

一种基于局部和判别特性的降维算法

张国印, 楼宋江, 王庆军, 程慧杰

(哈尔滨工程大学 计算机科学与技术学院, 哈尔滨 150001)

摘 要: 提出了一种基于 LPP 和 LDA 的降维算法。该算法不仅考虑了 LPP 能保持局部邻近关系属性, 还考虑了 LDA 能使降维后的数据更易于分类属性, 并且该算法是线性的, 很容易将新样本映射到目标空间。在人脸识别中的实验验证了算法的正确性和有效性。

关键词: 维数约简; 局部保持投影; 线性判别分析; 人脸识别

中图分类号: TP391

文献标志码: A

文章编号: 1001-3695(2009)09-3324-02

doi:10.3969/j.issn.1001-3695.2009.09.035

Dimension reduction algorithm based on locality and discriminant characteristics

ZHANG Guo-yin, LOU Song-jiang, WANG Qing-jun, CHEN Hui-jie

(College of Computer Science & Technology, Harbin Engineering University, Harbin 150001, China)

Abstract: This paper proposed a dimension reduction algorithm based on LPP and LDA, which not only utilized that LPP could preserve locality, but also took into account that the LDA made the data more discriminant and thus made classification more easier. The proposed algorithm was linear, which could map the out of sample data to the target space easily. Experiments in face recognition validate the correctness and effectiveness of the algorithm.

Key words: dimension reduction; locality preserving projection(LPP); linear discriminant analysis(LDA); face recognition

0 引言

在生物特征识别系统中, 人脸识别由于具有成本较低、不侵犯使用者的隐私权、应用前景广阔等特点, 受到了极大的关注, 是当前模式识别、图像处理和计算机视觉等学科非常活跃的研究热点^[1]之一。

正是因为人脸识别的这些特点, 在过去的几年时间里, 出现了很多人脸识别算法。其中, 基于子空间的算法起到了主导作用, 最著名的是主成分分析^[2] (principal component analysis) 和线性判别分析^[3] (linear discriminant analysis)。主成分分析算法的主要思想是通过估计数据的二阶统计性质来发现数据的本征线性维数, 它是在全局最小重构误差的条件下把高维观察数据投影到低维空间上, 通过求解数据点协方差矩阵的最大几个特征值所对应的特征向量张成的子空间而得到。线性判别分析算法就是求出一个线性子空间, 使得所有样本在这个子空间中, 类内样本散度最小, 类间样本的散度最大, 用它降维后的低维嵌入坐标非常有利于进行样本的分类。

上述两种方法是线性算法, 但人脸空间更可能存在非线性子空间中。最近相关人员提出了几种流形学习算法, 主要用于非线性降维, 如等距映射 (ISOMAP)^[4]、局部线性嵌入算法 (LLE)^[5]、拉普拉斯特征映射 (LE)^[6] 等。但是它们都不能解决 out of sample 问题, 即不能直接映射新数据, 因此不能直接用于人脸识别。基于以上问题, He 等人^[7] 提出了一种线性算法 LPP, 该算法是拉普拉斯特征映射的线性版本, 能保持数据

集的局部邻近关系, 又能较容易地映射新样本。

LPP 虽然能保持数据的局部结构, 但是没有利用数据集的类别信息, 从而不利于分类问题。本文基于 LPP 和 LDA 的优点, 既考虑 LPP 的局部保持特性, 同时又利用 LDA 的易于分类问题这一特点, 提出了一种基于局部特性和判别特性的降维算法。

1 基于局部特性和判别特性降维算法

1.1 LPP 简介

LPP 目标是保持数据之间的相似性, 即原始空间上相邻的数据点在投影空间上也保持相应的邻近关系。假定数据集 $X = \{x_1, x_2, \dots, x_n\}$ 含有 n 个样本点, 每个样本点 x_i 属于 D 维欧氏空间, 即 $x_i \in R^D$, 且样本数据来自一本征维数为 d ($d < D$) 的非线性流形。对于原始空间中的一数据点 x_i , 寻找一投影 W , 使得 $y = W^T x$, 通过求解以下最小化问题得到最佳投影:

$$\begin{aligned} W_{\text{opt}} &= \arg \min_W \sum_{ij} (y_i - y_j)^2 S_{ij} = \\ &= \arg \min_W \sum_{ij} (W^T x_i - W^T x_j)^2 S_{ij} = \\ &= \arg \min_W W^T X L X^T W \end{aligned} \quad (1)$$

其中: S_{ij} 刻画了两数据 x_i, x_j 之间的相似性。

$$S_{ij} = \begin{cases} \exp(-\|x_i - x_j\|^2/t) & x_i, x_j \text{ 为邻近点} \\ 0 & \text{其他} \end{cases}$$

附加约束条件 $W^T X D X^T W = 1$, 拉普拉斯算子 $L = D - S$ 。

其中: D 为对角矩阵, 其元素为对称矩阵 S 的列向量 (或行向

收稿日期: 2008-09-17; 修回日期: 2009-02-26

作者简介: 张国印 (1962-), 男, 哈尔滨人, 教授, 博士, 主要研究方向为嵌入式、信息安全、模式识别; 楼宋江 (1982-), 男, 博士研究生, 主要研究方向为模式识别、机器学习 (lousongjiangac@163.com)。

量)元素之和,即 $D_{ij} = \sum_j S_{ij}$ 。最后可以通过以下特征方程求解:

$$XLX^T w = \lambda XD X^T w \tag{2}$$

假定 $w_0, w_1, \dots, w_{d-1}$ 为特征方程式(2)的 d 个解,它们所对应的特征值分别为 $\lambda_0, \lambda_1, \dots, \lambda_{d-1}$,那么可以得到如下映射: $x_i \rightarrow y_i = W^T x_i$ 。其中: y_i 为 d 维向量; W 为变换矩阵。

1.2 LDA 简介

对于分类问题,希望能找到某个投影方向,使得不同类的数据在投影后能尽量分开。LDA 算法就是找到这样一投影方向,使得数据在低维空间中同类数据尽量靠近,不同类数据尽量分离,从而保留丰富的辨别信息,使数据具有最大的可分性。

给定分别属于 J 类的 n 个数据样本 $X = \{x_1, x_2, \dots, x_n\}$, C_j 表示第 j 类元素构成的集合, n_j 表示属于第 j 类元素的个数, $\mu_j = (1/n_j) \sum_{x_i \in C_j} x_i$ 表示第 j 类的均值, μ 表示样本的总体均值。

令 S_b 、 S_w 分别为类间散度矩阵和类内散度矩阵,即

$$S_b = (1/n) \sum_{i=1}^J n_i (\mu_i - \mu) (\mu_i - \mu)^T \tag{3}$$

$$S_w = (1/n) \sum_{j=1}^J \sum_{x_i \in C_j} (x_i - \mu_j) (x_i - \mu_j)^T \tag{4}$$

问题归结为求以下特征方程:

$$S_b w = \lambda S_w w \tag{5}$$

设 $w_1, w_2, \dots, w_d$ 是对应于矩阵 $S_w^{-1} S_b$ 的前 d 个最大特征值的特征向量,则令 $W = [w_1, w_2, \dots, w_d]$,投影后的数据为 $y_i = W^T x_i, i \in n$ 。

1.3 基于局部和判别特性的降维算法

基于以上的简单介绍可以看出 LPP 能有效降低数据维数,并能保持数据间的邻近关系,LDA 能找到一投影,使得投影后的数据更加有利于分类。本节将 LPP 和 LDA 的这两个特性结合起来,形成既保持数据点之间的关系又有利于分类的新降维算法。

找到一投影,使得目标函数 $W^T (S_w + \sigma XLX^T) W / (W^T S_b W)$ 最小(S_w 、 S_b 、 L 定义同上),并受到一限制条件 $W^T XD X^T W = 1$ 。所以问题最后归结为以下优化问题:

$$\begin{cases} W_{opt} = \arg \min_W W^T (S_w + \sigma XLX^T) W / (W^T S_b W) \\ \text{s. t.} & W^T XD X^T W = 1 \end{cases} \tag{6}$$

这里假设 S_b 可逆,即假设 S_b 非奇异,则该优化问题可以简化成

$$\begin{cases} W_{opt} = \arg \min_W W^T S_b^{-1} (S_w + \sigma XLX^T) W \\ \text{s. t.} & W^T XD X^T W = 1 \end{cases} \tag{7}$$

通过 Lagrange 乘数法,令

$$\begin{aligned} L(W) &= W^T S_b^{-1} (S_w + \sigma XLX^T) W - \\ &\quad \lambda (W^T XD X^T W - 1) \end{aligned} \tag{8}$$

对 W 求偏导,得到

$$\partial L(W) / \partial W = 2 S_b^{-1} (S_w + \sigma XLX^T) W - 2 \lambda XD X^T W \tag{9}$$

令 $\partial L(W) / \partial W = 0$,即

$$S_b^{-1} (S_w + \sigma XLX^T) W - \lambda XD X^T W = 0 \tag{10}$$

最后求解以下广义特征方程

$$S_b^{-1} (S_w + \sigma XLX^T) w = \lambda XD X^T w \tag{11}$$

假定 $w_0, w_1, \dots, w_{d-1}$ 为式(11)的 d 个解,它们所对应的特征值分别为 $\lambda_0, \lambda_1, \dots, \lambda_{d-1}$,那么可以得到如下映射: $x_i \rightarrow y_i = W^T x_i$ 。其中: y_i 为 d 维向量; W 为变换矩阵。

在求解过程中, S_b 有可能是奇异的,为了解决该问题,可以先通过 PCA 算法,使得 S_b 变成非奇异。所以具体算法步骤如下:

- a) 对原数据先进行主成分分析,一方面既可以消除噪声,又可以避免 S_b 的奇异性,映射图像 X 到 PCA 子空间,即 $Y = W_{PCA}^T X$ 。
- b) 根据式(3)(4)求得 S_w 、 S_b ,以及拉普拉斯 L 和 D 。
- c) 根据式(10)解决特征函数问题,得到特征值和所对应的特征向量,求得 W_{opt} 。
- d) 令 $W = W_{PCA} W_{opt}$,根据 $Y = W^T X$,将图像 X 映射到目标子空间,求得目标图像。

2 实验结果及其分析

本算法采用公共人脸数据库 ORL 库来验证算法的正确性和有效性。实验过程为:先将人脸库中的图像进行光照预处理;再利用本文所提出的算法进行低维人脸特征提取;最后使用 NN(nearest neighborhood)分类器进行识别。实验中比较了该算法与 PCA、LDA、LPP 的识别率。

ORL 人脸数据库包含 40 人,每人有 10 张人脸图片,共 400 张图片。其中有一些图像是在不同时期拍摄的,人的表情和脸部细节有着不同程度的变化,如睁眼和闭眼,是否带有眼镜等。人脸姿态也有一定程度的变化,深度旋转和平面选择可达 20°,人脸大小也有多达 10% 的变化。图 1 是 ORL 人脸数据库中的第一个人的样本。



图1 ORL人脸图像的样本

为了有效地计算,所有图像都缩放成 32×32 像素,即每一图像可以看成是 $D = 32 \times 32 = 1\,024$ 维向量。参数 σ 是决定局部信息和类别信息对识别率贡献的大小,这个很难确定,在实验中,取 $\sigma = 1$,即假设局部信息和类别信息同等重要。每人分别随机选择 3、4、5、6、7 幅图像用做训练样本,其他图像用于测试样本,重复 20 次,最后取平均值作为识别结果。该算法(姑且称它为 DRLD)与著名算法 PCA、LDA、LPP 进行了比较,实验结果如表 1 所示。

表 1 算法在 ORL 人脸库中的识别结果

训练样本数	PCA/%	LDA/%	LPP/%	DRLD/%
3	78.2	84.7	82.1	88.7
4	83.7	90.8	87.6	91.3
5	86.8	93.7	89.8	94.5
6	89.1	95.6	93.3	96.2
7	92.4	96.9	94.1	98.3

实验结果表明,LDA 算法的识别率高于 PCA,这是因为 LDA 考虑了类别信息,更加有利于分类问题。而 LPP 虽然高于 PCA,但是没有 LDA 理想,原因是虽然考虑了人脸的局部信息,但是没有考虑类别信息,所以在人脸识别上的效果不如

有监督的学习算法 LDA。DRLD 算法优于 (下转第 3329 页)

没有标出的数据,则 A 与 C 的并集 $A + C$ 表示满足要求的数据, B 与 C 的并集 $B + C$ 表示系统标出的数据。那么系统标注的准确率(P)、召回率(R)、 F 值可定义如下:

$$P = C / (B + C)$$
$$R = C / (A + C)$$
$$F_{\beta} = (\beta^2 + 1) \times P \times R / (\beta^2 \times P + R)$$

上述三个指标是通常使用的评测指标, F_{β} 表示准确率和召回率的平衡, β 通常取 1,即 F_1 值。本文后面实验将采用这三个指标进行结果的评测。

根据测试集和训练集的不同关系,可以将评测分为封闭测试和开放测试。为了能够充分评价基于层叠条件随机场的古文断句与句读标记算法的效果,本文做了四组实验。其中前两组实验是本文算法的封闭测试与开放测试,如表 4 所示;后两组是传统条件随机场模型的封闭测试与开放测试,如表 5 所示。封闭测试时,训练数据为全部古文语料,测试数据是其中随机抽取 20% 组成的子集。开放测试时,训练数据为随机抽取 80% 组成的集合,剩下的 20% 作为测试数据。实验结果表明,基于层叠条件随机场模型的方法将底层模型的断句信息作为特征引入句读标记过程,增加了句读标记阶段可用的特征参数;另一方面,基于层叠模型的方法针对不同阶段的任务设计不同的标记集,便于更好地描述语言特点。因此句子切分和标点标注的性能均有明显提升,尤其是标点符号标注的召回率得到了较大的提高。

表 4 层叠模型实验结果

测试类型	准确率/%		召回率/%		F_1 值/%	
	切分	标点	切分	标点	切分	标点
封闭测试	98.47	92.31	94.57	90.41	96.48	91.35
开放测试	72.15	68.58	70.70	66.78	71.42	67.67

表 5 传统模型实验结果

测试类型	准确率/%		召回率/%		F_1 值/%	
	切分	标点	切分	标点	切分	标点
封闭测试	95.47	88.36	92.57	83.68	94.00	85.96
开放测试	69.15	52.37	66.70	45.97	67.90	48.96

(上接第 3325 页)其他算法,这是因为它不仅考虑了人脸的局部信息,而且还考虑了类别信息,使不同类之间的可区分度加强了。显然随着训练样本数的增加,算法的识别率提高了。

3 结束语

本文提出了一种基于局部特性和判别特性的一种新颖数据降维算法,在一定程度上解决了流形学习方法应用于模式分类中存在的问题。该算法不仅保持了数据点之间的局部邻近关系,而且也考虑了数据之间的类别关系。它是一种线性算法,能将新数据映射到目标空间,即解决流形问题中的 out of sample 问题。算法中有一参数 σ ,它是确定类别信息与局部信息对分类问题的贡献大小。如何确定参数 σ ,如何将算法扩展成非线性使其能更好地用于实际应用(在此可以用核技巧 kernel trick,先将数据映射到高维空间,再利用本文提出的算法),以及将算法用于其他领域是今后的一个研究方向。

参考文献:

[1] ZHAO W, CHELLAPPA R, ROSENFELD A, et al. Face recognition: a literature survey, CVL technical report[R]. Maryland: University

6 结束语

本文研究了 CRF 模型在古文句子切分与标点符号自动标注的应用,设计了能够更好描述古文语言特点的六字位标记集,提出了基于层叠条件随机场模型的算法。实验结果证明,基于层叠条件随机场模型的算法的性能较传统方法有了明显提升,是一种比较理想的古文断句与句读标注方法。

古文可用的特征比较少,如何利用已有信息发掘新的特征,进一步提高古文断句和句读标注算法的性能将是下一步努力的方向。

参考文献:

[1] CHAROENPORNSAWAT P, SORNLERTLAMVANICH V. Automatic sentence break disambiguation for Thai[C]//Proc of ICCPOL '01. 2001:231-235.

[2] 胡俊峰,俞士汶. 唐宋诗之计算机辅助深层研究[J]. 北京大学学报:自然科学版,2001,37(5):727-733.

[3] 陈天莹,陈蓉,潘璐璐,等. 基于前后文 n-gram 模型的古汉语句子切分[J]. 计算机工程,2007,33(2):192-193.

[4] LAFFERTY J, McCALLUM A, PEREIRA F. Conditional random field: probabilistic models for segmenting and labeling sequence data[C]//Proc of the 18th International Conference on Machine Learning. San Francisco: Morgan Kaufmann Publishers, 2001: 282-289.

[5] 刘群,张华平,俞鸿魁,等. 基于层叠隐马模型的汉语词法分析[J]. 计算机研究与发展, 2004,41(8):1421-1429.

[6] 赵海,揭春雨. 基于有效子串标注的中文分词[J]. 中文信息学报,2007,21(5):8-13.

[7] ZHAO Hai, HUANG Chang-ning, LI Mu. An improved Chinese word segmentation system with conditional random field[C]//Proc of the 15th SIGHAN Workshop on Chinese Language Processing. Sydney: [s. n.], 2006:162-165.

[8] NOCEDAL J, WRIGHT S J. Numerical optimization[M]. New York: Springer, 1999:194-200.

of Maryland, 2000.

[2] TURK M, PENTLAND A. Eigenfaces for recognition[J]. Cognitive Neurosci, 1991,3(1):71-86.

[3] BELHUMEUR P N, KRIEGMAN D J. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 1997,19(7):711-720.

[4] TENENBAUM J B, DeSILVA V, LANGFORD J C. A global geometric framework for nonlinear dimensionality reduction[J/OL]. [2000-12-22]. <http://www.sciencemag.org/content/vol290/issue5500/>.

[5] ROWIES S, SAUL L. Nonlinear dimensionality reduction by locally linear embedding[J/OL]. [2000-12-22]. <http://www.sciencemag.org/cgi/content/full/290/5500/2323>.

[6] BELKIN M, NIYOGI P. Laplacian eigenmaps for dimensionality reduction and data representation[J]. Neural Computation, 2003,15(6):1373-1396.

[7] HE Xiao-fei, NIYOGI P. Locality preserving projections[C]//Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2004:327-334.