第 37 卷 第 24 期
Vol.37 No.24

计 算 机 工 程
Computer Engineering

2011 年 12 月
December 2011

·人工智能及识别技术· 文章编号：1000—3428(2011)24—0193—02 文献标识码：A 中图分类号：TP301.6

基于成对约束和稀疏保留的数据降维算法

王颖静，王正群，张国庆，俞振洲

(扬州大学信息工程学院，江苏 扬州 225009)

摘 要：结合以成对约束形式给出的监督信息和无监督信息，提出一种基于成对约束和稀疏保留的数据降维算法。通过成对约束信息进行鉴别分析，利用稀疏表示方法保留数据集在变换空间中的全局稀疏结构。实验结果表明，与传统特征抽取算法相比，该算法的识别效果更好，需要调节的参数更少，且鲁棒性较高。

关键词：稀疏保留；机器学习；特征提取；人脸识别

Dimensionality Reduction Algorithm Based on Pair-wise Constraints and Sparsity Preserving

WANG Ying-jing, WANG Zheng-qun, ZHANG Guo-qing, YU Zhen-zhou

(Collage of Information Engineering, Yangzhou University, Yangzhou 225009, China)

【Abstract】 This paper presents a dimensionality reduction algorithm based on pair-wise constraints and sparsity preserving. It combines some supervised information in the form of pair-wise constraints and large number of unsupervised information. It uses pair-wise constraints to discriminant analysis and uses sparse representation to preserve the sparse reconstructive structure in the transformed space. Compared with the traditional feature extraction method, this algorithm has a better recognition impact, lower parameters, and better robustness.

【Key words】 sparsity preserving; machine learning; feature extraction; face recognition

DOI: 10.3969/j.issn.1000-3428.2011.24.064

1 概述

随着大量高维数据的快速累积，如数字图像、金融时间序列以及基因微阵列等，降维已成为许多数据挖掘任务的一种基础工具。根据样本数据中是否存在监督信息，现有的降维方法可大致分为有监督学习方法和无监督学习方法。

有监督学习和无监督学习是机器学习领域中常用的 2 种算法，然而这 2 种算法都有缺陷。有监督学习必须对所有的学习样本做好类别分类，为提高其推广能力，需要获得大量的学习样本；无监督学习是一种自动学习方式，其分类效果往往不尽如人意。半监督学习将 2 种算法结合起来，利用较少的监督信息来提高分类性能。现有的半监督学习大致可以分为 3 类，即半监督分类、半监督回归、半监督聚类。

半监督分类是从监督学习角度出发，考虑带标签训练样本不足时，如何利用大量无标签样本信息辅助分类器训练[1]。近年来，以成对约束作为监督信息的半监督分类或降维方法颇受人们关注。其中比较有代表性的是文献[2]提出的将成对约束信息与数据的流形结构结合起来构造优化函数，并应用于线性和非线性降维分析。文献[2]是利用局部线性嵌入(Locally Linear Embedding, LLE)算法来保持数据的流形结构，该算法需要确定 3 个参数，即嵌入后的维数、样本近邻的个数以及调整参数。降维的质量与这 3 个参数有很大的关系[3]。出于对算法鲁棒性以及识别性能的考虑，本文将成对约束信息与稀疏保留结合起来，提出一种新的数据降维算法 DR-PCS(Dimension Reduction Based on Pairwise Constraints and Sparity)。

2 相关算法

在本节中将回顾一些通过成对约束来分类的方法。成对约束方法首先是由 Wagstaff 等人提出的，旨在提高无监督聚类算法的性能。正约束规定：如果 2 个样本属于正约束，那么这 2 个样本在聚类时必须分配到同一个聚类中；负约束规定：如果 2 个样本属于负约束，那么这 2 个样本在分类时必须分配到不同的聚类中。

文献[4]提出优化基追踪目标函数如下：

$$\boldsymbol{W}^* = \arg\max_{\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}=\boldsymbol{I}} \frac{tr(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{S}_b\boldsymbol{W})}{tr(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{S}_w\boldsymbol{W})} \tag{1}$$

其中，$\boldsymbol{S}_w$ 是通过正约束计算得到的协方差矩阵；$\boldsymbol{S}_b$ 是通过负约束计算得到的协方差矩阵。利用式(1)的鉴别分析方法，并提出一种二分查找的迭代算法来解决这个优化问题。其基本思想是：最小化正约束集中的点对距离，最大化负约束集中的点对距离。文献[2]在此基础上考虑保留数据集的流形结构，即同时使用成对约束以及局部线性嵌入来构造目标函数以获得变换 $y=f(\boldsymbol{x})$ ($\boldsymbol{x}$ 为向量)：

$$f^* = \arg\max_{f} \frac{\sum_{(\boldsymbol{x}_i,\boldsymbol{x}_j)\in\boldsymbol{D}} \left\| f(\boldsymbol{x}_i)-f(\boldsymbol{x}_j) \right\|^2}{\sum_{(\boldsymbol{x}_i,\boldsymbol{x}_j)\in\boldsymbol{P}} \left\| f(\boldsymbol{x}_i)-f(\boldsymbol{x}_j) \right\|^2 + \alpha J_{\mathrm{LLE}}(f,\boldsymbol{X})} \tag{2}$$

其中，$\boldsymbol{P}$ 是正约束集；$\boldsymbol{D}$ 是负约束集；惩罚项 $J_{\mathrm{LLE}}(f,\boldsymbol{X})$ (矩阵训练样本 $\boldsymbol{X}$)用于保留数据集在变换空间中的拓扑结构；参数 $\alpha \geqslant 0$，用于平衡相似点对距离以及惩罚项之间的关系。

基金项目：国家自然科学基金资助项目(60875004)；江苏省自然科学基金资助项目(BK2009184)；江苏省高校自然科学基金资助项目(10KJB510027, 07KJB520133)

作者简介：王颖静(1987－)，女，硕士研究生，主研方向：机器学习；王正群，教授、博士；张国庆、俞振洲，硕士研究生

收稿日期：2011-07-19 **E-mail**：wyj0314@126.com

$J_{\mathrm{LLE}}(f,\boldsymbol{X})$ 的表达形式如下式所示：

$$J_{\mathrm{LLE}}(f,\boldsymbol{X})=\sum_{i=1}^{n}\left\|f(\boldsymbol{x}_i)-\sum_{\boldsymbol{x}_j\in N_k(\boldsymbol{x}_i)}\boldsymbol{s}_{ij}^{*}f(\boldsymbol{x}_j)\right\|^2=$$
$$tr(\boldsymbol{Y}(\boldsymbol{I}-\boldsymbol{S}^{*})^{\mathrm{T}}(\boldsymbol{I}-\boldsymbol{S}^{*})\boldsymbol{Y}^{\mathrm{T}})=tr(\boldsymbol{YEY}^{\mathrm{T}}) \tag{3}$$

其中，$\boldsymbol{E}=(\boldsymbol{I}-\boldsymbol{S}^{*})^{\mathrm{T}}(\boldsymbol{I}-\boldsymbol{S}^{*})$；$\boldsymbol{Y}=[\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_n]$ 是变换后样本。

3 本文算法

3.1 图像的稀疏表示

稀疏表示的概念源于视神经网络的研究，是对只有一小部分神经元同时处于活跃状态的多维数据的神经网络的表示方法[5]。由于稀疏表示的良好性能，很快被应用于人脸识别研究。文献[6]提出的稀疏保留映射算法(Sparisty Preserving Projects, SPP)能取得较好的实验效果。对训练样本 $\boldsymbol{X}$，通过解决下面改动的 L1 范数问题，对每个样本 $\boldsymbol{x}_i$ 求出对应的稀疏重构权重向量 $\boldsymbol{s}_i$：

$$\min_{\boldsymbol{s}_i}\|\boldsymbol{s}_i\|_1$$
$$\text{s.t. } \boldsymbol{x}_i=\boldsymbol{X}\boldsymbol{s}_i \tag{4}$$

3.2 算法过程

假设矩阵 $\boldsymbol{X}=[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n]$ 表示包含一组正约束以及一组负约束的 n 个训练样本数据集，那么降维的主要目的则是要找出一个投影向量 $\boldsymbol{W}=[w_1,w_2,\cdots,w_d]$，使得数据的低维表示 $\boldsymbol{Y}=\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}$，可以保存数据原有的稀疏结构以及成对约束信息。为此，定义目标函数如下式，通过最大化该目标函数来获得投影矩阵 $\boldsymbol{w}$：

$$J(\boldsymbol{w})=\frac{\sum_{i,j=1}^{n}\left\|\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_j\right\|^2\boldsymbol{C}_{ij}}{\sum_{i,j=1}^{n}\left\|\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_j\right\|^2\boldsymbol{M}_{ij}+\alpha J_{sp}(\boldsymbol{w},\boldsymbol{X})} \tag{5}$$

其中，矩阵 $\boldsymbol{C}$ 和 $\boldsymbol{M}$ 的定义如下：

$$\boldsymbol{C}_{ij}=\begin{cases}1 & \text{如果}\boldsymbol{x}_i\text{和}\boldsymbol{x}_j\text{属于负约束}\\ 0 & \text{其他}\end{cases} \tag{6}$$

$$\boldsymbol{M}_{ij}=\begin{cases}1 & \text{如果}\boldsymbol{x}_i\text{和}\boldsymbol{x}_j\text{属于正约束}\\ 0 & \text{其他}\end{cases} \tag{7}$$

其中，$J_{sp}(\boldsymbol{w},\boldsymbol{X})$ 是惩罚项，用于保留数据在变换空间中的稀疏重构结构；参数 $\alpha(\alpha\geqslant 0)$ 用于保持相似对距离与惩罚项之间的平衡。

$J_{sp}(\boldsymbol{w},\boldsymbol{X})$ 的定义如下：

$$J_{sp}(\boldsymbol{w},\boldsymbol{X})=\sum_{i=1}^{n}\left\|\boldsymbol{x}_i-\sum_{j=1}^{n}\boldsymbol{s}_{ij}\boldsymbol{x}_j\right\|^2=\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{S}'\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W} \tag{8}$$

其中，$\boldsymbol{S}'=\boldsymbol{I}-\boldsymbol{S}-\boldsymbol{S}^{\mathrm{T}}+\boldsymbol{S}^{\mathrm{T}}\boldsymbol{S}$；$\boldsymbol{S}=[\boldsymbol{s}_1,\boldsymbol{s}_2,\cdots,\boldsymbol{s}_n]$；$\boldsymbol{s}_i$ 是对样本 $\boldsymbol{x}_i$ 求出的对应的稀疏重构权重向量。

在式(5)中：

$$\sum_{i,j=1}^{n}\left\|\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_j\right\|^2\boldsymbol{C}_{ij}=\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{L}_C\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W} \tag{9}$$

且有 $\boldsymbol{L}_C=\boldsymbol{D}_C-\boldsymbol{C}$，$\boldsymbol{D}_C$ 是一个对角阵，其第 i 个对角元素为 $\sum_{j=1}^{n}\boldsymbol{C}_{ij}$。同样有：

$$\sum_{i,j=1}^{n}\left\|\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_j\right\|^2\boldsymbol{M}_{ij}=\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{L}_M\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W} \tag{10}$$

结合式(8)、式(9)和式(10)，目标函数式(5)可以重写为：

$$J(\boldsymbol{w})=\frac{\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{L}_C\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}}{\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{L}_M\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}+\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{S}'\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}}=$$
$$\frac{\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{L}_C\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}}{\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}(\boldsymbol{L}_M+\boldsymbol{S}')\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}} \tag{11}$$

4 实验结果与分析

4.1 Feret 库中不同训练样本个数的识别率比较

本节实验采用 Feret 人脸库，包含 200 个人，每个人分别有 7 幅图像，每幅图像的分辨率为 80×80 像素。在实验中，随机抽取每类人脸的 m(m=3, 4)个样本组成训练样本集，将剩余的数据作为测试数据。在固定训练样本数目的情况下，每次实验均随机产生 10 组不同的训练样本集，并统一采用最近邻分类器进行分类识别，实验结果取 10 次实验的平均值。另外，在实验过程中随机选取 50%的训练样本作为成对约束信息，调节参数 α 选为 0.2。选取 10 个~100 个投影轴进行特征抽取，最后比较本文算法和其余算法的分类识别效果。限于篇幅，选取部分数据罗列。图 1 描述 4 个训练样本下 4 种算法在 Feret 人脸库上的识别率比较。

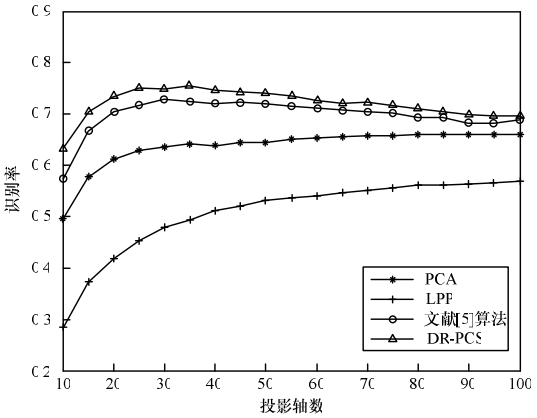

图 1 Feret 库上 4 种算法的识别率比较

由图 1 可知，当训练样本数选为 4 时，DR-PCS 在最近邻分类器下的识别性能均优于 PCA(Principal Component Analysis)、LPP(Locality Preserving Projections)以及文献[5]算法。当投影轴数达到 20~40 的阶段时识别率达到最大值且逐渐趋于稳定。

4.2 Feret 库上不同约束对的识别率比较

本节实验仍然采用 Feret 人脸库，随机选取每类人脸的第 1 幅、第 2 幅、第 5 幅、第 7 幅图像作为训练样本，并将剩余的图像作为测试样本，通过改变成对约束对的个数来进行识别率比较。在实验中，将选取 100 对~400 对成对约束对进行识别率比较，调节参数 α 仍选为 0.2。表 1 描述不同成对约束数目下 DR-PCS 算法在 Feret 人脸库上的最佳识别率及对应的投影轴。

表 1 不同约束对的识别率比较

约束数/对	最佳识别率/(%)	投影轴数
100	70.67	65
200	77.33	70
300	83.83	25
400	86.17	35

4.3 AR 库的最佳识别率比较

本节实验采用 AR 人脸库，该库包含 120 个人，每个人 26 幅图像，这些人脸图像是在不同时期、光照、姿势、表情、遮挡等条件下拍摄的。每幅图像的分辨率为 50×40 像素。在本实验中，仅考察无遮挡的人脸图像。取每个人的第 14 幅~第 20 幅图像作为训练样本，第 1 幅~第 7 幅图像作为测试样本。另外选取 50%的训练样本作为成对约束对，参数 α 选为 0.2。然后选取 10 个~100 个投影轴进行特征抽取，最后比较

(下转第 197 页)