基于稀疏表示方法, 将稀疏表示方法用到高维度空间, 并利用核方法在高维度空间进行稀疏表示, 用其计算核稀疏表示系数, 同时研究了核稀疏保持投影算法(Kernel sparsity preserve projection, KSPP)。将 KSPP 引入到典型相关分析算法(Canonical correlation analysis, CCA), 研究了基于核稀疏保持投影的典型相关分析算法(Kernel sparsity preserve canonical correlation analysis, K-SPCCA)。在多特征手写体数据库和人脸图像数据库上分别证实了本文提出方法的可靠性和有效性。

**关键词:** 特征提取; 核稀疏表示; 核稀疏保持投影; 典型相关分析

**中图分类号:** TP751      **文献标志码:** A

## Canonical Correlation Analysis Algorithm Based on Kernel Sparsity Preserve Projection

Zhang Rong, Sun Quansen

(School of Computer Science and Engineering, Nanjing University of Science & Technology, Nanjing, 210094, China)

**Abstract:** The key of pattern recognition is feature extraction. Fusion of feature is an important complement of feature extraction, and it has been proved to be important to improve discrimination. Here, the sparse representation method is studied by introducing sparse representation into a high dimensional feature space and utilizing kernel trick to make sparse representation in the space. The kernel sparse representation coefficients with kernel sparse representation are utilized, then kernel sparsity preserve projection (KSPP) subspace. Moreover KSPP is brought into canonical correlation analysis (CCA), then kernel sparsity preserve canonical correlation analysis (KSPCCA) is studied. The proposed algorithm is reliable and validated on the multiple feature database and face database.

**Key words:** feature extraction; kernel sparse representation; kernel sparsity preserve projection (KSPP); canonical correlation analysis (CCA)

## 引 言

随着压缩感知理论的迅速发展, 在 20 世纪 90 年代, 稀疏表示理论应运而生, 稀疏表示理论就是从原始的高维样本数据中提取低维空间的特征, 使得这个投影满足稀疏性, 使得这个表示系数的非 0 个数

最少。通常使用过完备字典<sup>[1]</sup>的设计来求解稀疏表示系数,其原理是:字典中的所有元素选取全部的原始高维样本数据,然后在字典中尽量选择最少的基本元素来表示原始高维的样本数据,从而使这个表示系数里不是 0 的个数越少越好。Qiao<sup>[2]</sup>等提出了一种基于稀疏表示的特征提取的方法——稀疏保持投影(Sparsity preserve projection, SPP),而最近侯书东等<sup>[3,4]</sup>受稀疏算法的启发,将 SPP 引入到典型相关分析(Canonical correlation analysis, CCA)中,提出了基于稀疏保持的典型相关分析(Sparsity preserve canonical correlation analysis, SPCCA),这种算法不仅约束了样本之间的稀疏重构性,同时实现了两组不同特征之间的特征融合,从而增强了该特征提取方法的识别率。Sun 通过原始样本数据提取两种不同的特征,以这两组不同的特征作为典型相关分析的两组输入数据,在 CCA 准则下最大化这两组特征的相关性,得到两组投影矢量,通过不同的策略,融合这两组投影矢量,对原始样本数据进行新特征提取工作,用于样本数据的分类。CCA 是一种线性的特征提取模型,其计算能力有限。一般而言,现实世界中很多复杂的应用问题,比如 Shen<sup>[5]</sup>等提出了正交化 CCA,还有核的 CCA<sup>[6]</sup>都是比线性函数更富有表达能力的假设空间。核方法的原理是:通过某种未知的非线性映射关系,将原始样本数据映射到高维特征空间或者无穷维的特征空间,然后在新的空间中进行线性形式的操作。尽管通过这种非线性的映射,存在以下问题:非线性映射函数形式的不确定和函数的某些参数不确定等,但是这些问题都可以通过核函数技术的内积运算加以解决。20 世纪 90 年代以来,随着核方法在支持向量机(SVM)<sup>[7,8]</sup>中的成功应用,一大批核化的算法相继被提出,核技术在模式识别领域得到了很大扩展和成功运用,主要包括两方面:新方法主要用于支持向量机中,但是这里使得 SVM 结构风险最小;传统方法改造提出了很多有效地稳定的算法,主要包括核主元分析(Kernel PCA, KPCA)<sup>[9]</sup>、核主元回归(Kernel PCR, KPCR)、核部分最小二乘法(Kernel PLS, KPLS)、核 Fisher 判别分析(Kernel fisher discriminator, KFD)<sup>[10,11]</sup>以及核独立主元分析(Kernel independent component analysis, KICA),这些不同的方法在模式识别中都取得了不错的、甚至意想不到的结果。这些核方法在处理非线性问题时都获得了成功。根据 Cover T 的模式可分性定理,当模式分类问题映射到高维空间时,会比在低维度空间更加容易可识别或者说可分<sup>[12]</sup>。受该定理的启发,本文运用核方法(所谓“核技巧”<sup>[8,13]</sup>,就是在算法中出现样本内积时进行核函数代入),将线性的 SPCCA 推广到高维特征空间,提出了核化的 SPCCA(Kernelized SPCCA, KSPCCA)。最后利用优化策略将它们引入 CCA 理论。

## 1 核稀疏保持投影的典型相关分析

### 1.1 稀疏表示理论<sup>[14]</sup>

假设存在包含  $c$  类的训练样本,每类样本集合表示为  $X_i = [x_1^i, x_2^i, \dots, x_{n_i}^i] \in R^{p \times n_i}$ ,其中  $p$  表示每类样本的特征维数,  $n_i$  表示第  $i$  类的训练样本数目。则第  $i$  类的任一样本  $y \in R^p$  可表示为

$$y = \alpha_1^i x_1^i + \dots + \alpha_{n_i}^i x_{n_i}^i + \dots + \alpha_1^j x_1^j + \dots + \alpha_{n_j}^j x_{n_j}^j + \dots + \alpha_1^c x_1^c + \dots + \alpha_{n_c}^c x_{n_c}^c \quad (1)$$

其中  $\alpha_j^i$  表示第  $i$  类的第  $j$  个样本的权重系数,然后将每类的数据矩阵,组合成如下向量矩阵  $\mathbf{X}$ ,有

$$\mathbf{X} = [\mathbf{X}_1, \dots, \mathbf{X}_i, \dots, \mathbf{X}_c] = [x_1^1, \dots, x_{n_1}^1, \dots, x_1^i, \dots, x_{n_i}^i, \dots, x_1^c, \dots, x_{n_c}^c] \quad (2)$$

所以如果测试样本的类别不知道,就通过所有的原始样本数据的线性组合,表示测试样本  $y$ ,即

$$y = \mathbf{X}\boldsymbol{\alpha} \in R^p \quad (3)$$

其中  $\boldsymbol{\alpha} = [\alpha_1^1, \dots, \alpha_{n_1}^1, \dots, \alpha_1^i, \dots, \alpha_{n_i}^i, \dots, \alpha_1^c, \dots, \alpha_{n_c}^c]^T$  表示系数向量。

当式(3)中字典  $\mathbf{X}$  的基本元素较多时,即原始样本数据很多,系数向量  $\boldsymbol{\alpha}$  就会只包含很少数的有用的系数,其他的大部分系数都为 0,因而稀疏表示的最后就是要求出解稀疏系数向量  $\boldsymbol{\alpha}$ ,满足式(3)的解中选择  $\boldsymbol{\alpha}$  最稀疏的解,也就是求解式(4)的约束条件的最小化问题,这里采用  $L_1$  范数来加以限制

$$\arg \min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1$$

$$\text{s. t. } \|y - \mathbf{X}\boldsymbol{\alpha}\| \quad (4)$$

式(4)可以变为下面的正则化表现形式

$$\min_{\boldsymbol{\alpha}} \|y - \mathbf{X}\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1 \quad (5)$$

图 1 给出了只有两组特征变量情况下在  $L_1$  范数和  $L_2$  范数限制下的优化问题求解图示,图 1 中正方形虚线表示  $L_1$  范数的优化,圆形实线表示  $L_2$  范数的优化,从图中可以很容易地看出,只有范数才能达到稀疏的目的。

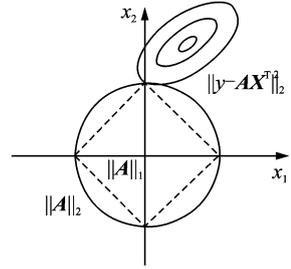


图 1 稀疏表示的求解示意图

Fig. 1 Schematic diagram of solving sparse representation

### 1.2 稀疏保持投影算法

SPP 算法较好地保持了全局样本间的稀疏重构性。文献[4]中给出了 SPP 算法的优化目标函数,通过数学推导可得到广义本征方程。

假设同一模式下的一组训练样本,记为  $\{\mathbf{x}_i\}_{i=1}^N$ ,  $N$  表示样本数,其中  $\mathbf{x}_i \in R^p$ ,  $p$  表示每个样本的特征维数。令  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in R^{p \times N}$  表示所有的训练样本数据。通过如下公式的最小化问题,求解每一个样本  $\mathbf{x}_i$  的稀疏重构权重向量  $\mathbf{s}_i$ , 求解过程同样采用  $L_1$  范数,即有

$$\begin{aligned} \min_{\mathbf{s}_i} \|\mathbf{s}_i\|_1 \\ \text{s. t. } \mathbf{x}_i = \mathbf{X}\mathbf{s}_i \end{aligned} \quad (6)$$

式中:  $\mathbf{s}_i = [s_{i1}, \dots, s_{i,j-1}, 0, s_{i,j+1}, \dots, s_{iN}]^T$  是一个  $N$  维的向量代表  $\mathbf{x}_i$  的稀疏重构权重向量,第  $i$  个稀疏系数为零,说明在求解稀疏权重向量时已经排除了自己;  $s_j^i (i \neq j)$  表示其他所有的训练样本  $\mathbf{x}_j$  对重构样本  $\mathbf{x}_i$  的权重系数,最后用每一个样本的稀疏重构向量构造训练样本的稀疏重构邻接矩阵,即

$$\mathbf{S} = [\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \dots, \tilde{\mathbf{s}}_N]^T \quad (7)$$

其中  $\tilde{\mathbf{s}}_i$  是式(6)的最优解。假设降维后的特征是  $\omega^T \mathbf{x}_i$ , 定义如下优化目标函数,来寻求能够保持最佳权重向量  $\mathbf{s}_i$  的投影  $\omega$ , 即有

$$\min_{\omega} \sum_{i=1}^N \|\omega^T \mathbf{x}_i - \omega^T \mathbf{X}\tilde{\mathbf{s}}_i\|^2 \quad (8)$$

通过数学推导,可以将式(8)化简为

$$\begin{aligned} \sum_{i=1}^N \|\omega^T \mathbf{x}_i - \omega^T \mathbf{X}\tilde{\mathbf{s}}_i\|^2 &= \omega^T \left( \sum_{i=1}^N (\mathbf{x}_i - \mathbf{X}\tilde{\mathbf{s}}_i)(\mathbf{x}_i - \mathbf{X}\tilde{\mathbf{s}}_i)^T \right) \omega = \\ \omega^T \left( \sum_{i=1}^N (\mathbf{X}\mathbf{e}_i - \mathbf{X}\tilde{\mathbf{s}}_i)(\mathbf{X}\mathbf{e}_i - \mathbf{X}\tilde{\mathbf{s}}_i)^T \right) \omega &= \omega^T \mathbf{X} \left( \sum_{i=1}^N (\mathbf{e}_i - \tilde{\mathbf{s}}_i)(\mathbf{e}_i - \tilde{\mathbf{s}}_i)^T \right) \mathbf{X}^T \omega = \\ \omega^T \mathbf{X} \left( \sum_{i=1}^N \mathbf{e}_i \mathbf{e}_i^T - \tilde{\mathbf{s}}_i \mathbf{e}_i^T - \mathbf{e}_i \tilde{\mathbf{s}}_i^T + \tilde{\mathbf{s}}_i \tilde{\mathbf{s}}_i^T \right) \mathbf{X}^T \omega &= \omega^T \mathbf{X} (\mathbf{I} - \mathbf{S}^T - \mathbf{S} + \mathbf{S}^T \mathbf{S}) \mathbf{X}^T \omega \end{aligned}$$

其中  $\mathbf{e}_i$  表示第  $i$  个元素为 1, 其他均为 0 的列向量。在求解过程中为了避免出现退化的解,给定约束条件,  $\omega^T \mathbf{X}\mathbf{X}^T \omega = 1$ , 因此,最终 SPP 的广义本征方程为

$$\mathbf{X}(\mathbf{S}^T + \mathbf{S} - \mathbf{S}^T \mathbf{S}) \mathbf{X}^T \omega = \lambda \mathbf{X}\mathbf{X}^T \omega \quad (9)$$

求解式(9),取得  $d$  个最大特征值所对应的投影向量  $\omega_i (i=1, \dots, d)$ , 即为所求的投影。

由于矩阵  $\mathbf{X}\mathbf{X}^T$  一般都不可逆,因而式(9)一般不能直接求解,即常说的小样本问题。为了解决小样本问题,就要求在特征提取前对原始样本数据进行 PCA 操作,降低原始样本数据的特征维数,从而避免  $\mathbf{X}\mathbf{X}^T$  奇异。

### 1.3 核稀疏保持投影

通过核方法将原始的样本数据映射到高维特征空间,改变了原始样本的分布情况,从而将原始样本数据中的非线性问题转化为线性问题。因此,在高维特征空间中,针对新分布的样本情况进行稀疏表示

的工作,就可以得到高维特征空间的核稀疏表示系数;利用得到的核稀疏表示系数,构造高维特征空间中样本数据的邻接矩阵来提取数据特征,称为核稀疏保持投影(KSPP)。

### 1.3.1 核稀疏表示

假设原始样本数据通过非线性映射  $\varphi$  映射到高维特征空间,则令  $\Phi(\mathbf{X}) = [\Phi(x_1), \dots, \Phi(x_N)]$ , 它就表示高维特征空间中的样本情况,那么高维特征空间中进行稀疏表示的求解,具体形式为

$$\begin{aligned} & \arg \min_s \|\mathbf{s}\|_1 \\ & \text{s. t. } \Phi(y) = \Phi(\mathbf{X})\mathbf{s} \end{aligned} \quad (10)$$

由于  $\Phi(\mathbf{X})$  和  $\Phi(x_i)$  均未知,所以式(10)不能直接求解。在约束条件的两边同时乘上  $\Phi(\mathbf{X})^T$ , 即

$$\begin{aligned} & \arg \min_s \|\mathbf{s}\|_1 \\ & \text{s. t. } \Phi(\mathbf{X})^T \Phi(y) = \Phi(\mathbf{X})^T \Phi(\mathbf{X})\mathbf{s} \end{aligned} \quad (11)$$

在高维特征空间中的内积,同样也是通过核函数来求得,两个样本数据  $x_1, x_2$ , 映射到高维特征空间后的内积形式为:  $\Phi(x_1)^T \Phi(x_2) = K(x_1, x_2)$ , 则

$$\Phi(\mathbf{X})^T \Phi(\mathbf{X}) = [\Phi(x_1), \dots, \Phi(x_N)]^T [\Phi(x_1), \dots, \Phi(x_N)] = \begin{bmatrix} K(x_1, x_1) & \cdots & K(x_1, x_N) \\ \vdots & \ddots & \vdots \\ K(x_N, x_1) & \cdots & K(x_N, x_N) \end{bmatrix} \quad (12)$$

并且

$$\Phi(\mathbf{X})^T \Phi(y) = [\Phi(x_1), \dots, \Phi(x_N)]^T \Phi(y) = \begin{bmatrix} K(x_1, y) \\ K(x_2, y) \\ \vdots \\ K(x_N, y) \end{bmatrix} \quad (13)$$

当核函数  $K(x_1, x_2)$  给定时,  $\Phi(\mathbf{X})^T \Phi(\mathbf{X})$  和  $\Phi(\mathbf{X})^T \Phi(y)$  就可以根据式(12,13)求得,从而求解(11)的最优化问题。

### 1.3.2 核稀疏保持投影

利用上一小节得到的高维特征空间中的核稀疏表示系数,建立样本数据的邻接矩阵,并用于特征提取工作。同样地,通过某种未知的非线性映射  $\varphi$ , 将原始的样本数据  $x_i \in R^p (i=1, \dots, N)$  映射到高维特征空间,得到映射后的样本数据  $\varphi(x_i) (i=1, \dots, N)$ 。对于现在的高维特征空间任一个样本数据  $\varphi(x_i) (i=1, \dots, N)$ ; 也 同样地,在高维特征空间中用除自己以外的所有样本数据求解自己的稀疏表示系数。其中  $p$  表示样本的特征维数,  $N$  表示训练样本数,则  $\varphi(\mathbf{X}) = [\varphi(x_1), \dots, \varphi(x_N)]$  表示所有高维特征空间的样本数据。仿照 SPP 理论的方法,可以得到 KSPP 的优化问题模型,即

$$\begin{aligned} & \min_s \|\mathbf{s}_i\|_1 \\ & \text{s. t. } \Phi(x_i) = \Phi(\mathbf{X})\mathbf{s}_i \end{aligned} \quad (14)$$

式中:  $\Phi(\mathbf{X}) = [\Phi(x_1), \dots, \Phi(x_N)]$  表示高维特征空间的所有训练样本矩阵;  $\mathbf{s}_i = [s_{i1}, \dots, s_{i,j-1}, 0, s_{i,j+1}, \dots, s_{iN}]^T$  表示高维特征空间中的稀疏表示系数向量;  $\mathbf{s}_j (i \neq j)$  表示高维特征空间中训练样本  $\Phi(x_j)$  对重构  $\Phi(x_i)$  的权重。 $\Phi(\mathbf{X})$  和  $\Phi(x_i)$  都是未知的,也不能求解,继续在约束条件的两边同时乘上  $\Phi(\mathbf{X})^T$ , 即

$$\begin{aligned} & \arg \min_s \|\mathbf{s}_i\|_1 \\ & \text{s. t. } \Phi(\mathbf{X})^T \Phi(x_i) = \Phi(\mathbf{X})^T \Phi(\mathbf{X})\mathbf{s}_i \end{aligned} \quad (15)$$

利用核函数计算高维空间数据的内积,可得

$$\Phi(\mathbf{X})^T \Phi(\mathbf{X}) = [\Phi(x_1), \dots, \Phi(x_N)]^T [\Phi(x_1), \dots, \Phi(x_N)] = \begin{bmatrix} K(x_1, x_1) & \cdots & K(x_1, x_N) \\ \vdots & \ddots & \vdots \\ K(x_N, x_1) & \cdots & K(x_N, x_N) \end{bmatrix} = K \quad (16)$$

和

$$\Phi(\mathbf{X})^T \Phi(\mathbf{x}_i) = [\Phi(\mathbf{x}_1), \dots, \Phi(\mathbf{x}_N)]^T \Phi(\mathbf{x}_i) = \begin{bmatrix} K(\mathbf{x}_1, \mathbf{x}_i) \\ K(\mathbf{x}_2, \mathbf{x}_i) \\ \vdots \\ K(\mathbf{x}_N, \mathbf{x}_i) \end{bmatrix} \quad (17)$$

这样就可以求解式(15)的最优化问题,得到核稀疏表示系数向量  $\tilde{\mathbf{s}}_i (i=1, \dots, N)$ , 它们组合起来就可以得到稀疏重构邻接矩阵  $\mathbf{S} = [\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \dots, \tilde{\mathbf{s}}_N]^T$ 。这样就可以建立 KSPP 的优化目标函数,则有

$$\min_{\mathbf{w}} \sum_{i=1}^N \|\mathbf{w}^T \Phi(\mathbf{x}_i) - \mathbf{w}^T \Phi(\mathbf{X}) \tilde{\mathbf{s}}_i\|^2 \quad (18)$$

通过类似 SPP 的推导,可以得到

$$\sum_{i=1}^N \|\mathbf{w}^T \Phi(\mathbf{x}_i) - \mathbf{w}^T \Phi(\mathbf{X}) \tilde{\mathbf{s}}_i\|^2 = \mathbf{w}^T \Phi(\mathbf{X}) (\mathbf{I} - \mathbf{S}^T - \mathbf{S} + \mathbf{S}^T \mathbf{S}) \Phi(\mathbf{X})^T \mathbf{w}$$

同样地,给定约束条件  $\mathbf{w}^T \Phi(\mathbf{X}) \Phi(\mathbf{X})^T \mathbf{w} = 1$ , 这样优化准则可以变成

$$\min_{\mathbf{w}} \frac{\mathbf{w}^T \Phi(\mathbf{X}) (\mathbf{I} - \mathbf{S}^T - \mathbf{S} + \mathbf{S}^T \mathbf{S}) \Phi(\mathbf{X})^T \mathbf{w}}{\mathbf{w}^T \Phi(\mathbf{X}) \Phi(\mathbf{X})^T \mathbf{w}} \quad (19)$$

等价于最大化准则

$$\max_{\mathbf{w}} \frac{\mathbf{w}^T \Phi(\mathbf{X}) (\mathbf{S}^T + \mathbf{S} - \mathbf{S}^T \mathbf{S}) \Phi(\mathbf{X})^T \mathbf{w}}{\mathbf{w}^T \Phi(\mathbf{X}) \Phi(\mathbf{X})^T \mathbf{w}} \quad (20)$$

最大化准则可以转化为求解广义本征方程

$$\Phi(\mathbf{X}) (\mathbf{S}^T + \mathbf{S} - \mathbf{S}^T \mathbf{S}) \Phi(\mathbf{X})^T \mathbf{w} = \lambda \Phi(\mathbf{X}) \Phi(\mathbf{X})^T \mathbf{w} \quad (21)$$

$\mathbf{w}$  可以表示为

$$\mathbf{w} = \Phi(\mathbf{X}) \boldsymbol{\alpha} \quad (22)$$

将式(22)代入式(21),得到

$$\Phi(\mathbf{X}) (\mathbf{S}^T + \mathbf{S} - \mathbf{S}^T \mathbf{S}) \Phi(\mathbf{X})^T \Phi(\mathbf{X}) \boldsymbol{\alpha} = \lambda \Phi(\mathbf{X}) \Phi(\mathbf{X})^T \Phi(\mathbf{X}) \boldsymbol{\alpha} \quad (23)$$

式(23)等价于

$$\Phi(\mathbf{X})^T \Phi(\mathbf{X}) (\mathbf{S}^T + \mathbf{S} - \mathbf{S}^T \mathbf{S}) \Phi(\mathbf{X})^T \Phi(\mathbf{X}) \boldsymbol{\alpha} = \lambda \Phi(\mathbf{X})^T \Phi(\mathbf{X}) \Phi(\mathbf{X})^T \Phi(\mathbf{X}) \boldsymbol{\alpha} \quad (24)$$

从而,广义本征方程可以化简为

$$K(\mathbf{S}^T + \mathbf{S} - \mathbf{S}^T \mathbf{S}) K \boldsymbol{\alpha} = \lambda K^2 \boldsymbol{\alpha} \quad (25)$$

求解式(25),取得  $d$  个最大特征值所对应的投影向量  $\boldsymbol{\alpha}_i (i=1, \dots, d)$ 。为了归一化 KSPP 的投影向量  $\mathbf{w}_i = \Phi(\mathbf{X}) \boldsymbol{\alpha}_i (i=1, \dots, d)$ , 就要使得

$$\mathbf{w}_i^T \mathbf{w}_i = \boldsymbol{\alpha}_i^T \Phi(\mathbf{X})^T \Phi(\mathbf{X}) \boldsymbol{\alpha}_i = \boldsymbol{\alpha}_i^T K \boldsymbol{\alpha}_i = 1 \quad (26)$$

则

$$\mathbf{w}_i = \frac{\Phi(\mathbf{X}) \boldsymbol{\alpha}_i}{\sqrt{\boldsymbol{\alpha}_i^T K}} \quad (27)$$

高维特征空间样本  $\Phi(\mathbf{x})$ , 经过 KSPP 投影后的第  $i$  个新特征为

$$\mathbf{w}_i^T \Phi(\mathbf{x}) = \left( \frac{\Phi(\mathbf{X}) \boldsymbol{\alpha}_i}{\sqrt{\boldsymbol{\alpha}_i^T K}} \right)^T \Phi(\mathbf{x}) = \frac{\boldsymbol{\alpha}_i^T \Phi(\mathbf{X})^T \Phi(\mathbf{x})}{\sqrt{\boldsymbol{\alpha}_i^T K}} = \frac{\boldsymbol{\alpha}_i^T}{\sqrt{\boldsymbol{\alpha}_i^T K}} \begin{bmatrix} K(\mathbf{x}_1, \mathbf{x}) \\ K(\mathbf{x}_2, \mathbf{x}) \\ \vdots \\ K(\mathbf{x}_N, \mathbf{x}) \end{bmatrix} \quad (28)$$

综上所述,KSPP 算法可归纳为如下步骤:

(1) 解约束优化问题(15),求出核稀疏表示系数  $\tilde{\mathbf{s}}_i (i=1, \dots, N)$ , 从而得到核稀疏重构的邻接矩阵

$$\mathbf{S} = [\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \dots, \tilde{\mathbf{s}}_N]^T。$$

(2) 求解广义本征方程(25)对应的前  $d$  个最大本征值所对应的本征向量  $\boldsymbol{\alpha}_i (i=1, \dots, d)$ 。

$$(3) \boldsymbol{\alpha} = \left[ \frac{\boldsymbol{\alpha}_1}{\sqrt{\boldsymbol{\alpha}_1^T \mathbf{K}}, \frac{\boldsymbol{\alpha}_2}{\sqrt{\boldsymbol{\alpha}_2^T \mathbf{K}}, \dots, \frac{\boldsymbol{\alpha}_d}{\sqrt{\boldsymbol{\alpha}_d^T \mathbf{K}}} \right]。$$

对于样本  $\mathbf{x}$ , 其新的 KSP 特征为  $\boldsymbol{\alpha}^T \begin{bmatrix} K(\mathbf{x}_1, \mathbf{x}) \\ K(\mathbf{x}_2, \mathbf{x}) \\ \vdots \\ K(\mathbf{x}_N, \mathbf{x}) \end{bmatrix}$ 。

#### 1.4 核稀疏保持典型相关分析(KSPCCA)

典型相关分析(CCA)<sup>[6]</sup>的优化准则函数为

$$J_{\text{CCA}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{\boldsymbol{\alpha}^T \mathbf{S}_{xy} \boldsymbol{\beta}}{\sqrt{\boldsymbol{\alpha}^T \boldsymbol{\alpha} \cdot \boldsymbol{\beta}^T \boldsymbol{\beta}}} \quad (29)$$

式中:  $\mathbf{S}_{xx}, \mathbf{S}_{yy}$  分别表示  $x, y$  两组随机变量组内的协方差矩阵,  $\mathbf{S}_{xy}$  表示  $x, y$  两组随机变量之间协方差矩阵, 则 CCA 可表述为如下优化问题的解, 即

$$\begin{aligned} & \max_{\boldsymbol{\alpha}, \boldsymbol{\beta}} \boldsymbol{\alpha}^T \mathbf{S}_{xy} \boldsymbol{\beta} \\ \text{s. t. } & \boldsymbol{\alpha}^T \mathbf{S}_{xx} \boldsymbol{\alpha} = \boldsymbol{\beta}^T \mathbf{S}_{yy} \boldsymbol{\beta} = 1 \end{aligned} \quad (30)$$

核稀疏保持典型相关分析(KSPCCA)就是将核稀疏保持(KSP)技术, 引入到 CCA 算法中。首先给定  $N$  对样本  $(x_i, y_i)_{i=1}^N \in R^p \times R^q$ , 设有非线性映射  $\Phi: x \rightarrow \Phi(x)$  和  $\psi: y \rightarrow \psi(y)$ , 将原始样本映射到高维特征空间, 记  $\Phi(\mathbf{X}) = [\Phi(x_1), \dots, \Phi(x_N)]$  与  $\psi(\mathbf{Y}) = [\psi(y_1), \dots, \psi(y_N)]$ , 分别表示高维特征空间中的样本矩阵, 则 KSPCCA 的解向量  $\mathbf{w}_\Phi$  和  $\mathbf{w}_\psi$  可以表示为两组样本的线性组合, 即  $\mathbf{w}_\Phi = \Phi(\mathbf{X})\boldsymbol{\alpha}$ ,  $\mathbf{w}_\psi = \psi(\mathbf{Y})\boldsymbol{\beta}$  同时成立, 这里的  $\boldsymbol{\alpha}$  和  $\boldsymbol{\beta}$  不同于式(30)中的  $\boldsymbol{\alpha}, \boldsymbol{\beta}$  (即原始输入空间的解向量), 为了简便, 就依旧采用这样的记号。由式(30)可知, KSPCCA 的目标函数可以表示为

$$\begin{aligned} \mathbf{w}_\Phi^T \Phi(\mathbf{X}) \psi(\mathbf{Y})^T \mathbf{w}_\psi &= (\Phi(\mathbf{X})\boldsymbol{\alpha})^T \Phi(\mathbf{X}) \psi(\mathbf{Y})^T (\psi(\mathbf{Y})\boldsymbol{\beta}) = \\ & \boldsymbol{\alpha}^T \Phi(\mathbf{X})^T \Phi(\mathbf{X}) \Phi(\mathbf{X})^T \psi(\mathbf{Y})^T \psi(\mathbf{Y}) \boldsymbol{\beta} = \boldsymbol{\alpha}^T \mathbf{K}_x \mathbf{K}_y \boldsymbol{\beta} \end{aligned} \quad (31)$$

则目标函数可改为

$$\max_{\boldsymbol{\alpha}, \boldsymbol{\beta}} \mathbf{w}_\Phi^T \Phi(\mathbf{X}) \psi(\mathbf{Y})^T \mathbf{w}_\psi = \max_{\boldsymbol{\alpha}, \boldsymbol{\beta}} \boldsymbol{\alpha}^T \mathbf{K}_x \mathbf{K}_y \boldsymbol{\beta} \quad (32)$$

相应地, 由 CCA 的约束条件可得 KSPCCA 的约束条件  $\boldsymbol{\alpha}^T \mathbf{K}_x (\mathbf{I} - \mathbf{R}^T - \mathbf{R} + \mathbf{R}^T \mathbf{R}) \mathbf{K}_x \boldsymbol{\alpha} = 1, \boldsymbol{\beta}^T \mathbf{K}_y (\mathbf{I} - \mathbf{S}^T - \mathbf{S} + \mathbf{S}^T \mathbf{S}) \mathbf{K}_y \boldsymbol{\beta} = 1$ , 因此, KSPCCA 的约束优化问题可表示为

$$\begin{aligned} & \max_{\boldsymbol{\alpha}, \boldsymbol{\beta}} \boldsymbol{\alpha}^T \mathbf{K}_x \mathbf{K}_y \boldsymbol{\beta} \\ \text{s. t. } & \boldsymbol{\alpha}^T \mathbf{K}_x (\mathbf{I} - \mathbf{R}^T - \mathbf{R} + \mathbf{R}^T \mathbf{R}) \mathbf{K}_x \boldsymbol{\alpha} = 1 \\ & \boldsymbol{\beta}^T \mathbf{K}_y (\mathbf{I} - \mathbf{S}^T - \mathbf{S} + \mathbf{S}^T \mathbf{S}) \mathbf{K}_y \boldsymbol{\beta} = 1 \end{aligned} \quad (33)$$

式中:  $\mathbf{R} = [\mathbf{r}_1, \dots, \mathbf{r}_N]^T$  和  $\mathbf{S} = [\mathbf{s}_1, \dots, \mathbf{s}_N]^T$  分别是两组特征样本集  $X$  和  $Y$  的稀疏重构权重矩阵。

利用 Lagrange 乘子法, 不难得到如下关于  $\boldsymbol{\alpha}, \boldsymbol{\beta}$  的广义本征特征值问题

$$\begin{pmatrix} & \mathbf{K}_x \mathbf{K}_y \\ \mathbf{K}_y \mathbf{K}_x & \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix} = \lambda \begin{pmatrix} \mathbf{K}_x (\mathbf{I} - \mathbf{R}^T - \mathbf{R} + \mathbf{R}^T \mathbf{R}) \mathbf{K}_x & \\ & \mathbf{K}_y (\mathbf{I} - \mathbf{S}^T - \mathbf{S} + \mathbf{S}^T \mathbf{S}) \mathbf{K}_y \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix} \quad (34)$$

因此只需利用式(34), 即可求解得  $\boldsymbol{\alpha}, \boldsymbol{\beta}$ , 之后取  $d$  个最大特征值所对应的特征向量  $\boldsymbol{\alpha}_i (i=1, \dots, d)$  和  $\boldsymbol{\beta}_i (i=1, \dots, d)$ 。KSPCCA 算法可归纳为如下步骤

(1) 解约束优化问题式(15),分别对两组样本集  $X$  和  $Y$  求核稀疏表示系数  $r_i (i=1, \dots, N)$  和  $s_i (i=1, \dots, N)$ , 从而分别得到两组样本核稀疏重构邻接矩阵  $R = [r_1, \dots, r_N]^T$  和  $S = [s_1, \dots, s_N]^T$ 。

(2) 求解广义本征方程(34)对应的前  $d$  个最大特征值所对应的特征向量  $\alpha_i (i=1, \dots, d)$  和  $\beta_i (i=1, \dots, d)$ 。

(3)  $\alpha = \left[ \frac{\alpha_1}{\sqrt{\alpha_1^T}}, \frac{\alpha_2}{\sqrt{\alpha_2^T}}, \dots, \frac{\alpha_d}{\sqrt{\alpha_d^T}} \right]$ ,  $\beta = \left[ \frac{\beta_1}{\sqrt{\beta_1^T}}, \frac{\beta_2}{\sqrt{\beta_2^T}}, \dots, \frac{\beta_d}{\sqrt{\beta_d^T}} \right]$ 。对于样本  $x$  和  $y$ , 其新的 KSPCC-CA 特征分别为

$$\alpha^T \begin{bmatrix} K_x(x_1, x) \\ K_x(x_2, x) \\ \vdots \\ K_x(x_N, x) \end{bmatrix} \quad \beta^T \begin{bmatrix} K_y(y_1, y) \\ K_y(y_2, y) \\ \vdots \\ K_y(y_N, y) \end{bmatrix} \quad (35)$$

## 2 实验与分析

### 2.1 手写体识别实验

本实验采用多特征手写体数据库 (Multiple feature database), 该数据库总共包括 0~910 个数字手写体, 总共包括 6 组不同的特征, 每类 200 个样本, 共 2 000 个样本。在实验中, 对于每种特征组合, 在每类数字手写体中随机选取 100 个样本数字作为训练样本, 余下的 100 个样本作为测试样本, 独立随机 10 次实验, 这样的随机试验独立进行 10 次, 记录其平均识别率。实验中, 对于 KSPCCA 中的核函数, 选取了高斯核  $K(x, y) = \exp(-\|x - y\|^2/t)$ 。这里选取的数据库的特征维数是 10, 从而作为高斯核的参数  $t$ , 即  $t=10$ 。将 KSPCCA 与 CCA<sup>[6]</sup>, SPCCA<sup>[14]</sup> 等算法做对比, 其结果如表 1 所示。

实验结果表明, 在不同的数据组合实验中, 共有 13 次组合方式的识别率 KSPCCA 优于 SPCCA 和 CCA; 相反, SPCCA 只有 2 次表现最优。这说明在多数情况下, KSPCCA 的识别能力优于 SPCCA 和 CCA, 同时, 也说明核方法对于线性学习器的性能具有一定的提升作用。

### 2.2 人脸识别实验

为了进一步验证 KSPCCA 的识别性能, 在 Yale 人脸数据库上进行人脸识别实验。Yale 人脸库数据集主要反映了人脸的表情、光照的不同, 该数据库总共包含有 15 个人, 每个人有 11 幅灰度, 在每人 11 幅图像中选取 6 幅作训练样本, 其余 5 幅作测试, 这样训练集的样本数为 90, 测试集样本数为 75。表 2 列出了 CCA, SPCCA 和 KSPCCA 3 种方法在 Yale 数据集上的识别率。从实验结果来看, 稀疏表示的方法明显增加了人脸识别的性能, 大大提高了人脸识别率; 然而核方法的加入, 虽然识别率有

表 1 多特征手写体数据库上的识别率

Tab. 1 Rate of recognition on multiple feature database

No. #	数据集组合方式		各种方法的识别率		
	X	Y	CCA	SPCCA	KSPCCA
1	mfeat_fac	mfeat_fou	0.876 5	0.952 0	0.973 0
2	mfeat_fac	mfeat_kar	0.929 8	0.959 0	0.964 0
3	mfeat_fac	mfeat_mor	0.813 8	0.869 0	0.918 0
4	mfeat_fac	mfeat_pix	0.914 5	0.970 0	0.976 0
5	mfeat_fac	mfeat_zer	0.863 3	0.924 0	0.935 0
6	mfeat_fou	mfeat_kar	0.925 6	0.968 0	0.941 0
7	mfeat_fou	mfeat_mor	0.783 5	0.827 0	0.897 0
8	mfeat_fou	mfeat_pix	0.835 7	0.970 0	0.914 0
9	mfeat_fou	mfeat_zer	0.843 6	0.844 0	0.876 0
10	mfeat_kar	mfeat_mor	0.805 7	0.903 0	0.959 0
11	mfeat_kar	mfeat_pix	0.932 5	0.963 0	0.981 0
12	mfeat_kar	mfeat_zer	0.946 5	0.965 0	0.963 5
13	mfeat_mor	mfeat_pix	0.814 5	0.933 0	0.961 0
14	mfeat_mor	mfeat_zer	0.763 3	0.812 0	0.847 0
15	mfeat_pix	mfeat_zer	0.815 5	0.960 0	0.969 0

表 2 Yale 数据集上的识别率

Tab. 2 Rate of recognition on data set of Yale face

识别方法	识别率
CCA	0.723 0
SPCCA	0.820 0
KSPCCA	0.839 0

所提高,但是没有明显的提升,这就说明选取核与核参数是一个值得深入探讨的问题,即如何根据样本集的复杂程度(如样本容量、特征维数以及类别数等)并结合数据集的先验知识(如人脸的表情、光照等)启发式地选取合适的核与核参数,如果采用不适当的核反而会造造成过拟合而导致识别率下降。

### 3 结束语

本文中的 KSPCCA 的实验采用的是高斯核函数,实际上也可以使用其他形式的核函数,如多项式核、Sigmoid 核等,核与核参数的选取是数据依赖的,与样本集的复杂度以及数据集的先验知识有关。尽管已经有很多这方面的探索,然而目前核与核参数的选取仍然是一个悬而未决的问题,因此还需进一步思考和探索。

#### 参考文献:

- [1] Wright J, Yang A Y, Ganesh A, et al. Robust face recognition via sparse representation[J]. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2009, 31(2): 210-227.
- [2] Qiao L S, Chen S C, Tan X Y. Sparsity preserving projections with application to face recognition[J]. *Pattern Recognition*, 2010, 43(1): 331-341.
- [3] 侯书东, 孙权森. 稀疏保持典型相关分析及在特征融合中的应用[J]. *自动化学报*, 2012, 38(4): 659-665.  
Hou Shudong, Sun Quansen. Sparsity preserving canonical correlation analysis with application in feature fusion[J]. *Acta Automatica Sinica*, 2012, 38(4): 659-665.
- [4] 侯书东. 基于相关投影分析的特征提取研究及在图像识别中的应用[D]. 南京: 南京理工大学, 2012.  
Hou Shudong. Feature extraction based on correlation projection analysis and its application to image recognition[D]. Nanjing: Nanjing University of Science and Technology, 2012.
- [5] Shen Xiaobo, Sun Quansen, Yuan Yunhao. Orthogonal canonical correlation analysis and its application in feature fusion [M]. *Information Fusion (FUSION)*, 2013 16th International Conference on. [S.l.]: IEEE, 2013.
- [6] Melzer T, Reiter M, Bischof H. Appearance models based on kernel canonical correlation analysis[J]. *Pattern Recognition*, 2003, 36(9): 1961-1971.
- [7] Mairal J, Bach F, Ponce J, et al. Discriminative learned dictionaries for local image analysis[J]. *Proceedings of IEEE CVPR*, 2008, 413(2): 1-8.
- [8] Mairal J, Bach F, Ponce J, et al. Supervised dictionary learning[J]. *Proceedings of NIPS*. 2009, 21: 1033-1040.
- [9] Schölkopf B, Smola A, Müller K-R. Nonlinear component analysis as a kernel eigenvalue problem[J]. *Neural Computation*, 1998, 10: 1299-1319.
- [10] Mika S, Ratsch G, Weston J, et al. Fisher discriminant analysis with kernels[J]. *IEEE Neural Networks for Signal Processing Workshop*. 1999, 9: 41-48.
- [11] Baudat G, Anouar F. Generalized discriminant analysis using a kernel approach[J]. *Neural Computation*, 2000, 12(10): 2385-2404.
- [12] Cover T M. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition[J]. *IEEE Transactions on Electronic Computers*, 1965, 14: 326-334.
- [13] Schölkopf B, Smola A, Müller K-R. Nonlinear component analysis as a kernel eigenvalue problem[J]. *Neural Computation*, 1998, 10: 1299-1319.
- [14] 侯彬. 基于稀疏表示的典型相关分析算法研究[D]. 南京: 南京理工大学, 2013.  
Hou bin. Canonical correlation analysis algorithm based on sparse representation[D]. Nanjing: Nanjing University of Science and Technology, 2013.

#### 作者简介:



张荣(1989-),男,硕士研究生,研究方向:图像处理与识别, E-mail: 15850564451@126.com.



孙权森(1963-),男,教授,研究方向:模式识别理论与应用、遥感信息系统理论与应用、图形图像技术与应用、多媒体信息处理技术以及医学影像处理与分析。

