

自适应半监督边界费舍尔分析

姜伟^{1,2} 杨炳儒¹ 隋海峰¹

(北京科技大学信息工程学院 北京 100083)¹ (辽宁师范大学数学学院 大连 116029)²

摘要 基于图的半监督算法已经成功地应用于人脸识别中, 算法不仅考虑带标签数据而且利用一致性的假设。传统的算法一致性约束是定义在原特征空间中, 但是在原特征空间中定义的一致性不是最好的。提出了自适应半监督边界费舍尔分析算法, 它将一致性约束定义在原特征空间和期望低维特征空间中。在 CMU PIE 和 YALE-B 数据库上进行了实验, 结果表明自适应半监督边界费舍尔分析算法在人脸识别率上有显著的提高。

关键词 判别结构, 半监督, 边界费舍尔分析

中图法分类号 TP181 文献标识码 A

Adaptive Semi-supervised Marginal Fisher Analysis

JIANG Wei^{1,2} YANG Bing-ru¹ SUI Hai-feng¹

(School of Information Engineering, University of Science and Technology Beijing, Beijing 100083, China)¹

(School of Mathematics, Liaoning Normal University, Dalian 116029, China)²

Abstract Graph based semi-supervised methods have successfully used in face recognition. These algorithms not only consider the label information, but also utilize a consistency assumption. Conventional algorithms assumed that the consistency constraint is defined on the original feature space. However, the original feature space is not the best for defining consistency. We proposed adaptive semi-supervised marginal fisher analysis (ASMFA) by which the consistency constraint is defined in the original feature space and the expected low-dimensional feature space. Experimental results on the CMU PIE and YALE-B databases demonstrate that ASMFA brings significant improvement in face recognition accuracy.

Keywords Discriminant structure, Semi-supervised, Marginal fisher analysis

1 概述

子空间学习把高维空间中松散分布的样本通过线性或非线性变换压缩到一个低维的子空间中, 使样本在该低维子空间中分布更紧凑, 更有利于分类。两个经典的子空间降维算法是主成分分析(PCA)^[1] 和线性判别分析(LDA)^[2]。若数据分布于全局线性结构上, 两个算法能够获得数据集真实的内蕴几何结构。若数据点分布于高度非线性的几何结构上, 两种算法将不能胜任, 原因是算法忽视了数据的局部几何结构。流形学习算法能够揭示数据内在的非线性结构, 寻找高维数据在低维空间中的紧致嵌入。但是, 流形学习算法只能在给定的数据上降维, 不能同时给出具有泛化能力的非线性映射。针对这个问题, 提出了多个线性化版本的算法, 比较有代表性的是局部保持投影(LPP)^[3]。LPP 算法是非线性方法 Laplacian Eigenmap 的线性近似, 能够较好地保持原数据的局部结构。LPP 具有局部保持的特性, 但没有考虑类别信息。基于 LPP 的线性化框架, 引入鉴别分析的思想, 衍生出图嵌入框架中的 MFA^[4] 算法、LDE^[5] 算法, 其在鉴别分析中既考虑局

部近邻关系, 又能进行线性映射。上面的算法要么是无监督的, 要么是有监督的。但是, 在许多学习任务中, 对无标签数据进行大量的标注, 会耗费很大的代价。如果只使用少量的标记样本, 那么训练出的学习系统很难具有良好的泛化能力。如果只使用未标记样本, 则浪费了标记样本中所提供的有用信息。因此, 无论是监督算法还是无监督算法都不能获得令人满意的结果, 由此, 半监督学习应运而生。由维数危机理论可以知道, 样本在高维空间中总是非常稀疏的。为了得到理想的泛化性能, 可以将数据的流形几何结构作为正则项来平滑问题的解。Cai 等^[6] 提出半监督判别分析(SDA), 其不仅考虑带标签的数据, 而且利用一致性的假设, 通过在原特征空间构建所有样本的近邻图来估计数据空间的几何结构, 然后将其作为正则化项添加到 LDA 的目标函数中。Dong 等^[7] 提出的半监督判别分析算法是在期望低维特征空间构建样本的近邻图。我们所提出的新算法是根据原特征空间数据的分布与期望的低维特征空间数据的分布构建近邻图来估计数据的几何结构, 然后将其作为正则化项添加到 MFA 目标函数中。由于图的邻域关系需在原空间与期望的低维空间中确定, 因

到稿日期: 2010-04-06 退修日期: 2010-09-06 本文受国家自然科学基金项目(60875029)资助。

姜伟(1969—), 男, 博士生, 副教授, 主要研究方向为数据挖掘、流形学习、模式识别, E-mail: swxxjw@yahoo.com.cn; 杨炳儒(1943—), 男, 教授, 博士生导师, 主要研究方向为知识工程、数据挖掘和人工智能; 隋海峰(1974—), 男, 博士生, 主要研究方向为知识工程、数据挖掘和人工智能。

此需用一个循环程序来确定图的邻域关系,我们把它称为自适应半监督边界费舍尔分析。

2 算法描述

2.1 边界费舍尔分析(MFA)

MFA 使用类内图 G_w 和类间图 G_b 对数据流形的几何结构和判别结构构建模。对于每个数据点 x_i ,首先找到 k 个近邻,记为 $N(x_i)=\{x_i^1, \dots, x_i^k\}$ 。令 $l(x_i)$ 为数据点 x_i 的类别,则 $N(x_i)$ 自然分为两个子集, $N_b(x_i)$ 和 $N_w(x_i)$ 。 $N_w(x_i)$ 包含与 x_i 类别相同的近邻,而 $N_b(x_i)$ 包含与 x_i 不同类别的近邻,即

$$N_w(x_i)=\{x_j^l \mid l(x_j^l)=l(x_i), 1 \leq j \leq k\} \quad (1)$$

$$N_b(x_i)=\{x_j^l \mid l(x_j^l) \neq l(x_i), 1 \leq j \leq k\} \quad (2)$$

易知, $N_b(x_i) \cap N_w(x_i)=\emptyset$ 且 $N_b(x_i) \cup N_w(x_i)=N(x_i)$ 。对 G_w 中的每个点 x_i ,连接顶点 x_i 和 $N_w(x_i)$ 对应的顶点。对 G_b 中的每个点 x_i ,连接顶点 x_i 和 $N_b(x_i)$ 对应的顶点。定义类内图和类间图的权值

$$W_{b,ij}=\begin{cases} 1, & \text{if } x_i \in N_b(x_j) \text{ or } x_j \in N_b(x_i) \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

$$W_{w,ij}=\begin{cases} 1, & \text{if } x_i \in N_w(x_j) \text{ or } x_j \in N_w(x_i) \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

MFA 的目标函数为

$$\arg \max_a \frac{\operatorname{Tr}(a^T X L_b X^T a)}{\operatorname{Tr}(a^T X L_w X^T a)} \quad (5)$$

D_w, D_b 是一个对角阵,它们对角线上各项是相应的 W_w, W_b 列的和, $D_{w,ii}=\sum_j W_{w,ij}$, $D_{b,ii}=\sum_j W_{b,ij}$, $L_w=D_w-W_w, L_b=D_b-W_b$ 。

2.2 自适应半监督边界费舍尔分析

2.2.1 正则化项的确定

已知 m 个数据点 x_1, x_2, \dots, x_m 分布在紧致黎曼流形 M 上,对于每个数据点 x_i , $N(x_i)=\{x_i^1, \dots, x_i^k\}$ 为 x_i 的 k 个近邻。 $N_w(x_i)$ 为 x_i 类别相同的近邻, $N_b(x_i)$ 为 x_i 不同类别的近邻, $N_u(x_i)=N(x_i)-N_w(x_i)-N_b(x_i)$ 为 x_i 的 k 近邻但没有类别标签数据点的集合。

易知, $N_b(x_i) \cap N_w(x_i) \cap N_u(x_i)=\emptyset$ 且 $N_b(x_i) \cup N_w(x_i) \cup N_u(x_i)=N(x_i)$ 。设 a 是由原空间向低维期望特征空间的投影向量,数据点在低维期望特征空间的投影为 $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m$,每一个数据点 \bar{x}_i 的 k 个近邻记为 $N^{apt}(\bar{x}_i), N_w^{apt}(\bar{x}_i)$ 为 \bar{x}_i 类别相同的近邻, $N_b^{apt}(\bar{x}_i)$ 为 \bar{x}_i 不同类别的近邻, $N_u^{apt}(\bar{x}_i)=N^{apt}(\bar{x}_i)-N_w^{apt}(\bar{x}_i)-N_b^{apt}(\bar{x}_i)$ 为 \bar{x}_i 的 k 近邻但没有类别标签数据点的集合。同样, $N_w^{apt}(\bar{x}_i) \cap N_w^{apt}(\bar{x}_i) \cap N_u^{apt}(\bar{x}_i)=\emptyset$ 且 $N_w^{apt}(\bar{x}_i) \cup N_w^{apt}(\bar{x}_i) \cup N_u^{apt}(\bar{x}_i)=N^{apt}(\bar{x}_i)$ 。数据点 x_i, \bar{x}_i 与它们的 k 近邻之间建立带权近邻图为 $G^{apt}(x_i, \bar{x}_i), G^{apt}(\bar{x}_i, \bar{x}_i)$ 的权矩阵 $W_i^{apt}(x_i, \bar{x}_i)$ 定义为

$$W_i^{apt}(x_i, \bar{x}_i)=\begin{cases} \gamma, & \text{if } x_i \in N_w(x_j) \text{ or } x_j \in N_w(x_i) \\ \gamma, & \text{if } x_i \in N_w(x_j) \text{ and } x_j \in N_u(x_i) \text{ and} \\ & \bar{x}_i \in N^{apt}(\bar{x}_j) \text{ and } (\forall \bar{x}_s \in N^{apt}(\bar{x}_j) \\ & \bar{x}_s \in N_w^{apt}(\bar{x}_j)) \\ \lambda, & \text{if, } x_i \in N_w(x_j) \text{ and } x_i \in N_u(x_j) \text{ and} \\ & \bar{x}_i \in N^{apt}(\bar{x}_j) \text{ and } (\forall \bar{x}_s \in N^{apt}(\bar{x}_j) \\ & \bar{x}_s \in N_w^{apt}(\bar{x}_j)) \\ 1, & x_i \in N_u(x_j) \text{ or } x_j \in N_u(x_i) \\ & \text{and } \bar{x}_i \in N_u^{apt}(\bar{x}_j) \text{ or } \bar{x}_j \in N_u^{apt}(\bar{x}_i) \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

由式(6)所定义的相似矩阵记为 W^{apt} 。目标函数为

$$\min Tr(a^T \tilde{X} L^{apt}(a) \tilde{X}^T a) \quad (7)$$

$D^{apt}(a)$ 为对角矩阵,对角元素为相似矩阵 W^{apt} 对应的列或行上的元素之和, $D_{ii}^{apt}=\sum_j W_{ji}^{apt}$, 拉普拉斯矩阵 $L^{apt}(a)=D^{apt}(a)-W^{apt}(a)$ 。

2.2.2 目标函数的确定与算法描述

综合式(10)、式(12)得到目标函数:

$$\arg \max_a \frac{\operatorname{Tr}(a^T X L_b X^T a)}{\operatorname{Tr}(a^T X L_w X^T a)+\mu \operatorname{Tr}(a^T \tilde{X} L^{apt}(a) \tilde{X}^T a)} \quad (8)$$

由于 L^{apt} 的计算由投影向量 a 决定,反之亦然,因此目标函数没有闭式解。我们提出了一个迭代的求解算法来优化这个目标函数。其核心思想是固定 a ,求出一个最优的 L^{apt} ,然后固定 L^{apt} 求出一个最优的 a 。不停地迭代,可以得到一个局部最优解。

给定一个 a^{t-1} ,计算 $L^{apt}(a^t)$ 时用 a^{t-1} 近似 a^t ,因此目标函数修改为

$$\arg \max_a \frac{\operatorname{Tr}(a^{t-1 T} X L_b X^T a^t)}{\operatorname{Tr}(a^{t-1 T} X L_w X^T a^t)+\mu \operatorname{Tr}(a^{t-1 T} X L^{apt}(a^{t-1}) X^T a^t)} \quad (9)$$

投影矩阵是如下广义特征值问题的最大特征值对应的特征向量集

$$X L_b X^T a^t=\lambda' [X L_w X^T+\mu \tilde{X} L^{apt}(a^{t-1})] \tilde{X}^T a^t \quad (10)$$

式中, μ 是一个常数, $0 \leq \mu \leq 1$,通过交叉验证确定。

算法描述如下:

数据 $X=\{x_1, x_2, \dots, x_l, x_{l+1}, \dots, x_n\} \subset R^m$, 其中 m 表示观测空间的维数, l 表示带标签的样本的数目, n 表示数据集 X 中全部样本的数目。带标签的样本 $x_i \mid_{i=1}^N \in R^m$, 无标签样本 $x_i \mid_{i=l+1}^n$, 降维后的维数为 d , 循环次数为 T_{\max} 。

1) 初始化 $a^0=I$ 。

2) For $t=1, 2, \dots, T_{\max}$ Do

根据 a^{t-1} 计算权矩阵 $W^{apt}(a^{t-1})$;

根据 $W^{apt}(a^{t-1})$ 计算拉普拉斯矩阵 $L^{apt}(a^{t-1})$;

由式(10)计算投影矩阵 a^t 。

3) 输出投影矩阵 $a=a^t$ 。

3 实验与分析

实验采用了两个比较著名的人脸数据库,分别是 CMU PIE 人脸库和 Yale-B 人脸库。图 1 显示了 CMU PIE, Yale-B 人脸数据库中的部分人脸样本。在实验中,所有人脸图像都经过了标准化处理,即首先标定出人脸图像中两眼的位置并在该位置将图像对齐,然后剪切出图像中的面部区域并将其缩放成统一的 32×32 大小。图像最终按列堆叠成 1024 维的向量。



图 1 顶行是 CMU PIE 数据集,底行是 Yale-B 数据集

实验首先利用这些算法将数据降低至低维空间,然后用简单的最近邻分类器对得到的低维特征进行分类。为了避免小样本问题,采用 PCA 算法对数据进行降维。对于 MFA 算法,所保留的 PCA 主分量的个数为 $m-N_c$,这里 m 是训练

(下转第 262 页)

- Multiplayer Computer Games[J]. The Electronic Library, 2002, 20(2):87-97
- [5] 张渊. 空间二叉树排序查找算法及其在网络游戏中的应用[J]. 计算机应用, 2007, 27: 356-359
- [6] Assiots M, Tzanov V. A Distributed Architecture for MMORPG[C]// Proceedings of 5th ACM SIGCOMM Workshop on Network and System Support for Games. Singapore, October 2006
- [7] Bettstetter C, Hartenstein H, Xavier Pérez-Costa. Stochastic Properties of The Random Waypoint Mobility Model; Epoch Length, Direction Distribution, and Cell Change Rate[C]// Proceedings of the 5th ACM International Workshop on Modeling Analysis and Simulation of Wireless and Mobile Systems. USA, September 2002
- [8] Yuan Feng. Windows Graphics Programming: Win32 GDI and DirectDraw[M]. Hewlett-Packard Professional Books, 2001
- [9] Hsiao T Y, Yuan S M. Practical Middleware for Massively Multiplayer Online Games[J]. IEEE Internet Computing, 2005, 9(5):47-54

(上接第 253 页)

数的个数, N_c 是类数。

3.1 在数据集 CMU PIE 上的实验

CMU PIE 人脸数据库共包含 68 人在不同的姿态和光照条件下的共 41368 幅图片, 包含各种光照和表情的变化。实验中选取了所有 68 人在 5 种近正面姿态(C05, C07, C09, C27, C29)下的一种近正面姿态(C27), 每个人 49 幅图片。包含各种光照和表情的变化, 实验从每人的图像中随机抽取 30 幅作为训练样本, 其余的作为测试样本。为得到随机的训练集、测试集划分, 实验重复了 10 轮。每轮实验中, 又从 30 幅训练图像中随机选取 $L(L=2,3)$ 幅作为有标签样本, 其余的作为无标签样本。对每个 L 值实验又重复了 10 次。最终的识别率取其平均。当 α 分别取 0.01, 0.02, 0.04, 0.06, 0.08 时, 最优识别率的变化曲线如图 2 所示。

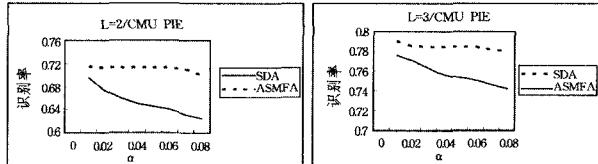


图 2 最优识别率随 α 变化曲线

SDA 算法随着 α 的不同取值识别率变化非常明显, 而 ARSMFA 算法对 α 具有鲁棒性。当 α 取 0.01 时, 最优识别率随特征维数变化曲线如图 3(a)、(b) 所示。

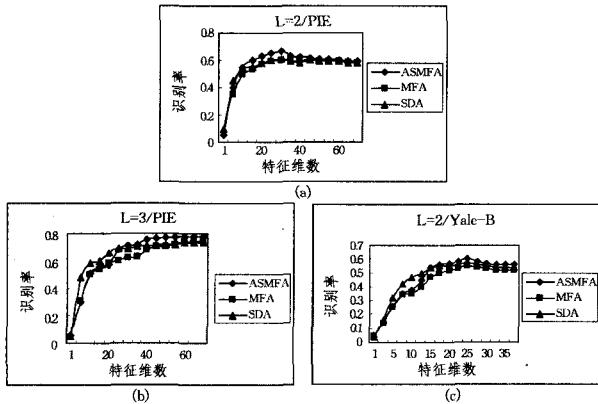


图 3 最优识别率随特征维数变化曲线

3.2 在数据集 Yale-B 上的实验

Yale-B 人脸数据库共包含 28 个人在 9 种姿态、64 种光照下的 16128 幅人脸图像, 每人 576 幅图片。从数据库中选择 10 人, 每人选择 100 幅图像。实验从每人 100 幅图像中随机抽取 30 幅作为训练样本, 其余的作为测试样本, 为得到随机的训练集、测试集划分, 实验重复了 10 轮。每轮实验中, 又

从 30 幅训练图像中随机选取 $L(L=2,3)$ 幅作为有标签样本, 其余的作为无标签样本, 对每个 L 值实验又重复了 10 次。最优识别率随特征维数变化曲线如图 3(c) 所示。

由实验结果可知, ASMFa 算法的识别性能超过 MFA 算法, 说明了无标签数据能够被用来提高人脸识别算法的性能。ASMFa 算法的识别性能也超过 SDA, 说明了基于原空间和期望的低维空间融合所建立的正则化项能够提高识别性能。SDA 随着 α 的不同取值发生了显著的变化, 新算法对 α 的取值有很好的鲁棒性。在 Yale-B 和 CMU PIE 人脸数据库的实验表明, ASMFa 能够充分利用样本的类别信息和样本原空间与期望的低维空间的局部几何结构, 可以得到比 MFA 和 LDA 算法更优的性能。

结束语 人脸图像是高维数据, 获得足够的带标签的训练样本是不可能的。当每类训练样本很少时, 进行人脸识别不能得到理想的识别效果。随着数字图像技术的发展, 收集大量的无标签的人脸图像却是可能的, 因此, 人脸识别可以看作是一个半监督学习问题。研究表明人脸数据处在一个具有较低维数的流形上。传统算法在原空间中定义正则化项, 而本文在半监督思想基础上探索出一种新的正则化项的创建方法, 这种将半监督的思想与流形学习结合用于人脸识别中取得了很好的效果。

参 考 文 献

- [1] 谢永华, 严云洋, 杨静宇. 分块 PCA 鉴别特征抽取能力的分析研究[J]. 计算机科学, 2006, 33(3): 155-159
- [2] Sets D L, Weng J. Using Discriminant Eigenfeatures for image Retrieval[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 1996, 18(8): 831-836
- [3] He Xiaofei, Yan Shuicheng, Hu Yuxiao, et al. Face Recognition Using Laplacianfaces[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2005, 27(3): 328-340
- [4] Yan Shuicheng, Xu Dong, Zhang Benyu, et al. Graph Embedding and Extensions: A General Framework for Dimensionality Reduction[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2007, 29(1): 40-51
- [5] Chen H T, Chang H W, Liu T L. Local discriminant embedding and its variants[J]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, 2(1): 846-853
- [6] Cai Deng, He XiaoF, Han J W. Semi-supervised discriminant analysis[C]// Proc. of the IEEE Int'l Conf. on Computer Vision. Rio de Janeiro, 2007
- [7] Xu Dong, Yan Shui-cheng. Semi-supervised Bilinear Subspace Bilinear Subspace Learning[J]. IEEE Trans. on Image Processing, 2009, 19(7): 1671-1676