中国科学: 信息科学 2016 年 第 46 卷 第 2 期: 179–192
SCIENTIA SINICA Informationis
《中国科学》杂志社 SCIENCE CHINA PRESS

论 文

空间结构化欧拉核及其应用

刘爽, 陈松灿*

南京航空航天大学计算机科学与技术学院, 南京 210016
* 通信作者. E-mail: s.chen@nuaa.edu.cn

收稿日期: 2015–05–20; 接受日期: 2015–06–08; 网络出版日期: 2016–01–22
国家自然科学基金 (批准号: 61170151) 和高等学校博士点基金 (批准号: 20133218110032) 资助项目

摘要 自 Yang 等将向量主成分分析 (1D-PCA) 推广至面向图像的 2D-PCA 以来, 众多基于向量的 1D 形式算法被相继推广至对应的 2D 形式. 因利用了图像/矩阵空间特定的结构先验知识, 在处理矩阵型数据时 2D 算法自然导致了较传统 1D 算法更好的学习性能, 这与 “没有免费午餐定理” 相符. 但 2D 算法仍有其不足, 主要表现在: (1) 2D 算法几乎都是线性的, 因此对非线性数据处理的能力有限; (2) 2D 算法的空间结构信息利用仍不够充分. 针对第一个不足, 本文利用核方法进行改进, 但相对于 1D 算法, 2D 算法因难以利用表示定理而导致核化困难, 因此本文绕过表示定理, 通过改变度量获得一个简洁的核化方法; 针对第二个不足, 本文采用在核空间直接对空间结构信息进行补偿的方法. 但这需要在核空间中描述矩阵数据的空间结构, 如果使用隐式核进行核化可能会导致矩阵数据空间结构扭曲, 从而使对空间结构信息的描述和利用变得困难; 如果使用显式核进行核化, 会导致维数灾难而失去隐式核的优势. 若核映射是一个显式、等维且各分量非耦合的映射, 就能自然地描述出矩阵数据在核空间中的结构. 幸运的是, 存在众多符合以上要求的显式核 (如 Hellinger 核和欧拉核) 和隐式加性核 (如 Intersection 核、JS 核和 χ^2 核) 的近似显式形式. 因欧拉核形式上的简洁性及其良好的行为, 本文以欧拉核作为样例, 首次尝试进行矩阵的核化及其在核空间的空间结构信息补偿. 尽管存在若干空间结构信息的补偿方法, 如空间结构信息约束、图像距离度量等, 本文围绕现有的图像欧氏距离加以阐述, 从而为矩阵或图像数据构建出对应的空间结构化欧拉核. 最后将其应用于典型 2D 算法并通过实验验证了其有效性.

关键词 2D 算法 核方法 空间结构信息 欧拉核 图像欧氏距离

1 引言

传统的模式识别算法都是基于向量的 1D 算法, 即使是具有结构的矩阵数据也要将其向量化后进行处理, 这就忽略了矩阵数据内在的空间结构信息. 因此, 对于矩阵数据为了利用数据内在的空间结构信息, 最简单的方法就是直接处理矩阵数据. Yang 等 [1] 首先将传统 PCA 推广到了 2D-PCA, 并在人脸识别上取得了较 PCA 算法更好的性能. 受 2D-PCA 启发, 众多基于向量的 1D 形式算法被相继推广至对应的 2D 形式, 如 2D-LDA[2], 基于矩阵度量学习的 KISSM[3] 等. 对于矩阵数据, 2D 算法之

引用格式: 刘爽, 陈松灿. 空间结构化欧拉核及其应用. 中国科学: 信息科学, 2016, 46: 179–192, doi: 10.1360/N112015-00111

所以很大程度上优于传统 1D 算法在于其利用了与模式相关的部分空间结构信息, 这也印证了 "没有免费午餐定理", 即只有利用更多与问题相关的信息才能获得更好的学习性能. 尽管如此, 2D 算法仍有其不足, 具体表现为: (1) 2D 算法几乎都是线性的, 因此对非线性数据处理能力有限; (2) 2D 算法的空间结构信息利用仍不够充分, 如 2D-PCA/LDA 只考虑了单纯行 (或列) 间的相关信息. 为了克服上述不足, 已有众多改进的 2D 算法.

针对第一个不足, 一种方法是直接对矩阵数据非线性化, 如矩阵指数 [4], 但这类做法相对固定, 很难适应学习任务的多样性. 另一种方法是针对矩阵数据核化, 但与 1D 算法可借助表示定理进而通过核方法实现非线性化相比, 2D 算法难以利用表示定理而较难核化. 目前矩阵核化的相关工作较少, 据我们检索所知, 仅有 2D-kLDA (2DKDA)[5] 和 2D-kPCA[6]. 其中 2D-kLDA 通过对数据矩阵做 SVD 分解后进行核化, 此方法不仅复杂, 且核化对象不再是原数据, 而是左右奇异向量矩阵, 已很难再描述原数据在核空间的空间结构信息. 相比 2D-kLDA 的核化, 2D-kPCA 相对简单, 但仅限于对矩阵行 (或列) 单独核化, 核化方法类似 1D 核化. 此两种针对矩阵数据的核化均不是严格从表示定理 (解能由训练样本集的线性组合表示, 又称对偶表示) 出发, 使核化不自然. 本文尝试对面向矩阵数据的学习算法 (包含 2D-PCA, 2D-LDA 等) 进行核化, 关键在于避开表示定理的限制, 直接通过改变度量实现简洁的核化. 考虑到矩阵数据内的空间结构信息, 所采用的映射应保持矩阵数据内在的结构, 以方便在核空间对矩阵数据的空间结构信息的利用. 但通过数据隐式映射实现的非线性核化常会导致矩阵数据空间结构的扭曲, 使对空间结构信息的描述和利用变得复杂; 另一方面, 如果使用升维的显式映射进行核化, 通常会导致维数灾难而失去核方法的优势, 因此我们期望核映射既是显式的又是等维和分量间非耦合的.

幸运的是存在众多符合以上要求的显式核 (如 Hellinger 核、欧拉核 [7]) 和部分隐式加性核的近似显式表示 [8](如 χ^2 核、JS 核和 Intersection 核). 此类核映射保留了原数据的空间结构, 使得在核空间中对空间结构信息的直接描述和补偿成为可能. 因欧拉核形式上的简洁及其良好的行为 [7], 本文工作以欧拉核为基础展开讨论. 欧拉核首先被用于欧拉 PCA[7] 和欧拉聚类 [9], 其核映射基于欧拉变换将原实数域空间映射到复数域空间. 虽然变换后数据间距离的计算在复数域中进行, 但最终结果是一个基于余弦的有界实数, 从而保证了鲁棒性 [7].

针对第二个不足, 已出现了众多改进方法. 其中一类方法为针对 2D 算法的改进算法, 如, 为了同时利用矩阵数据的行和列间的相关信息所提出的双边改进算法 2D2-PCA[10] 和 Ye 的 2D-LDA[11] 等; 为了更好地描述矩阵数据的局部空间信息而将矩阵数据分块重排的 Block PCA[12]. 另一类方法为对矩阵数据内在的空间结构信息补偿, 大致可分为正则化方法 (如 SSSL[13]) 和利用图像距离的方法 [14~18]. 此类正则化方法主要针对向量进行, 因而能进行核化, 但对矩阵数据的向量化使计算复杂度提高, 同时正则化本身也增加了计算复杂性, 如其因子的调节, 这就使此类方法的应用受限. 而降低矩阵数据处理复杂性的一个可行方法是直接使用 2D 数据, 但这又使 2D 算法面临难核化的困境. 另一方面, 利用图像距离的相关方法在计算与使用上都比较简单, 其实现思想是设计出一种能够反映矩阵数据空间结构信息的距离度量. 此类方法最终也归结到基于向量数据的计算, 典型的图像欧氏距离 (IMED)[14] 即为一种基于向量马氏距离的度量方法. 但 Gao 等 [19] 利用了一个变换将基于向量的图像欧氏距离转化为基于矩阵形式的计算, 且这种变换可视为对数据的预处理, 而后用于 2D 算法. 但他们的工作并未涉及矩阵数据的非线性核化. 由此可见空间结构信息补偿虽然方法众多, 但本身并未解决矩阵数据的核化. 由此可见, 矩阵核化与矩阵空间信息补偿是两个不同的方面, 对矩阵数据的核化及其在核空间对空间结构信息的补偿可望提升算法的学习性能. 在具体的实现过程中, 考虑到核映射和空间结构信息补偿的多样性, 本文选用了欧拉变换/函数和图像欧氏距离作为典型加以阐述, 从而为

矩阵数据构建出了对应的空间结构化欧拉核. 最后将其应用于 2D-PCA 和 2D-LDA 等典型 2D 算法并通过实验验证了其有效性.

本文的第 2 节对欧拉核和图像欧氏距离进行简短的回顾; 第 3 节为本文的重点 —— 空间结构化欧拉核及其应用, 以 2D-PCA 和 2D-LDA 为例说明; 第 4 节以大量的实验验证改进算法的有效性; 第 5 节为本文的总结.

2 相关知识

2.1 欧拉核函数

欧拉核 [7,9] 是针对向量数据的核映射, 其最初思想是构造一个基于余弦的有界度量. 设原始向量数据为 $\boldsymbol{I}_k \in \mathbb{R}^p$, 首先将 $\boldsymbol{I}_k$ 各个分量的值归一化到 $[0,1]$, 记为 $\boldsymbol{x}_k$. 定义度量

$$d(\boldsymbol{x}_l,\boldsymbol{x}_q)=\sum_{c=1}^{p}\{1-\cos(\alpha\pi(x_l(c)-x_q(c)))\}, \tag{1}$$

其中, $x_k(i)$ 表示向量 $\boldsymbol{x}_k$ 的第 i 个分量. 实验已证明了此度量具有良好的鲁棒性 [7]. 巧合的是, 此度量等价于样本在欧拉核空间 2 范数 (F 范数) 的平方. 对向量数据 $\boldsymbol{x}_k$ 定义欧拉核映射

$$\boldsymbol{z}_k=\varphi(\boldsymbol{x}_k)=\frac{1}{\sqrt{2}}\begin{bmatrix}\mathrm{e}^{\mathrm{i}\alpha\pi x_k(1)}\\ \cdots\\ \mathrm{e}^{\mathrm{i}\alpha\pi x_k(q)}\end{bmatrix}=\frac{1}{\sqrt{2}}\mathrm{e}^{\mathrm{i}\alpha\pi\boldsymbol{x}_k}, \tag{2}$$

当 $0<\alpha<2$ 时, 此映射为双射. 设 $\boldsymbol{\theta}_k=\alpha\pi\boldsymbol{x}_k$, 则距离度量为

$$\begin{aligned}\|\boldsymbol{z}_l-\boldsymbol{z}_q\|_F^2&=\frac{1}{2}\left\|(\cos(\alpha\pi\boldsymbol{x}_l)+\mathrm{i}\sin(\alpha\pi\boldsymbol{x}_l))-(\cos(\alpha\pi\boldsymbol{x}_q)+\mathrm{i}\sin(\alpha\pi\boldsymbol{x}_q))\right\|_F^2\\&=\sum_{c=1}^{p}\{1-\cos(\theta_l(c)-\theta_q(c))\}=d(\boldsymbol{x}_j,\boldsymbol{x}_q).\end{aligned} \tag{3}$$

由 (3) 式可知此度量有界, 从而保证了欧拉核的优良性质.

2.2 图像欧氏距离 (IMED)

图像欧氏距离 [14] 是一种针对向量数据的马氏距离度量, 其思想是在距离度量中考虑原矩阵数据中各分量的位置信息, 从而补偿矩阵数据拉成向量后丢失的空间结构信息. 设图像为 $X_k\in\mathbb{R}^{m\times n}$, 则相应度量矩阵 $G\in\mathbb{R}^{mn\times mn}$ 定义如下:

$$\begin{aligned}&G=\{g_{ij}\},\\&g_{ij}=\frac{1}{2\pi\sigma^2}\exp(-[(t-t')^2+(s-s')^2]/2\sigma^2),\end{aligned} \tag{4}$$

其中, $i=(t-1)m+s; j=(t'-1)m+s'; t,t'=1,2,\ldots,n; s,s'=1,2,\ldots,m$, t 和 t' 对应列号, s 和 s' 对应行号. 显然度量矩阵 G 正定, 由此定义图像的欧氏距离

$$d_{\mathrm{imed}}(X_l,X_q)=\mathrm{vec}(X_l-X_q)^{\mathrm{T}}G\mathrm{vec}(X_l-X_q)=(G^{0.5}\mathrm{vec}(X_l-X_q))^{\mathrm{T}}(G^{0.5}\mathrm{vec}(X_l-X_q)). \tag{5}$$

另一方面, 度量矩阵 G 具有明显的结构, 可以分解为行方向的关系矩阵和列方向的关系矩阵的 Kronecher 积, 分别代表行和列之间的空间结构关系, 即

$$\begin{aligned} G &= G_1 \otimes G_2, \\ G_1(t,t') &= \frac{1}{2\pi\sigma^2}\exp(-(t-t')^2/2\sigma^2),\ t,t'=1,2,\ldots,n, \\ G_2(s,s') &= \frac{1}{2\pi\sigma^2}\exp(-(s-s')^2/2\sigma^2),\ s,s'=1,2,\ldots,m. \end{aligned} \tag{6}$$

Gao 等 [19] 利用度量矩阵 G 的结构性, 将图像欧氏距离等价转化为相应的矩阵形式. 根据式

$$\mathrm{vec}(ABC) = (C^{\mathrm{T}} \otimes A)\mathrm{vec}(B), \tag{7}$$

则 (5) 式可写作

$$d_{\mathrm{imed}}(X_j, X_q) = \left\|G_2^{0.5}(X_j - X_q)G_1^{0.5}\right\|_F^2 = \left\|G_2^{0.5}X_jG_1^{0.5} - G_2^{0.5}X_qG_1^{0.5}\right\|_F^2. \tag{8}$$

这样, 可将 $G_2^{0.5}X_kG_1^{0.5}$ 看做是针对矩阵数据 X_k 的预处理且保持了原有矩阵数据的维数. 这样就能够使用 2D 算法进行处理. 此种变换将图像欧氏距离引入到了 2D 算法中, 对矩阵数据内在的空间结构信息进行补偿, 从而获得更好的学习性能.

3 空间结构化欧拉核及其对 2D 算法的改进

本节主要工作为, 从度量出发, 将欧拉核推广到 2D 领域, 并在此基础上进行空间信息补偿, 从而定义出空间结构化欧拉核. 然后将空间结构化欧拉核应用于典型 2D 算法的改进.

3.1 空间结构化欧拉核

首先将欧拉核映射推广到 2D 形式. 将矩阵数据 (如图像) 归一化到 $[0,1]$, 记为 X_k, 定义矩阵数据度量

$$d(X_l, X_q) = \sum_{i=1}^{m}\sum_{k=1}^{n}\{1-\cos(\alpha\pi X_l(i,k) - \alpha\pi X_q(i,k))\}, \tag{9}$$

其中, $X_k(i,j)$ 表示数据矩阵第 i 行第 j 列的分量. α 为超参数. 参考 2.1 小节内容, 根据 (9) 式所定义的度量, 定义矩阵数据欧拉核映射

$$Z_k' = \Phi(X_k) = \frac{1}{\sqrt{2}}\begin{bmatrix} \mathrm{e}^{\mathrm{i}\alpha\pi X_k(1,1)} & \cdots & \mathrm{e}^{\mathrm{i}\alpha\pi X_k(1,n)} \\ \vdots & \ddots & \vdots \\ \mathrm{e}^{\mathrm{i}\alpha\pi X_k(m,1)} & \cdots & \mathrm{e}^{\mathrm{i}\alpha\pi X_k(m,n)} \end{bmatrix} = \mathrm{e}^{\mathrm{i}\alpha\pi X_k}. \tag{10}$$

当 $0<\alpha<2$ 时, 此映射为双射. 令 $\Theta_k = \alpha\pi X_k$, 此时在核空间基于 F 范数的距离度量为

$$\begin{aligned} \left\|Z_l' - Z_q'\right\|_F^2 &= \frac{1}{2}\left\|(\cos(\alpha\pi X_l) + \mathrm{i}\sin(\alpha\pi X_l)) - (\cos(\alpha\pi X_q) + \mathrm{i}\sin(\alpha\pi X_q))\right\|_F^2 \\ &= \sum_{i=1}^{m}\sum_{k=1}^{n}\{1-\cos(\Theta_l(i,k) - \Theta_q(i,k))\}. \end{aligned} \tag{11}$$

由 (9)~(11) 式可知, 虽然是从度量角度出发, 但欧拉映射结果确实为一个核. 由 (10) 式可知, 欧拉核映射之后矩阵数据空间结构不变, 在此基础上引入图像欧氏距离, 根据 (7) 式定义空间结构化欧拉核

$$Z_k^{''}=\mathrm{H}(X_k)={G_2}^{0.5}\Phi(X_k){G_1}^{0.5}. \tag{12}$$

由 (6), (7), (10) 和 (12) 式可知, 此时距离度量为

$$\begin{aligned}
&\left\|Z_l^{''}-Z_q^{''}\right\|_F^2\\
&=\mathrm{vce}\Big(Z_l^{'}-Z_q^{'}\Big)^{\mathrm{H}}G\mathrm{vec}\left(Z_l^{'}-Z_q^{'}\right)\\
&=\frac{1}{2}\times\sum_{i=1}^{p}\sum_{k=1}^{p}g_{ik}(\cos(\theta_j(i))+i\sin(\theta_j(k))-\cos(\theta_q(i))-i\sin(\theta_q(i)))^H\\
&\quad\times(\cos(\theta_j(i))+i\sin(\theta_j(k))-\cos(\theta_q(k))-i\sin(\theta_q(k)))\\
&=\frac{1}{2}\times\sum_{i=1}^{p}\sum_{k=1}^{p}g_{ik}\left(\cos(\theta_l(i)-\theta_l(k))-\cos(\theta_l(i)-\theta_q(k))-\cos(\theta_q(i)-\theta_l(k))+\cos(\theta_q(i)-\theta_q(k))\right),
\end{aligned}\tag{13}$$

其中, $p=m\times n;\theta_k=\alpha\pi\times\mathrm{vec}(X_k)$. 经过 (12) 式核映射, 在计算空间结构化核空间的 F 范数相当于在欧拉核空间使用图像欧氏距离. 因在核空间嵌入矩阵数据内在的空间结构信息, 此方法称为空间结构化欧拉核.

3.2　空间结构化欧拉核针对 2D 算法的改进

3.1 小节提出的二维欧拉核和空间结构化欧拉核能够适用于多种 2D 算法的改进, 如 2D-LDA, KISSM, 2D-PCA 等. 这里主要以 2D-PCA 和 2D-LDA 加以说明. 为了使表达更加简洁清楚, 欧拉 2D-PCA 和空间结构化欧拉 2D-PCA 分别简称为 2D-EPCA 和 2D-IEPCA. 欧拉 2D-LDA 和空间结构化欧拉 2D-LDA 分别简称为 2D-ELDA 和 2D-IELDA.

3.2.1　2D-EPCA 和 2D-IEPCA 特征提取算法

由于欧拉核和空间结构化欧拉核映射是一个显式映射, 可直接使用 2D-PCA 算法运算, 因此计算速度相对于之前传统的核 2D-PCA 有了较大的提高. 而且核空间向原空间的映射只需通过复角值的计算, 数据重建简便. 另外 2D-PCA 算法是否进行中心化对结果影响不大, 因此在 2D-EPCA 和 2D-IEPCA 相关算法中没有进行数据的中心化. 2D-EPCA 和 2D-IEPCA 算法详见 "算法 1", 特征提取和数据重建算法详见 "算法 2".

3.2.2　2D-EPCA 和 2D-IEPCA 原像去噪算法

虽然 3.2.1 小节中数据的恢复可以看成一种去噪算法, 但并不是直接对原空间数据去噪. 因此本小节提出一种直接对原图像 (pre-image) 去噪算法. 由于欧拉核和空间结构化欧拉核均为显式映射, 因此为在原空间数据去噪带来了很大便利. 已经有人提出对于核 PCA(KPCA) 去噪的优化目标与算法 [20], 仿照核 PCA 的去噪, 对 2D 核 PCA 的去噪采用优化以下函数目标来实现

$$\tilde{X}=\arg\min_{\breve{X}}\left\|\phi(\breve{X})-BB^{\mathrm{T}}\phi(X)\right\|_F^2. \tag{14}$$

算法 1: 2D-EPCA 和 2D-IEPCA 算法

输入: N 个样本矩阵 $X_i, i=1,2,\ldots,N$, 投影矩阵行数 m, 超参数 α;

输出: 投影矩阵 B

1: 将 N 个样本数据 X_i 的各个数据点的值归一化到 $[0,1]$, 做核映射, 得到 Z_i:

  1)2D-EPCA 使用 (10) 式;

  2)2D-IEPCA 使用 (12) 式.

2: 计算所有数据散度矩阵:

  $G=\frac{1}{N}\sum\limits_{i=1}^{N}Z_iZ_{\mathrm{i}}^{\mathrm{T}}$.

3: 计算 G 的最大的 m 个特征值, 得到 Σ.

4: 计算最大的 m 个特征值对应的特征向量, 排列成矩阵后得到投影矩阵 B.

算法 2: 特征提取和数据重建算法

输入: 投影矩阵 B, 超参数 α, 数据 X;

输出: 核空间降维后的数据矩阵 Y, 原空间重建数据 $\tilde{X}$;

1: 将数据 X 的各个数据点的值归一化到 $[0,1]$, 做核映射, 得到 Z:

  1)2D-EPCA 使用 (10) 式;

  2)2D-IEPCA 使用 (12) 式.

2: 得到原数据在核空间的投影 $Y=BZ$.

3: 使用 $\tilde{Z}=B^{\mathrm{H}}\mathrm{Y}$ 在核空间重建数据.

4: 使用以下方法将重建后数据映射到原空间:

  1) 欧拉核: $\tilde{X}_i=\frac{\angle\tilde{Z}_i}{\alpha\pi}$;

  2) 空间结构化欧拉核: $\tilde{X}_i=\frac{\angle{G_2}^{-0.5}\tilde{Z}_i{G_1}^{-0.5}}{\alpha\pi}$.

其中, $\angle$ 表示计算矩阵数据每个数据点的复角值.

(1) 2D-EPCA 原像去噪算法. 假设待去噪数据为 X_k, 根据 (14) 式, 优化目标如下:

$$\begin{aligned}\tilde{X}_k&=\arg\min_{\breve{X}_k}||\Phi(\breve{X}_k)-BB^{\mathrm{H}}\Phi(X_k)||_F^2\\&=\arg\max_{\breve{X}_k}\mathrm{Re}(\mathrm{tr}(\Phi(\breve{X}_k)^{\mathrm{H}}BB^{\mathrm{H}}\Phi(X_k)))\\&=\arg\max_{\breve{X}_k}\sum_{i=1}^{m}\sum_{j=1}^{n}(\cos(\alpha\pi\breve{X}_k(i,j))\mathrm{Re}(T(i,j))+\sin(\alpha\pi\breve{X}_k(i,j))\mathrm{Im}(T(i,j))\\&=\arg\max f(\breve{X}_k),\end{aligned}\tag{15}$$

其中, $T=BB^{\mathrm{H}}\Phi(X_k)$. 由 (15) 式可知, $f(\breve{X}_k)$ 为多项连加, 且每一项均为非凸函数 (实际上是周期函数), 因此 $f(\breve{X}_k)$ 非凸. 一般对其优化仅能获得局部最优解, 但幸运的是 $\breve{X}_k$ 的所有变量间无交叉且不相关, 因此, 对任意单一变量, 只需使 $\cos(\alpha\pi\breve{X}_k(i,j))\mathrm{Re}(T(i,j))+\sin(\alpha\pi\breve{X}_k(i,j))\mathrm{Im}(T(i,j))$ 最大便能保证 $f(\breve{X}_k)$ 获得最大值. 显然, $\cos(\alpha\pi\breve{X}_k(i,j))\mathrm{Re}(T(i,j))+\sin(\alpha\pi\breve{X}_k(i,j))\mathrm{Im}(T(i,j))$ 是一个周期函数, 其极值点即为全局最优点, 因此可以通过梯度下降法求解, 且能够保证梯度下降法收敛到全局最

算法 3: 2D-ELDA 和 2D-IELDA 算法

输入: 分别属于 C 个类的 N 个样本数据矩阵 X_i, $i=1,2,....,N$, 投影矩阵行数 m, 超参数 α;

输出: 投影矩阵 B

1: 将 N 个样本数据 X_i 的各个数据点的值归一化到 $[0,1]$, 通过核映射, 得到 Z_i:

1)2D-ELDA 使用 (10) 式;

2)2D-IELDA 使用 (12) 式.

2: 计算类内散度矩阵:

$S_w=\frac{1}{N}\sum\limits_{j=1}^{C}\sum\limits_{X_k\in T_j}\left(Z_k-\bar{Z}_j\right)\left(Z_k-\bar{Z}_j\right)^{\mathrm{T}}$.

3: 计算类间散度矩阵:

$S_b=\sum\limits_{j=1}^{C}\left(\bar{Z}_j-\bar{Z}\right)\left(\bar{Z}_j-\bar{Z}\right)^{\mathrm{T}}$.

4: 计算 $S_W^{-1}S_b$ 的最大的 m 个特征值对应的特征向量, 排列成矩阵后获得投影矩阵 B.

其中, T_j 代表第 j 类, $\bar{Z}_j$ 代表第 j 类样本均值, $\bar{Z}$ 代表总体样本均值.

优点. $f(\breve{X}_k)$ 对每一个数据分量求导

$$\frac{\partial f(\breve{X}_k)}{\breve{X}_k(i,j)}=\alpha\pi(-\sin(\breve{X}_k(i,j))\mathrm{Re}(T(i,j))+\cos(\breve{X}_k(i,j))\mathrm{Im}(T(i,j))). \tag{16}$$

根据 (16) 式, 梯度方向为

$$\nabla f(\breve{X}_k)=\alpha\pi(-\mathrm{Im}(\Phi(\breve{X}_k))\odot\mathrm{Re}(T)+\mathrm{Re}(\Phi(\breve{X}_k))\odot\mathrm{Im}(T)). \tag{17}$$

(2) 2D-IEPCA 原像去噪算法. 优化目标

$$\begin{aligned}\tilde{X}_k&=\mathop{\arg\min}_{\breve{X}_k}||G_1^{0.5}\Phi(\breve{X}_k)G_2^{0.5}-BB^{\mathrm{H}}G_1^{0.5}\Phi(X_k)G_2^{0.5}||_F^2\\&=\mathop{\arg\min}_{\breve{X}_k}||G_1^{0.5}\Phi(\breve{X}_k)G_2^{0.5}||_F^2-2\mathrm{tr}(\mathrm{Re}(G_2^{0.5}\Phi(\breve{X}_k)^{\mathrm{H}}G_1^{0.5}BB^{\mathrm{H}}G_1^{0.5}\Phi(X_k)G_2^{0.5}))\\&=\mathop{\arg\min}_{\breve{X}_k}||G_1^{0.5}\Phi(\breve{X}_k)G_2^{0.5}||_F^2-2\mathrm{tr}(\mathrm{Re}(\Phi(\breve{X}_k)^{\mathrm{H}}G_1^{0.5}BB^{\mathrm{H}}G_1^{0.5}\Phi(X_k)G_2))\\&=\arg\min f'(\breve{X}_k),\end{aligned} \tag{18}$$

可用梯度下降法对 (18) 式进行求解, $f'(\breve{X}_k)$ 虽然是非凸函数, 但是有界, 因此梯度下降算法必然收敛, 为了简洁, 具体求解过程详见附录, 此处直接给出结果

$$\nabla f'(\breve{X}_k)=\alpha\pi\mathrm{Im}(\overline{\Phi(\breve{X}_k)}\odot(G_1\Phi(\breve{X}_k)G_2))+2\alpha\pi(\mathrm{Im}(\Phi(\breve{X}_k))\odot\mathrm{Re}(T')-\mathrm{Re}(\Phi(\breve{X}_k))\odot\mathrm{Im}(T')), \tag{19}$$

其中, $T'=G_1^{0.5}BB^{\mathrm{H}}G_1^{0.5}\Phi(X_k)G_2$, $\overline{\Phi(\breve{X}_k)}$ 为取矩阵每个分量共轭复数, $\odot$ 为两个矩阵对应分量相乘.

3.2.3 2D-ELDA 和 2D-IELDA 算法

由于欧拉核和空间结构化欧拉核映射是一个显式映射, 对于核空间数据可直接使用 2D-LDA[2] 算法进行降维运算. 为了与上文一致, 2D-LDA 算法采用左乘投影矩阵. 2D-ELDA 和 2D-IELDA 投影算法详见 “算法 3”, 特征提取算法详见 “算法 4”.

算法 4: 特征提取算法

输入: 投影矩阵 B, 超参数 α, 数据 X;

输出: 核空间降维后的数据矩阵 Y;

1: 将数据 X 的各个数据点的值归一化到 $[0,1]$, 通过核映射, 得到 Z:

1) 2D-ELDA 使用 (10) 式;

2) 2D-IELDA 使用 (12) 式.

2: 得到原数据在核空间投影 $Y = BZ$.

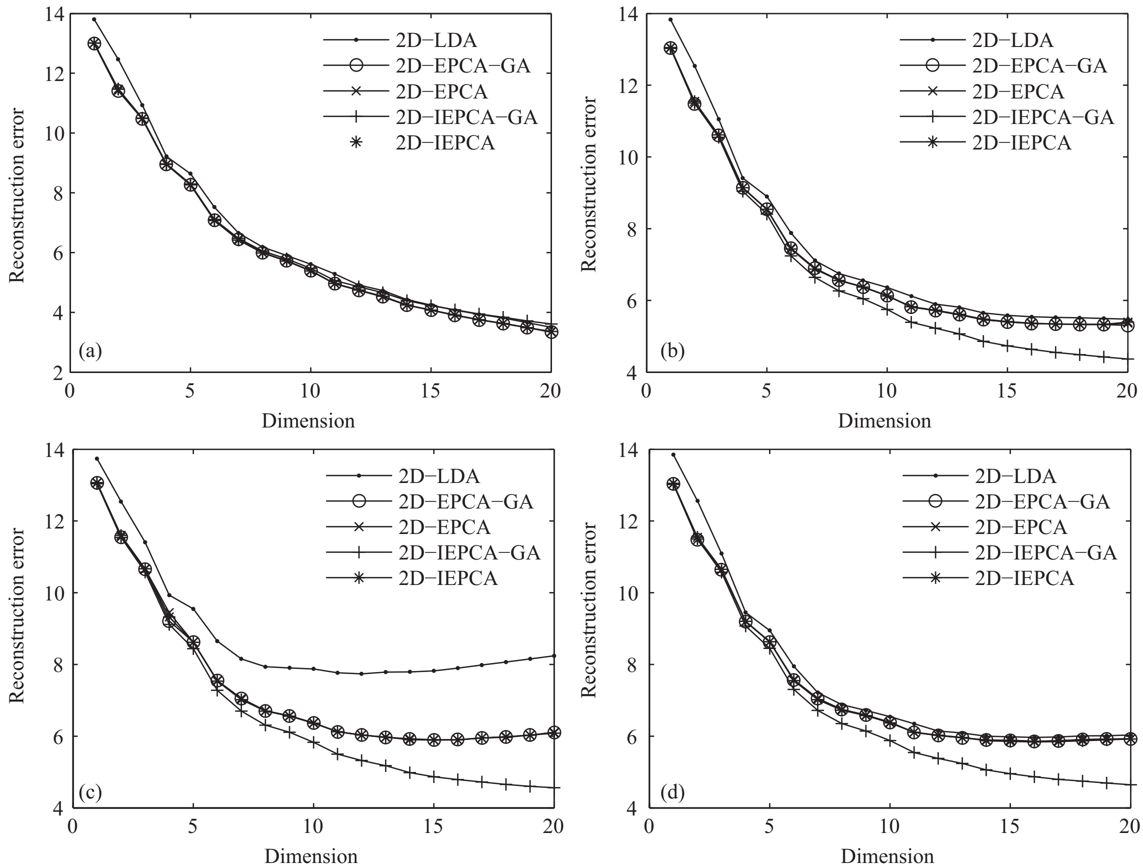


图 1 ORL 数据集重建误差

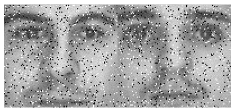
Figure 1 Reconstruction error on ORL dataset (a), with Gussian noise (b), salt & pepper noise (c) and speckle noise (d)

4 实验结果

本文实验分为两大部分, 第一部分为 2D-PCA, 2D-EPCA 和 2D-IEPCA 的数据重建 (去噪) 实验, 第二部分为 2D-LDA, 2D-ELDA 和 2D-IELDA 分类实验. 超参数选择为图像欧氏距离超参数 $\sigma = 1$, 欧拉核映射超参数 $\alpha = 1$. 实验中未对超参数进行调整, 即便如此, 改进后算法性能已优于原 2D 算法. 另外, 本文参考了 Gao 等 [19] 关于矩阵数据的图像欧氏距离的工作, 但其工作并没有被直接应用到 2D-PCA/2D-LDA 的改进中, 因此在实验中没有与 Gao 等的方法进行比较.

表 1　AR 数据集重建误差

Table 1　Reconstruction error on AR dataset

	Gaussian (0.01)		Speckle (0.04)		Salt & Pepper (0.1)	
	Image	Error	Image	Error	Image	Error
Original image		—		—		—
Noise image		—		—		—
2D-PCA		11.45		11.75		18.28
2D-EPCA-GA		11.16		11.49		14.15
2D-EPCA		11.16		11.5		14.31
2D-IEPCA-GA		**8.07**		**8.27**		**8.82**
2D-IEPCA		11.17		11.35		13.95

4.1　数据重建与去噪

本文中数据重建与去噪方法是相同的, 下文不再区分. 实验中主要对比有噪声和无噪声情况下的数据重建, 采用的重建方法为 2D-PCA, 2D-EPCA 中的普通算法 (见 3.1 节, 简称 2D-EPCA) 和梯度下降算法 (见 3.2 小节, 简称 2D-EPCA-GA) 以及 2D-IEPCA 中的普通算法 (见 3.1 小节, 简称 2D-IEPCA) 和梯度下降算法 (见 3.2 小节, 简称 2D-IEPCA-GA). 评价标准为还原后的图像与无噪声图像之差的 F 范数.

实验 1　采用 ORL 数据集, 共 40 个类, 每个类 10 张样本图像, 图像大小为 112×92. 其中每个类前 5 张图像 (共 200 张) 作为训练样本, 后 5 张 (共 200 张) 作为测试样本. 分别在无噪声数据、高斯噪声数据 (方差为 0.01)、椒盐噪声数据 (噪点比例为 0.1) 和乘性噪声数据 (方差为 0.04) 下进行数据的还原, 添加噪声时对训练样本和测试样本同时添加. 其结果如图 1 所示, 其中横轴表示降维维数 (投影矩阵行数), 纵轴为重建误差, 从实验结果可以看出, 在没有噪声的情况下, 2D-EPCA, 2D-EPCA-GA, 2D-IEPCA, 2D-IEPCA-GA 四种算法接近, 且略好于 2D-PCA 算法. 但当添加以上 3 种噪声后, 2D-IEPCA-GA 表现最为突出且大幅领先其他算法, 而 2D-EPCA, 2D-EPCA-GA, 2D-IEPCA 三种方法表现相近.

实验 2　为了进一步对比以上算法在有噪声情况下的数据重建能力 (去噪能力), 在 AR 数据集

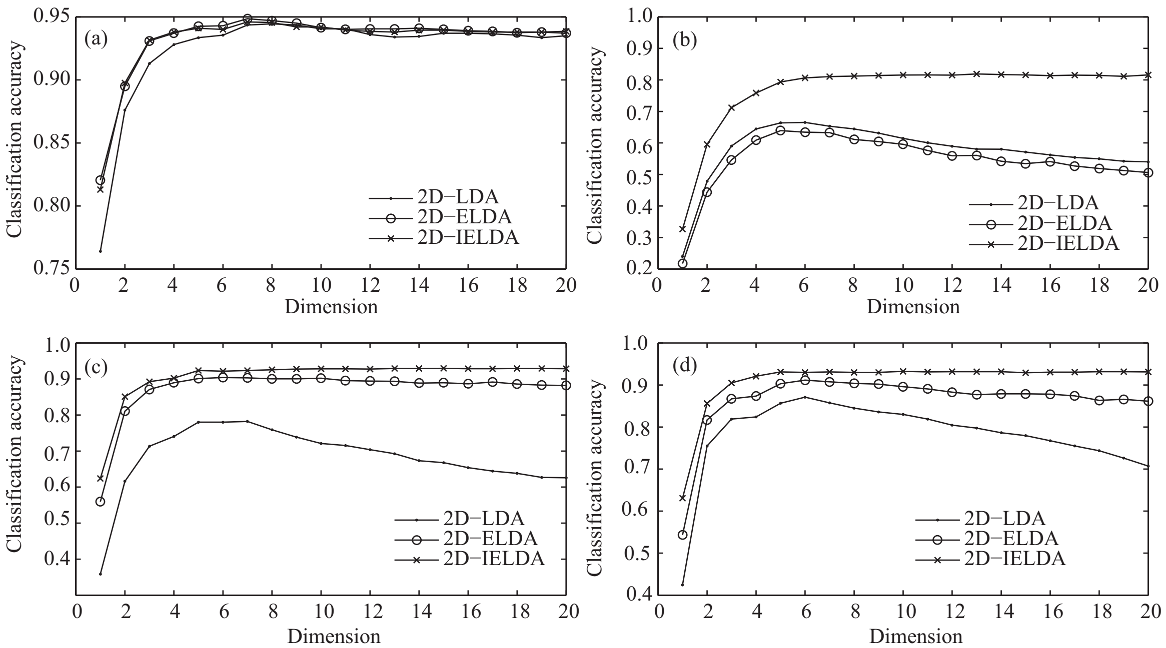


图 2 ORL 数据集分类准确率

Figure 2 Classification accuracy on ORL dataset (a), with Gaussian noise (b), salt & pepper noise (c) and speckle noise (d)

上进行进一步的对比. 实验中选用男女各 50 类的第一张照片, 截取 101×95 的人脸图像, 共 100 张. 以全部图像作为训练样本, 用前 5 张图像作为测试样本, 测试在高斯噪声数据 (方差为 0.01)、椒盐噪声数据 (噪点比例为 0.1) 和乘性噪声数据 (方差为 0.04) 下各算法的重建能力, 降维维数 (投影矩阵行数) 为 10. 结果如表 1 所示. 其结果与实验 1 基本相同, 当添加以上 3 种噪声后. 2D-IEPCA-GA 下重建误差最小, 去噪效果最好.

4.2 分类实验

此部分实验主要验证 2D-LDA, 2D-ELDA 和 2D-IELDA 的分类性能. 采用 ORL 数据集和 AR 数据集. 分类器采用基于 F 范数的最近邻分类器.

实验 3 本实验主要测试 2D-LDA, 2D-ELDA, 2D-IELDA 算法的分类性能和对噪声的鲁棒性. 实验采用 ORL 数据集, 共 40 个类, 每个类 10 张样本图像, 图像大小为 112×92. 其中每个类 5 张图片 (共 200 张) 作为训练样本, 剩余图像 (共 200 张) 作为测试样本. 分别测试无噪声、高斯噪声数据 (方差为 0.64)、椒盐噪声数据 (噪点比例为 0.5) 和乘性噪声数据 (方差为 3) 4 种情况. 添加噪声方法为对训练样本和测试样本同时添加, 且图像噪声已经比较严重. 以不同的样本划分方法测试 10 轮, 取正确率平均值. 分类结果如图 2 所示, 可以明显看到在有噪声情况下 2D-IEPCA 整体性能明显好于 2D-EPCA 和 2D-PCA, 在除高斯噪声以外的情况下, 2D-EPCA 好于 2D-PCA. 但需要强调的是, 这是在没有调整超参数的情况下获得的结果.

实验 4 本实验主要测试 2D-LDA, 2D-ELDA 和 2D-IELDA 算法对光照和遮挡的鲁棒性. 实验使用 AR 数据集, 共 100 个类 (男人和女人各 50 个类), 每类 26 张图片, 使用每类前 7 张作为训练样本 (有光照变化但无遮挡), 其余作为测试样本 (有墨镜、围巾和口罩遮挡以及光照变化). 实验只测试

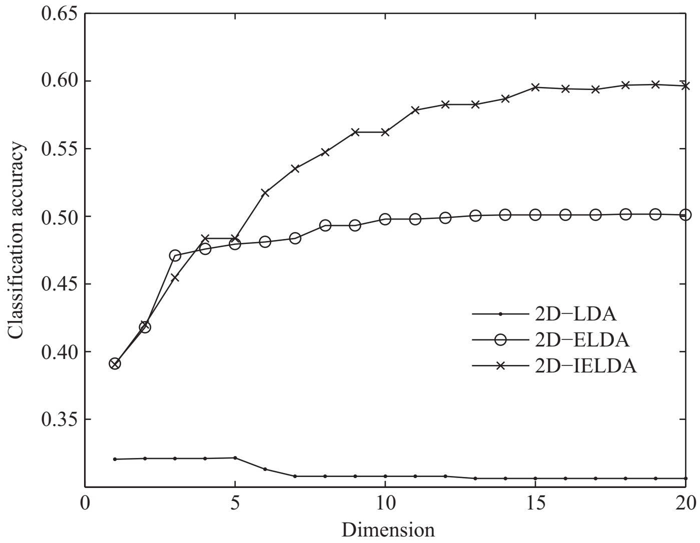


图 3 AR 数据集分类准确率

Figure 3 Classification accuracy on AR dataset

1 轮. 实验结果如图 3 所示, 可以看到 2D-IELDA 最好, 2D-LDA 最差且差距明显.

5 总结

核方法是处理非线性数据的一种重要方法, 但 2D 方法因难以利用表示定理而导致核化困难. 因此本文绕过表示定理, 直接通过改变度量来获得较简洁的针对 2D 方法的核化. 另一方面矩阵数据与向量数据的重要区别之一是其内部包含空间结构信息, 因此在针对矩阵数据核化时应考虑到空间结构信息的利用, 所以本文采用显式、等维和分量间非耦合的核映射. 核映射保留了原数据内的空间结构信息, 因此为后续空间结构信息的利用提供了可能. 更重要的是此种核映射可看作是对数据的预处理, 从而能将核化后的数据直接应用于现有的大多数 2D 算法的改进. 此方法可以看成是一个框架, 在针对 2D 算法改进时只需选择符合要求的核映射和利用空间结构信息的方法. 本文以欧拉映射核和图像欧氏距离为例, 提出 2D 欧拉核和空间结构化欧拉核并用于 2D-LDA 和 2D-PCA 算法的改进. 由实验结果可以看出, 当数据没有噪声污染时, 改进后的方法接近且整体好于原方法. 当数据遭到噪声污染后, 空间结构化欧拉核表现出了极大的优势.

参考文献

1 Yang J, Zhang D, Frangi A F, et al. Two-dimensional PCA: a new approach to appearance-based face representation and recognition. Pattern Analysis and Machine Intelligence, 2004, 26(1): 131–137

2 Li M, Yuan B. 2D-LDA: A statistical linear discriminant analysis for image matrix. Pattern Recognition Letters, 2005, 26(5): 527–532

3 Qian Q, Chen S C. Matrix metric learning algorithm based on likelihood ratio test with matrix normal distribution. Journal of Shandong University of Technology, 2012, 42(6): 37–42

4 Zhang T, Fang B, Tang Y , et al. Generalized discriminant analysis: a matrix exponential approach. IEEE Transactions on Systems Man and Cybernetics part B-CYBERNETICS, 2010, 40(1): 186–197

5 Yan S, Xu D, Zhang L, et al. Coupled kernel-based subspace learning. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, San Diego, 2005. 645–650

6 Zhang D Q, Chen S C, Zhou Z H. Recognizing face or object from a single image: Linear vs. kernel methods on 2d patterns. In: Dit-Yan Yeung, James T K, Ana Fred, et al, eds. Structural, Syntactic, and Statistical Pattern. Recognition 2006. Heidelberg: Springer Berlin Heidelberg, 2006. 889–897
7 Liwicki S, Tzimiropoulos G, Zafeiriou S, et al. Euler Principal Component Analysis. International Journal of Computer Vision February, 2013, 101: 498–518
8 Vedaldi A, Zisserman A. Efficient Additive Kernels via Explicit Feature Maps. in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, 2010. 3539–3546
9 Wu J S, Zheng W S, Lai J H. Euler Clustering. in: Proceedings of International Joint Conference on Artificial Intelligence, Beijing, 2013. 1792–1798
10 Zhang D Q, Zhou Z H: (2D)2PCA: two-directional two-dimensional PCA for efficient face representation and recognition. Neurocomputing, 2005, 69: 224–231
11 Ye J, Janardan R, Li Q. Two-Dimensional Linear Discriminant Analysis. Advances in Neural Information Processing Systems, 2005, 17: 1569–1576
12 Wang H X. Block principal component analysis with L1-norm for image analysis. Pattern Recognition Letters, 2012, 33(5): 537–542
13 Cai D, He X, Hu Y. Learning a Spatia Smooth Subspace for Face Recognition. in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, 2007. 1–7
14 Wang L W, Zhang Y, Feng J F. On the Euclidean distance of images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(8): 1334–1339
15 Huttenlocher D P, Klanderman G A, Rucklidge W J. Comparing Images Using the Hausdorff Distance. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1993, 15(9): 850–863
16 Simard P, Cun Y, Denker J S. Efficient Pattern Recognition Using a New Transformation Distance. In: Proceedings of Advances in Neural Information Processing Systems, Denver, 1992. 50–58
17 Bajcsy R, Lovacic S. Multiresolution Elastic Matching. Computer Vision, Graphics, and Image Processing, 1989, 46:1–21
18 Li J, Chen G, Chi Z. A Fuzzy Image Metric with Application to Fractal Coding. IEEE Transactions on Image Processing, 2002, 11(6): 636–643
19 Gao Q X, 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 Mika S, Sch ¨ olkopf B, Smola A, et al. Kernel pca and de-noising in feature spaces. Advances in Neural Information Processing Systems, 1999, 11(1): 536–542

附录 A 针对向量数据的空间结构化欧拉核

首先简单介绍针对向量数据的结构化欧拉核, 由于对 (19) 式的推导需要此过程的一些结果, 并且此部分也具有一定的实用意义, 因此在附录中简短介绍. 结合欧拉 PCA 和图像欧氏距离, 定义核映射与内积

$$\begin{aligned}&\eta(\boldsymbol{x}_k)=G^{0.5}\varphi(\boldsymbol{x}_k),\\&k(\boldsymbol{x}_l,\boldsymbol{x}_q)=\varphi(\boldsymbol{x}_l)^{\mathrm{H}}G\varphi(\boldsymbol{x}_q),\end{aligned}\tag{A1}$$

其中, $\boldsymbol{x}_k$ 为归一化到 $[0,1]$ 的向量数据, $\varphi(\boldsymbol{x}_k)$ 的定义见 (2) 式, G 的定义见 (4) 式. 此映射方法可作为针对数据的预处理, 直接带入到传统 PCA 算法, 获得投影矩阵 B. 根据文献 [20], 原空间去噪目标为

$$\begin{aligned}\tilde{x}&=\arg\min_{\breve{x}}\left\|G^{0.5}\varphi(\breve{x})-BB^{\mathrm{H}}G^{0.5}\varphi(x)\right\|_F^2\\&=\arg\min_{\breve{x}}\left\{k(\breve{x},\breve{x})-\mathrm{Re}(\varphi(\breve{x})G^{0.5}BB^{\mathrm{H}}G^{0.5}\varphi(x)\right\}.\end{aligned}\tag{A2}$$

令 $G^{0.5}BB^{\mathrm{H}}G^{0.5}\varphi(x)=T$, 则 (A2) 式可化为

$$\begin{aligned}\tilde{x}&=\arg\min_{\breve{x}}\left\{\left[\sum_{i=1}^{p}\sum_{k=1}^{p}g_{ik}\cos\left(\alpha\pi\left(\breve{x}(i)-\breve{x}(k)\right)\right)\right]-\mathrm{Re}\left(\varphi(\breve{x})^{\mathrm{H}}\right)\mathrm{Re}(T)+\mathrm{Im}\left(\varphi(\breve{x})^{\mathrm{H}}\right)\mathrm{Im}(T)\right\}\\&=\arg\min_{\breve{x}}\left\{\left[\sum_{i=1}^{p}\sum_{k=1}^{p}\left(g_{ik}\cos\left(\alpha\pi\left(\breve{x}(i)-\breve{x}(k)\right)\right)\right)\right]-\sum_{i=1}^{p}\left(\cos\left(\alpha\pi\breve{x}(i)\right)\mathrm{Re}\left(T(i)\right)+\sin\left(\alpha\pi\breve{x}(i)\right)\mathrm{Im}\left(T(i)\right)\right)\right\}\end{aligned}$$

$$= \arg\min_{\breve{x}} F(\breve{x}). \tag{A3}$$

上式可以用梯度下降方法求解

$$\frac{\partial F(\breve{x})}{\partial \breve{x}(i)} = \alpha\pi(-\sum_{k=1}^{p} g_{ik}\sin(\alpha\pi(\breve{x}(i)-\breve{x}(k)) + \sin(\alpha\pi\breve{x}(i))\mathrm{Re}(T(i)) - \cos(\alpha\pi\breve{x}(i))\mathrm{Im}(T(i))), \tag{A4}$$

则梯度为

$$\nabla F(\breve{x}) = \alpha\pi(\mathrm{diag}(-G\mathrm{Im}(-\varphi(\breve{x})^{\mathrm{H}} \otimes \varphi(\breve{x})) + \mathrm{Im}(\varphi(\breve{x}))\mathrm{Re}(T)^{\mathrm{T}} - \mathrm{Re}(\varphi(\breve{x}))\mathrm{Im}(T)^{\mathrm{T}}). \tag{A5}$$

附录 B　矩阵数据的空间结构化欧拉核梯度下降法求导过程

再看 (18) 式, 求导可分为前后两部分. 首先对其前半部分求导, 由 (7) 式可知

$$\mathrm{vec}(G_1^{0.5}\Phi(\breve{X})G_2^{0.5}) = (G_2^{0.5} \otimes G_1^{0.5})\varphi(\breve{x}) = G^{0.5}\varphi(\breve{x}), \tag{B1}$$

其中 $\breve{X} \in \mathbb{R}^{m\times n}, \varphi(\breve{x})$=vec$(\Phi(\breve{X})), \breve{x}$=vec$(\breve{X})$. 则

$$\left\|G_1^{0.5}\Phi(\breve{X})G_2^{0.5}\right\|_F^2 = \left\|G^{0.5}\phi(\breve{x})\right\|_F^2 = k(\breve{x},\breve{x}). \tag{B2}$$

因此, $\|G_1^{0.5}\Phi(\breve{X})G_2^{0.5}\|_F^2$ 对 $\breve{X}$ 的求导与 $\|G^{0.5}\phi(\breve{x})\|_F^2$ 对 $\breve{x}$ 的求导等价 (区别是结果的排列顺序不同), 且与 (A2) 式前半部分相同. 直接利用 (A2) 式求导结果. (B2) 式求导结果为 $\alpha\pi(\mathrm{diag}(-G\mathrm{Im}(-\varphi(\breve{x})^{\mathrm{H}} \otimes \varphi(\breve{x}))$. 又因为 $\varphi(\breve{x})^{\mathrm{H}} = [a_1, a_2, ..., a_p]$, 其中 $p = m \times n$, 则

$$\begin{aligned}
&\alpha\pi(\mathrm{diag}(-G\mathrm{Im}(-\varphi(\breve{x})^{\mathrm{H}} \otimes \varphi(\breve{x}))))\\
&= \alpha\pi\mathrm{Im}(\mathrm{diag}(-G(-\varphi(\breve{x})^{\mathrm{H}} \otimes \varphi(\breve{x}))))\\
&= \alpha\pi\mathrm{Im}(\mathrm{diag}(G[a_1\varphi(\breve{x}), a_2\varphi(\breve{x}), ..., a_p\varphi(\breve{x})]))\\
&= \alpha\pi\mathrm{Im}(\mathrm{diag}(\varphi(\breve{x})^{\mathrm{H}} \otimes (G\varphi(\breve{x}))).
\end{aligned} \tag{B3}$$

对以上结果做矩阵直拉的逆运算, 即进行重排, 便得到 $\|G_1^{0.5}\Phi(\breve{X})G_2^{0.5}\|_F^2$ 对 $\breve{X}$ 的求导结果, 为 $\alpha\pi\mathrm{Im}(\overline{\Phi(\breve{X}_k)} \odot (G_1\Phi(\breve{X}_k) G_2))$. 又因为 (18) 式后半部分与 (15) 式形式完全相同, 均为 $\mathrm{Re}(\mathrm{tr}(\Phi(\breve{X}_k)^{\mathrm{H}} T))$ 的形式, 其中 T 为已知的常数矩阵. 因此 (18) 式后半部分求导可直接利用 (15) 式的结果. 前后两部分求导相加便得到了 (19) 式.

Spatial Euler kernel and its applications

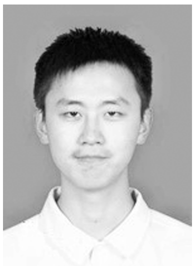
Shuang LIU [1] & Songcan CHEN[2*]

College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016*, China*

*E-mail: s.chen@nuaa.edu.cn

Abstract　Since the Principal Component Analysis (1D-PCA) extended to the image oriented 2D-PCA by Yang et al, a number of 1D methods have been extended to their corresponding 2D variants. Because of the use of natural spatial information of matrix-form data (e.g., an image), the 2D methods usually yield much better performance than their 1D versions, which is consistent with the theorem of "No Free Lunch". However, the 2D methods still suffer from two main drawbacks: (1) they are almost all linear, which might not match the nonlinear structure of actual data; and (2) the spatial information of data is not fully used in existing 2D methods. To address the first drawback, although the kernel trick is theoretically feasible, it is practically difficult since the representation theorem cannot be straightforwardly extended for 2D-form data. To this end, we in this work propose to obtain a simple kernelizing method by changing the measurement without the representation theorem. In view of the second shortcoming, we aim to use the spatial information of data in the nonlinear mapping feature space (or say kernel space). Unfortunately, it usually requires to describe the data in the nonlinear feature space (i.e., kernel space), which is generally implemented through data dimension-increasing as well as implicit kernel mapping. To fulfill this goal, it usually refers to the implicit or explicit kernel mapping, the former, however, might distort the spatial structure of data while the latter leads to dimension risk. As a result, we can preferably preserve the spatial structure of data naturally if we employ the form of explicit kernels in which the dimension is identical and each component is uncoupled. Fortunately, many explicit kernel (e.g., Hellinger and Euler kernels) and approximate explicit mathematical formulations of some implicit additive kernels (e.g., Intersection, JS and χ^2 kernels) meet the requirements. Considering the conciseness and good generalization of the Euler kernel, we in this work attempt to kernelize matrix-form data by Euler kernel and then in the mapped kernel space to compensate the spatial information. Although there exist various ways to compensating the spatial information, e.g., spatial structure information constraints and image distance metric, we in this work take the Image Euclidean Distance as an example to carry out the study and then develop matrix/image-oriented Spatial Euler Kernel. Finally, through experiments, we demonstrate the effectiveness of the proposed strategies.

Keywords 2D method, kernel trick, spatial information, Euler kernel, image Euclidean distance



Shuang LIU was born in 1989. He received the B.S. degree in computer science from Nanjing University of Aeronautics and Astronautics, China in 2013. Since 2013, he has been a M.S. candidate in computer science at Nanjing University of Aeronautics and Astronautics, and his current research interests include machine learning and pattern recognition.

Songcan CHEN was born in 1962. He received the B.S. degree from Hangzhou University (now merged into Zhejiang University), the M.S. degree from Shanghai JiaoTong University and the Ph.D. degree from Nanjing University of Aeronautics and Astronautics (NUAA) in 1983, 1985, and 1997, respectively. He joined in NUAA in 1986, and since 1998, he has been a full-time Professor with the Department of Computer Science and Engineering. He obtained Honorable Mentions of 2006, 2007 and 2010 Best Paper Awards of Pattern Recognition Journal respectively. His current research interests include pattern recognition, machine learning, and neural computing.