软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2018,29(12):3786−3798 [doi: 10.13328/j.cnki.jos.005392]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

基于切空间判别学习的流形降维算法*

王 锐, 吴小俊

(江南大学 物联网工程学院,江苏 无锡 214122)

通讯作者: 吴小俊, E-mail: xiaojun_wu_jnu@163.com

摘 要: 在基于图像集的流形降维问题中,许多算法的核心思想都是把一个高维的流形直接降到一个维数相对较低、同时具有的判别信息更加充分的流形上.投影度量学习(projection metric learning,简称 PML)是一种 Grassmann 流形降维算法.该算法是基于投影度量,并且使用 RCG(Riemannian conjugate gradient)算法优化目标函数,其在多个数据集上都取得了较好的实验结果,但是对于复杂的人脸数据集,如 YTC 其实验结果相对较差,只取得了 66.69%的正确率.同时,RCG 算法的时间效率较差.基于上述原因,提出了基于切空间判别学习的流形降维算法.该算法首先对于 PML 中的投影矩阵添加扰动,使其成为对称正定(symmetric positive definite,简称 SPD)矩阵;然后,使用 LEM(log-euclidean metric)将其映射到切空间中;最后,利用基于特征值分解的迭代优化算法构造判别函数,得到变换矩阵.对提算法在多个标准数据集上进行了实验验证,并取得了较好的实验结果,从而验证了该算法的有效性.

关键词: Grassmann 流形;降维;RCG;对称正定矩阵;LEM;特征值分解

中图法分类号: TP181

中文引用格式: 王锐,吴小俊.基于切空间判别学习的流形降维算法.软件学报,2018,29(12):3786−3798. http://www.jos.org.cn/1000-9825/5392.htm

英文引用格式: Wang R, Wu XJ. Manifold dimensional reduction algorithm based on tangent space discriminant learning. Ruan Jian Xue Bao/Journal of Software, 2018,29(12):3786−3798 (in Chinese). http://www.jos.org.cn/1000-9825/5392.htm

Manifold Dimensional Reduction Algorithm Based on Tangent Space Discriminant Learning

WANG Rui, WU Xiao-Jun

(School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China)

Abstract: Some good dimensional reduction algorithms based on image set have been developed. The core of these algorithms is performing a geometry-aware dimensionality reduction from the original manifold to a lower-dimensional, more discriminative manifold. Projection Metric Learning is a dimensional reduction algorithm that is based on Grassmann manifold. This algorithm, which is based on projection metric and RCG algorithm, has achieved better results on some benchmark datasets, but for some complicated face datasets, such as YTC, it has just obtained 66.69% classification accuracy. However, RCG algorithm has a poor performance of time efficiency. Based on the above reasons, a dimensional reduction algorithm based on the tangent space discriminant learning is presented. Firstly, perturbation is added to the projection matrix of PML to make it be a SPD matrix. Secondly LEM is adopted to map the element which lies on the SPD manifold to a tangent space, and then the iterative optimization algorithm based on eigen-decomposition is applied to find the discriminant function to obtain the transformation matrix. The experimental results on several standard datasets show the superiority of the proposed algorithm over other state-of-the-art algorithms.

Key words: Grassmann manifold; dimensional reduction; RCG; symmetric positive definite matrix; LEM; eigen-decomposition

* 基金项目: 国家自然科学基金(61373055, 61672265); 江苏省教育厅科技成果产业化推进项目(JH10-28); 江苏省产学研创新项目(BY2012059)

Foundation item: National Natural Science Foundation of China (61373055, 61672265); Industry Project of Provincial Department of Education of Jiangsu Province (JH10-28); Industry Oriented Project of Jiangsu Provincial Department of Technology (BY2012059)

收稿时间: 2017-03-01; 修改时间: 2017-05-18; 采用时间: 2017-09-13

在计算机科学的许多领域,降维技术始终都扮演着非常重要的角色,例如在计算机视觉、机器学习以及图像处理等领域.传统的降维方法有 PCA[1]和 LDA[2],这两种方法在人脸识别[3]、表情识别[4]、物体对象识别[5]、形状识别[6]等领域中有着广泛而又成熟的应用,但是这两种方法所处理的数据类型都是基于欧氏空间的.随着科技的进步与发展,科研人员发现在现代计算机科学领域中,我们处理的很多数据在所处空间并非都是传统的欧氏空间,而是处在流形上[7].流形又可以按照数据的性质和结构的不同分为 SPD(symmetric positive definite)流形[8]、Grassmann 流形[9]、Steifel 流形[10]等,传统的基于欧氏空间的分类以及识别问题在流形上又得到了新的研究与发展.但是在大数据时代,我们所采集到的图像较之以前变得更加复杂[11−13],这对传统的分类与识别算法是一项新的挑战.为了更好地抽取复杂图像的特征信息,基于图像集的分类与识别问题逐渐受到了研究人员的关注[11−14].

基于图像集的分类和识别问题相对于传统的基于单幅图像的问题在准确性以及有效性上有了较大的提高,同时,在基于图像集的分类任务中每一个图像集都是由很多张属于同一类的图片构成的.但是由于受到拍照时姿势和视角的改变、物体的非刚性形变、不同的光照条件、不同的场景、不同的分辨率等的影响,不同的图像集中所包含的图片在表观上又存在着一定的差别.WANG 等人提出了协方差判别学习(covariance discriminant learning,简称 CDL)[5],该方法首先对原始图像集利用协方差矩阵进行建模,使样本处在 SPD 流形之上,再利用 LEM 以及其核化形式将其映射到一个平滑空间(切空间)上;然后,利用传统的核判别分析(kernel discriminant learning,简称 KDA)[15]去学习.该方法在多个数据集上取得了较好的实验结果.Faraki 等人提出了一种解决基于对称半正定(symmetric positive semi-definite,简称 SPSD)矩阵的图像集的分类问题[16],当我们利用协方差矩阵去建模时,所得到的矩阵往往是对称半正定的,传统的解决方法是通过添加扰动量使其成为对称正定的,然后再利用基于 SPD 矩阵的一些算法去求解.但是在该方法中,作者把 SPSD 矩阵分解成一个 SPD 矩阵以及一个正交线性子空间,这样,再利用传统的核方法就可以很好地去实现分类任务.但是在解决图像集分类的传统方法中[5,11−14],一方面存在计算复杂性较高的问题,例如常常使用到的核函数;另一方面,原始流形的维数较高,冗余较大,在最终的精度和效率方面仍然有待提高.在传统的欧氏空间中遇到上述问题时,我们可以利用降维的方式对算法的计算量进行约减;但当数据所处的空间是非线性的黎曼流形时,欧氏空间中的降维算法就无法直接应用在其上.Harandi 等人提出了一种流形降维算法[8],该算法的目的是对一个原始的高维 SPD 流形直接进行降维,使其转变成一个维数相对较低、同时判别性更加充分的新的 SPD 流形.该方法首先定义了映射函数;然后定义了二值系数矩阵,作为目标函数中的判别系数;接着,基于仿射不变黎曼度量(affine invariant Riemannian metric,简称 AIRM)[17]去定义判别函数;最后,通过 RCG 算法去优化得到变换矩阵.该方法的计算量较小的同时使得降维后的流形具有更加充分的判别性.而 Huang 等人同样也提出了一种类似的降维算法——投影度量学习(projection metric learning,简称 PML)[9],该算法首先利用投影矩阵表示 Grassmann 流形上的元素;然后定义投影映射函数,接着,利用投影度量定义判别函数,最后,同样利用 RCG 去优化.但是该方法是通过优化 SPSD 矩阵 P 去间接得到最终所需的变换矩阵 W.该方法与 Harandi 等人提出方法的目的都是一样的,都是为了在原始流形上直接去降维,同时使得降维后流形的判别性更加充分,而且在优化目标函数时也都利用了 RCG 算法.

线性子空间是 Grassmann 流形的核心组成部分,同时,基于线性子空间的图像集分类以及识别问题也在受到人们的广泛关注[18−20],主要原因是由于利用线性子空间可以较为有效地降低计算的复杂性;同时,对于复杂图像集的建模问题,线性子空间可以很好地表征图像的特征信息.Grassmann 流形是非线性的结构,传统的基于欧氏空间的一些降维或者判别学习的方法都不能直接地应用在 Grassmann 流形上,针对该问题,很多学者也都提出了相应的算法[10,21−24],虽然这些算法都取得了很好的实验结果,但是它们都忽略了数据的流形性,即由线性子空间张成的 Grassmann 流形.因此,许多学者提出了基于流形降维的算法[8,9,25],这些算法的优点就是直接在流形上面进行降维,这样做一方面有效地利用了数据的流形结构,同时也使得降维后的流形具有更加充分的判别信息.PML 是一个经典的 Grassmann 流形降维算法,该算法利用的是投影度量,虽然取得了较好的实验结果,但是对于复杂的人脸数据集,如 YTC,其实验结果相对较差,从而可以说明:对于具有遮挡、光照以及形变等情形的图像的分类任务上,其判别性不够充分;同时,PML 使用的是 RCG 优化算法,该方法虽然可以较为真实地反映流形上

的梯度变化情况,求得的结果也很精确,但是该方法的效率很低,很难应用在实际生活中.鉴于此,我们提出了基于切空间判别学习的流形降维算法.该算法首先对 PML 中的投影矩阵添加扰动,使其成为 SPD 矩阵,也即使数据的分布空间从 Grassmann 流形转换到 SPD 流形;然后,利用 LEM 将其投影到对应的切空间中,这样,我们所计算出的测地线距离就会更加准确有效;最后,利用基于特征值分解的迭代优化算法去求解判别函数.我们在多个数据集上对该方法进行了验证,并取得了较好的实验结果,从而验证了该方法可以比较有效地克服原始方法的不足.

1 相关工作

1.1 投影度量学习(PML)

给定 m 个人脸视频序列,即 m 个人脸图像集$\{X_1,X_2,\dots,X_m\}$,其中,$X_i\in R^{D\times n_i}$ 表示第 i 个视频序列且含有的图像帧数为 n_i,而每个图像帧又可以看成是 D 维的特征向量.由于 $X_iX_i^T\approx Y_i\Lambda Y_i^T$,其中,$\Lambda$以及 Y_i 表示该矩阵的前 q 个最大的特征值以及对应的特征向量,因此,图像集 X_i 就可以由正交基矩阵 Y_i 张成的 q 维的线性子空间表示,其中,$Y_i\in R^{D\times q}$.

对于在 Grassmann 流形上由 Y_i 张成的线性子空间(其中,用 $Y_iY_i^T$ 表示该流形上的元素),一般性的投影映射 $f{:}G(q,D)\to G(q,d)$可以定义为

$$f(Y_iY_i^T)=W^TY_iY_i^TW=(W^TY_i)(W^TY_i)^T \tag{1}$$

其中,$W\in R^{D\times d}$ 是所需的变换矩阵,且其为列满秩的.通过该映射,原始的 D 维流形就可以被降维到新的 d 维的流形上.但是原始的 Y_i 是正交基矩阵,因此降维后为了保证流形结构不被改变,那么 W^TY_i 也必须是正交基矩阵.但是通过分析可以看出,W^TY_i 并非是正交的,所以我们可以利用正交分解对其进行处理:$W^TY_i=Q_iR_i$,其中,$Q_i\in R^{D\times q}$ 是 q 维的正交矩阵,而 $R_i\in R^{q\times q}$ 是可逆的上三角矩阵,因此我们可以得到:

$$Q_i=W^T(Y_iR_i^{-1})\to Y_i'=Y_iR_i^{-1} \tag{2}$$

这样,我们就可以利用 Y_i' 去代替 Y_i,从而保证了 W^TY_i' 的正交性.

对于 Grassmann 流形上的任意两个投影算子 $W^TY_i'Y_i'^TW$ 以及 $W^TY_j'Y_j'^TW$,其投影度量可以定义为

$$d_P^2(W^TY_i'Y_i'^TW,W^TY_j'Y_j'^TW)=2^{-1/2}\|W^TY_i'Y_i'^TW-W^TY_j'Y_j'^TW\|_F^2=2^{-1/2}tr(PA_{ij}A_{ij}^TP) \tag{3}$$

其中,$A_{ij}=Y_i'Y_i'^T-Y_j'Y_j'^T$,$P=WW^T$.由于 $W\in R^{D\times q}$ 是一个列满秩的,因此,P 则是一个 $D\times D$ 维的 SPSD 矩阵.

基于上述投影度量,就可以给出判别函数的定义:

$$P^*=\arg\min_P J(P)=\arg\min_P tr(PS_wP)-\alpha tr(PS_bP) \tag{4}$$

其中,S_w 以及 S_b 的定义是

$$S_w=\frac{1}{N_w}\sum_{i=1}^m\sum_{j:C_i=C_j}2^{-1/2}(A_{ij}A_{ij}^T) \tag{5}$$

$$S_b=\frac{1}{N_b}\sum_{i=1}^m\sum_{j:C_i\neq C_j}2^{-1/2}(A_{ij}A_{ij}^T) \tag{6}$$

在 S_w 以及 S_b 的定义中,N_w 表示同一类中用于计算距离散度的成对的样本点的个数,N_b 则表示不同类中用于计算距离散度的成对的样本点的个数.

1.2 优化

由前面的分析可知,该类问题最终转化为一个优化问题.在 PML 中,需要求解的参数是 P,其是一个 SPSD 矩阵,且公式(4)的偏导数为 $D_P(P_k)=2(S_w-\alpha S_b)P_k$.最后,利用 RCG 优化算法去求解.由于 RCG 算法是共轭梯度法在黎曼流形上的推广[26],因此在利用该算法去求解 P 时的步骤如下.

Riemannian Conjugate Gradient(RCG)算法.

输入:初始化的矩阵 P_0;

1. $H_0 \leftarrow 0$, $P \leftarrow P_0$.
2. 重复以下操作
3. $H_k \leftarrow \nabla_P J(P_k)+\eta\tau(H_{k-1},P_{k-1},P_k)$.
4. 沿着测地线γ,以 H_k 为搜索方向,通过点 $P_{k-1}=\gamma(k-1)$去寻找点 $P_k=\text{argmin}_P J(P)$.
5. $H_{k-1} \leftarrow H_k$, $P_{k-1} \leftarrow P_k$
6. 直到满足收敛条件

输出:优化后得到的 SPSD 矩阵 P.

2 基于切空间判别学习的流形降维

在 PML 中,用投影矩阵 $Y_iY_i^T$ 表示 Grassmann 流形上的元素,然后,在投影度量的基础上去构造判别函数.对于目标矩阵 W 的求解,是通过利用 RCG 算法对 SPSD 矩阵 P 进行优化而间接得到.但是对于复杂的人脸数据集,其判别性不够充分;同时,RCG 算法的时间效率也相对较差.对数欧氏度量(log-Euclidean metric,简称 LEM)[27]是 SPD 流形上常用的一种距离度量,其可以将 SPD 流形上的元素在一点处(该点是一个单位矩阵)映射到一个切空间,该映射过程可以表示为[5]

$$\Psi_{\log}:M \mapsto T_I, C \to \log(C) \tag{7}$$

其中,T_I 是单位矩阵 I 处的切空间,实际上是由 $D\times D$ 维的对称矩阵张成的一个向量空间.通过该映射, SPD 流形上任意两点间的距离就可以近似看成是切空间中任意两点的欧氏距离.对于 SPD 流形中的任意两点 X 和 Y,该度量的表达形式为[27]

$$d_{LE}=\|\log(X)-\log(Y)\|_F \tag{8}$$

其中,log 就是数学中的对数运算符,$\|\cdot\|_F$ 则表示矩阵的 F 范数.LEM 的优点之一是计算复杂度低,同时,该度量具有很完美的理论性质[28].鉴于上述情况,我们考虑对 PML 中的投影矩阵 $Y_iY_i^T$ 采用以前文献中常用的处理非满秩矩阵的方法[5,29],即添加扰动来使其成为 SPD 矩阵:

$$Y_iY_i^T = Y_iY_i^T + \frac{trace(Y_iY_i^T)}{\gamma} I \tag{9}$$

这样,我们就可以利用 LEM 去度量 SPD 流形上任意两点间的距离.通过该方式,我们就可以把变换矩阵 W 放到 SPD 流形中求解.

由于我们的目的是对原始的 Grassmann 流形直接进行降维,使其变成一个维数相对较低同时判别信息更加充分的 Grassman 流形,因此对于映射 $f:G(q,D)\to G(q,d)$,同样可以定义为

$$f(Y_iY_i^T) = W^TY_iY_i^TW = (W^TY_i)(W^TY_i)^T \tag{10}$$

其中,$W\in R^{D\times d}$ 即为我们所要求解的变换矩阵.由于 Y_i 是正交基矩阵,且为了保证降维后的流形仍然是 Grassmann 流形,因此 W^TY_i 也必须具有正交性,所以我们采取与文献[9]中一样的方式对其进行处理,即利用正交分解方法.为了便于后续的叙述,我们用 Y_i^o 代替原始的 Y_i.

2.1 切空间判别学习

对于任意两个 SPD 流形中的投影算子 $W^TY_i^oY_i^{oT}W$ 和 $W^TY_j^oY_j^{oT}W$,其 LEM 可表示为

$$d_W^2 = \|\log(W^TY_i^oY_i^{oT}W) - \log(W^TY_j^oY_j^{oT}W)\|_F^2 \tag{11}$$

由于上式对于 W 无解析解,因此为了方便和简化后续的计算,我们采用与文献[8,30]中一样的处理方式,即:

$$\log(W^TXW)=W^T\log(X)W \tag{12}$$

所以,公式(11)可以改写为

$$d_W^2 = \| W^T \log(Y_i^o Y_i^{oT}) W - W^T \log(Y_j^o Y_j^{oT}) W \|_F^2 \tag{13}$$

我们令 $A=\log(Y_i^o Y_i^{oT}), B=\log(Y_j^o Y_j^{oT})$,那么上式可以化简为

$$\left.\begin{aligned} d_W^2 &= \| W^T AW - W^T BW \|_F^2 \\ &= tr[(W^T AW - W^T BW)(W^T AW - W^T BW)^T] \\ &= tr[(W^T AW - W^T BW)(W^T A^T W - W^T B^T W)] \\ &= tr[W^T (A-B) W W^T (A-B)^T W] \\ &= tr(W^T f(W) W) \end{aligned}\right\} \tag{14}$$

其中,$f(w)=(A-B)WW^T(A-B)^T$.

判别函数的定义. 我们定义判别函数的目的是使得原始的高维流形在经过降维操作以后,在得到的新的流形上,同类样本之间的距离最小,而不同类样本之间的距离最大,因此,我们可以定义如下的目标函数:

$$W^* = \arg\min_W T(W) = \arg\min_W (T_w(W) - \alpha T_b(W)) \tag{15}$$

其中,α是平衡类内以及类间紧凑度的参数;T_w 和 T_b 分别表示类内和类间的平均距离散度,其定义为

$$T_w(W) = \frac{1}{N_w} \sum_{i=1}^{m} \sum_{j:C_i=C_j} tr(W^T f(W) W) \tag{16}$$

$$T_b(W) = \frac{1}{N_b} \sum_{i=1}^{m} \sum_{j:C_i \neq C_j} tr(W^T f(W) W) \tag{17}$$

其中,N_w 表示相同类中用于计算平均距离散度的成对的样本数,N_b 则表示不同类中用于计算平均距离散度的成对的样本数.

2.2 优 化

把公式(16)以及公式(17)代入公式(15)中,可得:

$$\left.\begin{aligned} W^* &= \arg\min_W (T_w(W) - \alpha T_b(W)) \\ &= \arg\min_W tr(W^T S_w W) - \alpha tr(W^T S_b W) \\ &= \arg\min_W tr[W^T (S_w - \alpha S_b) W] \\ &= \arg\min_W tr(W^T J(W) W) \end{aligned}\right\} \tag{18}$$

其中,S_w 以及 S_b 的定义为

$$S_w = \frac{1}{N_w} \sum_{i=1}^{m} \sum_{j:C_i=C_j} f(W) \tag{19}$$

$$S_b = \frac{1}{N_b} \sum_{i=1}^{m} \sum_{j:C_i \neq C_j} f(W) \tag{20}$$

且 $J(W)=S_w-\alpha S_b$.

由公式(18)可知,我们可以采用基于特征值分解的迭代优化算法[31]求解最终所需的变换矩阵 W.首先,给 W 赋一个初始值,该值既可以是随机数矩阵也可以是单位矩阵;然后计算公式(18);接着,取出 $J(W)$的前 d 个最小的特征值对应的特征向量,即一个新的 W;最后,用新的 W 更新 $J(W)$,直到公式(18)收敛为止.该过程可用算法 1 直观表示.

算法 1. 基于特征值分解的迭代优化算法.

输入:初始化的变换矩阵 $W \leftarrow I_{D\times d}$;

1. 重复以下操作:
2. 利用公式(19)以及公式(20)计算 $J(W)$;

3. 对变换矩阵 W 赋予新的值也即取出 $J(W)$的前 d 个最小的特征值对应的特征向量,然后用新的 W 去更新 $J(W)$;
4. 直到满足收敛条件

输出:优化后得到的变换矩阵 W.

为了解算法的收敛性,我们在 ETH-80 基准数据集上进行实验,观察随着迭代次数的增加,目标函数的误差变化情况,实验结果如图 1 所示.通过图 1 可以看出:基于特征值分解的迭代算法可以在较少的迭代次数内使得目标函数的误差有一个较大幅度的降低;同时,随着迭代次数的增加,目标函数的误差也逐渐趋于稳定.这说明所提算法具有较好的收敛性,同时也说明利用目标函数的误差作为判定迭代停止的条件是合理、有效的.

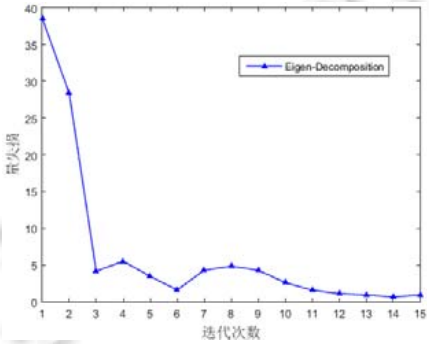


Fig.1 Convergence behavior of the proposed algorithm on ETH-80 dataset

图 1 在 ETH-80 数据集上所提算法的收敛性表现

经过前面的分析和计算,本文提出的算法的主要步骤可以归结为以下的算法 2.

算法 2. 基于切空间判别学习的流形降维.

输入:所有的 q 维的线性子空间也即 $span(Y_i)$;

1. $W \leftarrow I_{D\times d}$.
2. 重复以下步骤:
3. 用 Y_i^o 代替 Y_i,从而保证 $W^T Y_i^o$ 的正交性;
4. 利用公式(19)以及公式(20)计算 S_w 和 S_b;
5. 利用算法 1 去优化公式(18)中的变换矩阵 W;
6. 满足收敛条件时停止.

输出:优化后得到的变换矩阵 W

3 实验结果及分析

我们在两个不同的分类任务上验证本文提出算法的有效性,分别是人脸识别以及对象分类.对于人脸识别任务,我们选取的是被学者广泛使用的 Honda/UCSD[32]基准数据集,以及富有挑战性的 YouTube Celebrities (YTC)[5]基准数据集.对于对象分类任务,我们选取 ETH-80[33]基准数据集.为了更好地对比实验结果,验证本文提出的算法的合理性以及有效性,我们把本文提出的方法与 MSM(mutual subspace method)[18],DCC (discriminative canonical correlations)[19],AHISD(affine hull based image set distance)[20],CHISD(convex hull based image set distance)[20],MMD(manifold-manifold distance)[34],CDL(covariance discriminant learning)[5],GDA (grassmann discriminant analysis)[35],PML(projecion metric learning)[9]进行了比较,这些方法都是基于图像集的分类方法.其中,MSM,AHISD 以及 CHISD 都是无监督的基于子空间的分类算法,而剩下的方法都是有监督的.在 AHISD 以及 CHISD 中,每一个图像集都可以看成一个凸的几何区域,这样就可以通过计算任意两个凸的几何区域之间的距离去判别相似度.对于仿射包(affine hull),可以通过利用最小二乘法去计算最小的距离;而对于

凸包(convex hull),则可以利用 SVM 去分类.MMD 算法提出了一种计算流形之间距离的方法,该方法把流形划分成一些局部的线性模型,每个线性模型可以看成是一个子空间,这样,计算流形之间的距离就变成了计算子空间之间距离,而 DCC 算法则试图通过利用训练样本以及由典型相关分析得到的相似度函数去学习一个判别函数.CDL 算法是解决基于 SPD 流形的图像集分类问题的经典算法,该算法首先对原始图像集利用协方差矩阵进行建模,然后利用 LEM 及其核化形式将流形上的点映射到一个欧氏空间,最后利用基于欧氏空间核判别分析(kernel discriminant analysis,简称 KDA)[15]算法进行学习.对于分类任务,我们都希望通过相应的映射或者降维处理使得数据在新的空间中具有更加充分的判别性[8,9,26],GDA 算法则首先把一个 Grassmann 流形嵌入到一个高维的希尔伯特空间,然后通过 Fisher 线性判别准则去学习一个映射,从而使得原始的流形被映射到一个维度相对较低同时判别性又更加充分的空间.

在对本文提出的算法进行实验验证的过程中,线性子空间的维数 q、目标流形的维度 d、平衡系数α、扰动量γ以及迭代次数这几个参数对实验结果都会产生影响.对于这些参数值的确定,我们都是根据交叉验证的方式取得.同时,为了实验对比的公平性和严谨性,对于参与对比的算法,我们在本实验中使用的参数值与其完全一致.对于在 3 个数据集上的实验结果,除了算法 CDL 和 PML 是通过自己实现的以外,其余实验结果都是使用相关学者已经实现了的.

对于本文所有的实验,在用到的 3 个数据集上,都采用 10 次交叉验证的方式,然后把这 10 次实验的平均值作为最终的实验结果.

3.1 各数据集上的实验

3.1.1 Honda/UCSD 数据集

Honda 数据集是由含有 20 个不同的人的 59 个视频序列构成的.在该数据集中,每个类含有的图像集的个数是不相同的,同时,每个图像集中含有的图片数目也不尽相同,大约都在 300 张~500 张之间.在每个类中,不同的图像集所含有的图片在头部姿势、面部表情以及表观信息上等有着很大的差别,同时,图片的尺寸也都不完全相同.实验中,我们把图片的尺寸调整到 20×20,同时,在每个类中我们随机选取一个图像集用做训练,剩下的图像集用做测试.图 2 为部分 Honda 人脸图像,表 1 为实验结果.

Fig.2 Part of the face images of Honda dataset

图 2 部分 Honda 人脸图像

Table 1 Average result of each algorithm on Honda dataset

表 1 各算法在 Honda 上的平均实验结果

算法	Honda/UCSD
MSM[18]	92.50±2.74
DCC[19]	94.87±1.32
MMD[34]	94.87±1.16
AHISD[20]	89.74±1.85
CHISD[20]	92.31±2.12
CDL	98.13±2.64
PML	98.44±2.21
Proposed	**99.06±1.51**

3.1.2 ETH-80 数据集

ETH-80 数据集中含有 8 个类别,分别是苹果、奶牛、杯子、马、土豆、汽车、梨以及狗,每个类中含有 10

个图像集,每个图像集中又含有41个不同视角下的图片,而且图片的尺寸都是256×256.在实验中,我们把图片的尺寸同样调整为 20×20,而且在每个图像集中,我们随机选取 5 个图像集用做训练,剩下的用做测试.图 3 为部分 ETH-80 图像,表 2 是实验结果.

Fig.3 Part of the face images of ETH-80 dataset
图 3 部分 ETH-80 图像

Table 2 Average result of each algorithm on ETH-80 dataset
表 2 各算法在 ETH-80 上的平均实验结果

算法	ETH-80
MSM[18]	87.82±3.95
DCC[19]	90.75±4.42
MMD[34]	85.72±8.29
AHISD[20]	77.25±7.50
CHISD[20]	74.25±5.01
GDA[35]	92.25±4.16
PML	89.50±3.07
Proposed	**92.50±4.25**

3.1.3 YouTube 数据集

YouTube Celebrities(YTC)数据集是由包含 47 个类的 1 910 个视频片段组成的,每个视频片段又含有数以百计的图像帧.这些图片大部分都受到了噪声的影响而且分辨率较低;同时,在头部姿态、面部表情、以及表观信息上也有着很大的差别.在 YouTube 数据集中,每个类所含有的图像集的个数都是不同的;同时,图片的尺寸也不尽相同.实验中,我们同样把图片的尺寸调整为 20×20;同时,在每个类中,我们随机地选取 9 个图像集,其中 3 个用于训练,剩下的用于测试.图 4 是部分 YouTube 图像,表 3 是实验结果.

Fig.4 Part of the face images of YouTube dataset
图 4 部分 YouTube 人脸图像

Table 3 Average result of each algorithm on YouTube dataset
表 3 各算法在 YouTube 上的平均实验结果

算法	YouTube
MSM[18]	60.25±3.05
DCC[19]	65.48±3.51
MMD[34]	62.90±3.24
AHISD[20]	63.70±2.89
CHISD[20]	66.62±2.79
GDA[35]	65.02±2.91
CDL	68.42±3.67
PML	66.83±7.01
Proposed	**71.67±4.91**

3.2 实验结果及分析

从上述实验可以看出:不论在 Honda,ETH-80 还是 YouTube 数据集上,改进后的算法在分类正确率上都有较大提高.在 Honda 数据集上,文中提出的算法在分类正确率上达到了 99.06%,相对于 PML,CDL 等算法有了较大的提高;同时,其均方差为 1.51,比大部分对比算法都要小,说明文中提出的算法在人脸识别的任务上具有较好的有效性和鲁棒性.同时可以看出:文中提出的算法以及 PML 算法在 Honda 上的实验结果较 CDL 要好,主要原因是由于 CDL 通过在其切空间中利用 KDA 去学习,但是在该过程中,由于 KDA 的降维效应使得图像的特征信息出现了相应的丢失,同时,CDL 的降维过程是在欧氏空间中完成的,其相较于对流形直接进行降维,忽略了数据的流形性,因此在这个近似过程中,特征的判别性就相对不足.在 ETH-80 数据集上,改进后的算法在分类正确率上达到了 92.50%,较其他算法也有了较大的提高.同时,PML 算法在 ETH-80 数据集上也取得了 89.50%的分类正确率,较 MSM,MMD,AHISD 以及 CHISD 有了一定的提高.虽然其相对 DCC 以及 GDA 还有一定的差距,但是其均方差为 3.07 较前两者都较小,说明 PML 算法更加稳定.而本文提出方法的均方差为 4.25,虽然不是最低的,但其较大多数对比算法都是好的,说明其在处理对象分类的任务上仍然具有较好的鲁棒性和有效性.而在 YouTube 数据集上,本文提出的方法取得了 71.67%的分类结果,较几种对比算法也有了较大的提高;同时可以看出:其均方差为 4.91,虽然在所有的方法中仅比 PML 算法稍有优势,但与其他算法的差距不大,从而说明本文提出的方法在具有挑战性的人脸数据集上进行分类任务时,有效性以及鲁棒性仍然可以得到保证.对于 PML 算法,其在 YouTube 数据集上取得了 66.83%的分类正确率,较前几种算法也有一定提高,但是其均方差为 7.01,说明对于这种复杂人脸图像的分类任务,其鲁棒性不够.主要原因在于:当使用 *PM* 作为距离度量时,对于这种复杂的图像特征,其判别性还不够充分,从而对于处理一些高度压缩或者分辨率较低的图像时,其很难区分开.

通过在 3 个数据集上的实验可以看出:本文提出的算法在分类正确率上均优于其他算法;同时,本文提出的算法在鲁棒性上也要优于 PML 等算法.这说明对于 Grassmann 流形上的降维问题,在 SPD 流形的切空间中利用判别分析以及特征值迭代优化学习到的变换矩阵 W 较 PML 算法具有更加充分的判别性,从而不论是对于对象分类任务还是对于具有复杂特征的人脸识别任务,其都能表现出较好的有效性和鲁棒性.

3.3 参数的影响

从本文的算法过程可以看出,影响本文所提算法实验结果的参数有线性子空间的维数 q、降维的目标维数 d、平衡系数α、扰动量γ以及迭代次数 m.对于这些参数值的确定,我们都是通过交叉验证的方式取得,但是不同的参数取值会带来不同的实验结果,因此我们首先固定参数γ,m 以及 q,在 YouTube 数据集上去观察平衡系数α和目标维数 d 的变化对实验结果的影响,如图 5 所示.

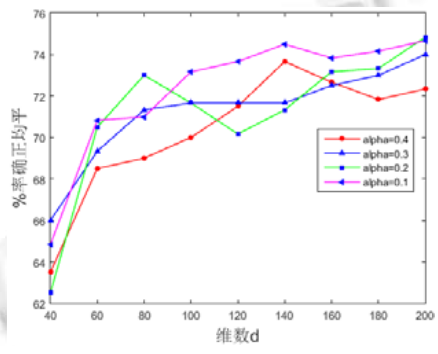


Fig.5 Effect of the parameter α and d on the accuracy of the proposed method on YouTube dataset

图 5 在 YouTube 数据集上参数α和 d 对实验结果的影响

从图 5 可以看出,不论是哪一个α值,随着目标维数的增加,实验结果总体是呈现上升的趋势.主要原因是:随着目标维数的增加,图像在降维过程中丢失的信息量就会减小,因此分类的正确率就会得到提高.同时可以看

出,当目标维数超过 100 后,分类正确率的整体变化是比较平稳的,可以说明本文提出的算法具有较好的鲁棒性.从图中可以看出:当我们固定α时,较好的维数值可以在 140~200 之间的取得;而当我们固定维数值后,α值取 0.1 或者 0.2 是相对较好的.

在本文提出的算法中,我们通过添加扰动量的方式去改变 $Y_iY_i^T$ 所处的流形,但是扰动量的添加会改变原始图像的特征信息,因此我们固定其他参数的值,在 ETH-80 和 YouTube 数据集上,通过实验观察不同的扰动量γ值对实验结果的影响,如图 6 所示.

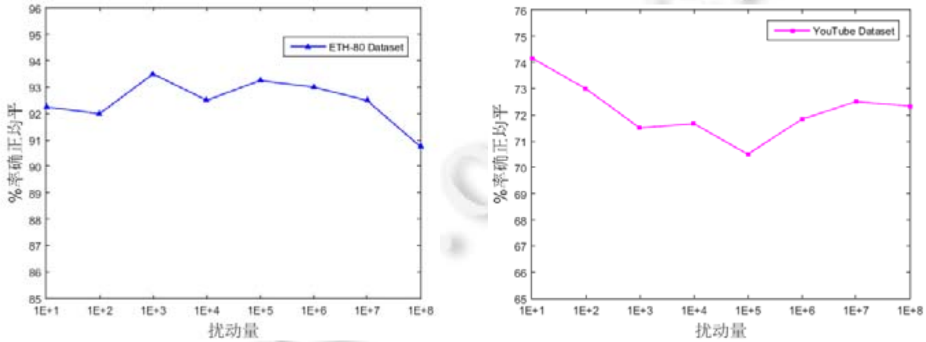


Fig.6 Effect of the parameter γ on the accuracy of the proposed method on ETH-80 and YouTube dataset

图 6 在 ETH-80 和 YouTube 上参数γ对实验结果的影响

从图 6 可以看出:随着扰动量γ值的变化,实验结果也出现了一定的波动.但是可以发现,实验结果的整体变化是比较平稳的,这说明本文提出的算法在参数值的改变对实验结果造成的影响上表现出了较好的鲁棒性.从图中可以看出:对于 ETH-80 数据集,扰动量γ的值在区间 1E+3~1E+6 之间取得时都是可以接受的;而对于 YouTube 数据集,在 1E+3~1E+7 之间取得时是比较理想的.

Grassmann 流形是由线性子空间张成的,因此线性子空间的维数 q 的取值对实验结果具有较大的影响,所以我们固定其他参数,在 Honda 数据集上进行实验,观察当 q 变化时实验结果的变化情况,实验结果如图 7 所示.

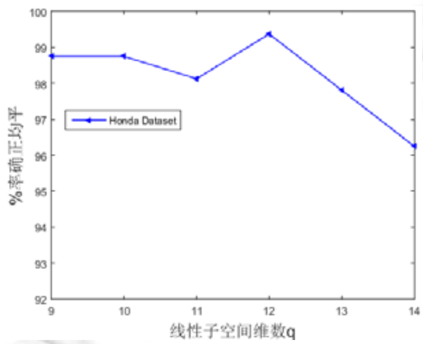


Fig.7 Effect of the parameter q on the accuracy of the proposed method on Honda dataset

图 7 在 Honda 数据集上参数 q 对实验结果的影响

从图 7 可以看出,实验结果整体呈现先增后减的平稳变化趋势.通过对图 7 的观察,最佳线性子空间维数 q 的取值为 12.由分析可知:如果线性子空间的维数过小,会导致 Y_i 中保留的有效的特征信息量就较少;而如果线性子空间的维数较大,那么 Y_i 中除了包含有效的信息量以外,还有部分冗余的信息量.这两种情况都会对实验结果产生影响.

通过前面的实验分析,我们可以看出,不同的参数在不同的取值下对实验结果的确有着不同程度的影响.对

于我们的优化任务,迭代次数也是很重要的参数,不同的迭代次数,对我们最终所获取的变换矩阵 W 可能也会有一定的不同,因此我们固定其他参数,在 YouTube 数据集上进行实验,观察实验结果的变化情况,如图 8 所示.

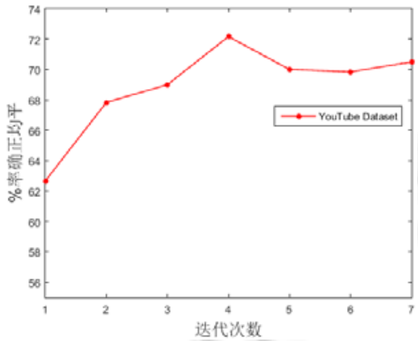


Fig.8 Effect of the number of iterations on the accuracy of the proposed method on YouTube dataset

图 8 在 YouTube 上迭代次数对实验结果的影响

通过图 8 我们可以看出:当迭代次数从 2 开始增加时,实验结果整体呈现比较平稳的变化趋势,然后趋于稳定;当迭代次数为 1 时,由于我们最初赋给变换矩阵 W 的初始值为单位矩阵或者随机数矩阵,因此最终求得的 W 没有经过充分的迭代更新,导致降维后流形的判别性较差,所以实验结果相对于较多的迭代次数下获得的 W 就较低;但是当迭代次数增加到一定值后,由于对 W 的更新已经达到收敛,所以实验结果就趋于了稳定.

4 总结与展望

本文提出了一种基于切空间判别学习的流形降维算法,该算法首先采用传统的针对非满秩对称矩阵的处理方式,也即添加扰动的方式,将 Grassmann 流形上的点变换到 SPD 流形;然后,利用 SPD 流形上常用的度量也即 LEM 将 SPD 流形上的点映射到一个切空间中;最后,在该切空间中,通过定义判别函数并利用基于特征值分解的迭代优化算法求得最终的变换矩阵.该算法在 3 个不同的数据集上所取得的较好的实验结果验证了该算法的可行性及有效性.虽然我们利用基于特征值分解的迭代优化算法较 RCG 算法加快了训练速度,但是在切空间中进行优化,其结果不一定能完全体现原空间中数据的流形性质,因此对于最终的实验结果会产生一定的影响.如何在快速优化 W 的同时又兼顾数据的流形性,也即寻找新的优化算法,以及如何构造新的判别函数,将是我们下一步研究的重点.

References:

[1] Wold S, Esbensen K, Geladi P. Principal component analysis. Chemometrics and Intelligent Laboratory Systems, 1987,2(1-3): 37–52.

[2] Balakrishnama S, Ganapathiraju A. Linear discriminant analysis—A brief tutorial. In: Proc. of the Institute for Signal and Information Processing. 1998. 18.

[3] Wright J, Yang AY, Ganesh A, *et al*. Robust face recognition via sparse representation. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2009,31(2):210–227.

[4] Shan C, Gong S, McOwan PW. Facial expression recognition based on local binary patterns: A comprehensive study. Image and Vision Computing, 2009,27(6):803–816.

[5] Wang R, Guo H, Davis LS, *et al*. Covariance discriminative learning: A natural and efficient approach to image set classification. In: Proc. of the 2012 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). IEEE, 2012. 2496–2503.

[6] Shu X, Wu XJ. A novel contour descriptor for 2D shape matching and its application to image retrieval. Image and Vision Computing, 2011,29(4):286–294.

[7] Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. Science, 2000,290(5500):2323–2326.

[8] Harandi M, Salzmann M, Hartley R. Dimensionality reduction on SPD manifolds: The emergence of geometry-aware methods. arXiv preprint arXiv:1605.06182, 2016.

[9] Huang ZW, *et al*. Projection metric learning on Grassmann manifold with application to video based face recognition. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2015.

[10] Edelman A, Arias TA, Smith ST. The geometry of algorithms with orthogonality constraints. SIAM Journal on Matrix Analysis and Applications, 1998,20(2):303−353.

[11] Arandjelovic O, Shakhnarovich G, Fisher J, *et al*. Face recognition with image sets using manifold density divergence. In: Proc. of the 2005 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR 2005). IEEE, 2005. 581−588.

[12] Beveridge JR, Zhang H, Flynn PJ, *et al*. The ijcb 2014 pasc video face and person recognition competition. In: Proc. of the 2014 IEEE Int'l Joint Conf. on Biometrics (IJCB). IEEE, 2014. 1−8.

[13] Wang W, Wang R, Huang Z, *et al*. Discriminant analysis on Riemannian manifold of Gaussian distributions for face recognition with image sets. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2015. 2048−2057.

[14] Lee KC, Ho J, Yang MH, *et al*. Video-Based face recognition using probabilistic appearance manifolds. In: Proc. of the 2003 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. IEEE, 2003. I-313−I-320.

[15] HBaudat G, Anouar F. Generalized discriminant analysis using a kernel approach. Neural Computation, 2000,12(10):2385.

[16] Faraki M, Harandi MT, Porikli F. Image set classification by symmetric positive semi-definite matrices. In: Proc. of the 2016 IEEE Winter Conf. on Applications of Computer Vision (WACV). IEEE, 2016. 1−8.

[17] Pennec X, Fillard P, Ayache N. A Riemannian framework for tensor computing. Int'l Journal of Computer Vision, 2006,66(1): 41−66.

[18] Yamaguchi O, Fukui K, Maeda K. Face recognition using temporal image sequence. In: Proc. of the '98 3rd IEEE Int'l Conf. on Automatic Face and Gesture Recognition. IEEE, 1998. 318−323.

[19] Kim TK, Kittler J, Cipolla R. Discriminative learning and recognition of image set classes using canonical correlations. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2007,29(6):1005−1018.

[20] Cevikalp H, Triggs B. Face recognition based on image sets. In: Proc. of the 2010 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). IEEE, 2010. 2567−2573.

[21] Absil PA, Mahony R, Sepulchre R. Optimization Algorithms on Matrix Manifolds. Princeton University Press, 2009.

[22] Cetingul HE, Vidal R. Intrinsic mean shift for clustering on Stiefel and Grassmann manifolds. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2009). IEEE, 2009. 1896−1902.

[23] Hamm J, Lee DD. Extended grassmann kernels for subspace-based learning. In: Proc. of the Advances in Neural Information Processing Systems. 2009. 601−608.

[24] Wong YC. Differential geometry of Grassmannmanifolds. Proc. of the National Academy of Sciences of the United States of America, 1967,57(3):589.

[25] Huang Z, Wang R, Shan S, *et al*. Log-Euclidean metric learning on symmetric positive definite manifold with application to image set classification. In: Proc. of the 32nd Int'l Conf. on Machine Learning (ICML 2015). 2015. 720−729.

[26] Cunningham JP, Ghahramani Z. Linear dimensionality reduction: Survey, insights, and generalizations. Journal of Machine Learning Research, 2015,16:2859−2900.

[27] Arsigny V, Fillard P, Pennec X, *et al*. Log-Euclidean metrics for fast and simple calculus on diffusion tensors. Magnetic Resonance in Medicine, 2006,56(2):411−421.

[28] Arsigny V, Fillard P, Pennec X, *et al*. Geometric means in a novel vector space structure on symmetric positive-definite matrices. SIAM Journal on Matrix Analysis and Applications, 2007,29(1):328−347.

[29] Uzair M, Mahmood A, Mian A, *et al*. A compact discriminative representation for efficient image-set classification with application to biometric recognition. In: Proc. of the 2013 Int'l Conf. on Biometrics (ICB). IEEE, 2013. 1−8.

[30] Cheng SH, Higham NJ, Kenney CS, *et al*. Approximating the logarithm of a matrix to specified accuracy. SIAM Journal on Matrix Analysis and Applications, 2001,22(4):1112−1125.

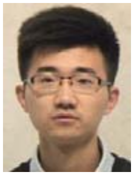
[31] Kokiopoulou E, Chen J, Saad Y. Trace optimization and eigenproblems in dimension reduction methods. Numerical Linear Algebra with Applications, 2011,18(3):565−602.

[32] Viola P, Jones MJ. Robust real-time face detection. Int'l Journal of Computer Vision, 2004,57(2):137−154.

[33] Leibe B, Schiele B. Analyzing appearance and contour based methods for object categorization. In: Proc. of the 2003 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. IEEE, 2003. II-409-15.

[34] Wang R, Shan S, Chen X, *et al*. Manifold-Manifold distance with application to face recognition based on image set. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2008). IEEE, 2008. 1−8.

[35] Hamm J, Lee DD. Grassmann discriminant analysis: A unifying view on subspace-based learning. In: Proc. of the 25th Int'l Conf. on Machine Learning. ACM Press, 2008. 376−383.



王锐(1992－),男,安徽寿县人,博士生,主要研究领域为模式识别,机器学习.

吴小俊(1967－),男,博士,教授,博士生导师,主要研究领域为人工智能,模式识别,计算机视觉.