文章编号:1001 - 2486(2005)02 - 0061 - 04

一种基于混合概率 PCA 模型的高光谱图像非监督分类方法*

吴 昊,郁文贤,匡纲要
(国防科技大学 电子科学与工程学院,湖南 长沙 410073)

摘 要:提出了一种在期望最大化(EM)算法框架下同时实现混合概率主成分分析(PPCA)降维和聚类的高光谱图像非监督分类方法。它根据不同类别应各有自己代表性的特征集,将通常意义下的特征抽取和模式分类合并在一步内完成,尽可能地保留了可分性;同时该方法具有概率模型的优点,更适合高维数据处理。采用仿真数据和真实数据进行的比较实验表明,该算法较一般不加区分地对所有原始数据进行 PCA 降维再分类的方法能得到更好的分类结果。

关键词:非监督分类;降维;混合概率主成分分析;期望最大化算法
中图分类号:TN958 **文献标识码**:A

An Unsupervised Hyperspectral Image Classification Method Based on the Mixture of Probabilistic PCA Modeling

WU Hao, YU Wen-xian, KUANG Gang-yao
(College of Electronic Science and Engineering, National Univ. of Defense Technology, Changsha 410073, China)

Abstract: An unsupervised hyperspectral image classification method simultaneously realizing the mixture of probabilistic PCA and clustering under the frame of EM algorithm is proposed. It is based on the fact that different class should have its own representative feature set, and it realizes feature extraction and classification in one step while preserving as much separability. It also possesses the advantages of PPCA model, which is more effective to high dimensional data processing. Applying the method to simulated data and real data shows that it can achieve better results compared with the method that applies PCA to all data without differentiation among classes.

Key words: unsupervised classification; dimensionality reduction; mixture of Probabilistic Principal Component Analysis (PPCA); EM (Expectation Maximization) algorithm

高光谱图像非监督分类技术可以揭示数据的固有结构,为监督分类、检测等处理提供必要的信息。但地物的复杂多变和高光谱数据的高维特性给高光谱分类技术提出了许多挑战。通常的方法是先对数据进行降维,再进行分类,而最常用的降维方法之一是 PCA 变换。这种方法比较直观,然而由于数据的复杂多变,并不一定满足 PCA 变换假定的椭球状分布,这样一个全局线性的 PCA 变换可能会丢失很多信息,给分类带来错误;而且传统 PCA 的定义由于缺乏一个相关的概率密度或生成模型而具有一定的局限性,尤其是在对高维数据或从大量的数据点寻找主成分方向时,在计算复杂度和数据缺失方面都会出现问题。文献[2]提出的一种概率 PCA(PPCA),具有较多优点,而且能进一步扩展至混合模式,扩展了 PCA 的应用范围。

本文将混合 PPCA 方法用于高光谱数据,并用 EM 算法[4]学习得到模型参数,从而对数据进行分类。算法在最大似然的框架下,基于一个特定形式的高斯隐变量模型构建概率 PCA,并形成混合模型,根据该模型分别对不同类的数据进行降维,同时实现了分类。由于保留了更多的可分性,算法可得到更准确的结果,尤其适用于地物光谱较相近的复杂场景。算法可以与其他的密度估计技术进行比较,并可以应用 Bayes 推导方法进行模型选择;在数据缺失时,算法可以在每次迭代中直接估计缺失信息的最大似然值,同样具有一般 EM 算法所有的优点。

* 收稿日期:2004 - 11 - 13
作者简介:吴昊(1976—),女,博士生。

1 混合概率 PCA

对于 d 维观测数据集 $\{t_n, n=1,2,\cdots,N\}$ 中的单个样本矢量 t，传统 PCA 是通过 $x = W^{\mathrm{T}}(t-\bar{t})$ 得到降维表示，x 即为 PCA 变换后的结果。其中，W 为变换矩阵，$\bar{t}$ 为样本均值，设 S 为样本协方差矩阵，则 W 由 S 的本征向量组成，即 $W=(w_1,w_2,\cdots,w_q)$，其中，$Sw_j=\lambda_j w_j$，$\lambda_j(j=1,\cdots,q)$ 为 S 的本征值，q 为降维后的维数，即主子空间维数。

隐变量模型与 PCA 之间有紧密的联系[2]。隐变量模型给出了观测数据 t 和隐变量 x 之间的关系，其中最常用的是因子分析，它描述的是线性关系：

$$t = Wx + \mu + \varepsilon \tag{1}$$

这里，W 为因子载荷，μ 为模型均值，ε 为误差项。通常定义 x 和 ε 服从高斯分布，即 $x \sim N(0,I)$ 以及 $\varepsilon \sim N(0,\Psi)$，且 Ψ 为对角阵，则 $t \sim N(\mu, C)$，模型协方差 $C = WW^{\mathrm{T}} + \Psi$。

1.1 概率 PCA 模型

PCA 可视为因子分析的特殊情形，当误差为各向同性，即 $\varepsilon \sim N(0,\sigma^2 I)$ 时，由(1)式可以建立一个由隐变量空间至观测数据主子空间的映射，将因子分析与 PCA 联系起来[2]：

$$P(t \mid x) = (2\pi\sigma^2)^{-d/2}\exp\left(-\frac{1}{2\sigma^2}\| t - Wx - \mu \|^2\right) \tag{2}$$

可以推导出隐变量 x 关于观测变量 t 的后验概率密度分布：

$$p(x \mid t) = (2\pi)^{-q/2}|\sigma^{-2}M|^{1/2}\exp\left\{-\frac{1}{2}[x - M^{-1}W^{\mathrm{T}}(t-\mu)]^{\mathrm{T}}(\sigma^{-2}M)[x - M^{-1}W^{\mathrm{T}}(t-\mu)]\right\} \tag{3}$$

其中，$M = W^{\mathrm{T}}W + \sigma^2 I$，维数为 $q \times q$，而 $C = WW^{\mathrm{T}} + \sigma^2 I$，维数为 $d \times d$。

由此得到单一的概率 PCA(PPCA)模型，在该模型下观测数据的对数似然函数为：

$$L(t) = \sum_{n=1}^{N}\ln\{p(t_n)\} = -\frac{N}{2}\{d\ln(2\pi) + \ln|C| + \mathrm{tr}(C^{-1}S)\} \tag{4}$$

各参数的最大似然估计为：$\mu_{ML} = \frac{1}{N}\sum_{n=1}^{N} t_n$，$S_{ML} = \frac{1}{N}\sum_{n=1}^{N}(t_n-\mu)(t_n-\mu)^{\mathrm{T}}$；将(4)式最大化可得：$W_{ML} = U_q(\Lambda_q - \sigma^2 I)^{1/2}R$，其中，$d \times q$ 阶矩阵 U_q 的列矢量为 S 的本征向量，对应 $q \times q$ 阶对角矩阵 Λ_q 中的特征值，R 为任意 $q \times q$ 阶正交旋转矩阵，实际应用中可以简化取 $R = I$；$\sigma^2_{ML} = \frac{1}{d-q}\sum_{j=q+1}^{d}\lambda_j$。由此用最大似然参数估计代替了通常的样本协方差矩阵特征值分解。另外，W 和 σ^2 除了用以上公式显示计算，还可以用 EM 迭代算法高效地求解。

PPCA 克服了传统 PCA 由于缺乏相关的概率密度或生成模型而带来的局限性。由概率模型得到的似然值使得该方法可与其它概率技术进行比较，便于进行统计测试并允许应用 Bayes 方法；除了降维，PPCA 也可用作通用的高斯密度模型，潜在的应用包括分类和异常检测。更重要的是，它可以扩展至混合形式，即混合 PPCA 模型，以适应数据分布更复杂的情形，有效克服了 PCA 作为线性投影的局限性；而且混合 PPCA 优于混合 PCA 对非线性结构建模的一般方法，即划分数据空间后对每个划分进行主子空间估计的方法，由于传统 PCA 定义的模糊性而造成了这类方法的定义形式不惟一；PPCA 不仅能得到定义清晰的惟一的混合算法，而且还具有概率密度函数带来的所有优点。

1.2 混合 PPCA 模型

将以上的 PPCA 扩展到混合形式，则观测数据集的混合模型对数似然函数为：

$$L = \sum_{n=1}^{N}\ln\{p(t_n)\} = \sum_{n=1}^{N}\ln\left\{\sum_{i=1}^{M}\pi_i p(t_n \mid i)\right\} \tag{5}$$

其中，$p(t \mid i)$ 是前述的单一 PPCA 模型，M 为混合成分数，π_i 为混合比例，$\pi_i \geqslant 0$ 且 $\sum\pi_i = 1$，每个混合成分的参数为 μ_i，W_i 和 σ_i^2。可由 E-step 和 M-step 推导出用 EM 算法求解混合 PPCA 的迭代公式[3]：

$$\begin{cases}\tilde{\pi}_i = \dfrac{1}{N}\sum\limits_n \boldsymbol{R}_{ni} \\ \tilde{\boldsymbol{\mu}}_i = \dfrac{\sum\limits_n \boldsymbol{R}_{ni}(\boldsymbol{t}_{ni} - \tilde{\boldsymbol{W}}_i < \boldsymbol{x}_{ni} >)}{\sum\limits_n \boldsymbol{R}_{ni}} \\ \tilde{\boldsymbol{W}}_i = \left[\sum\limits_n \boldsymbol{R}_{ni}(\boldsymbol{t}_{ni} - \tilde{\boldsymbol{\mu}}_i) < \boldsymbol{x}_{ni} >^{\mathrm{T}}\right]\left[\sum\limits_n \boldsymbol{R}_{ni} < \boldsymbol{x}_{ni}\boldsymbol{x}_{ni}^{\mathrm{T}} >\right]^{-1} \\ \tilde{\sigma}_i^2 = \dfrac{1}{d\sum\limits_n \boldsymbol{R}_{ni}}\left\{\sum\limits_n \boldsymbol{R}_{ni} \| \boldsymbol{t}_{ni} - \tilde{\boldsymbol{\mu}}_i \|^2 - 2\sum\limits_n \boldsymbol{R}_{ni} < \boldsymbol{x}_{ni} >^{\mathrm{T}} \tilde{\boldsymbol{W}}_i^{\mathrm{T}}(\boldsymbol{t}_n - \tilde{\boldsymbol{\mu}}_i) + \sum\limits_n \boldsymbol{R}_{ni}\mathrm{tr}(< \boldsymbol{x}_{ni}\boldsymbol{x}_{ni}^{\mathrm{T}} > \tilde{\boldsymbol{W}}_i^{\mathrm{T}}\tilde{\boldsymbol{W}}_i)\right\}\end{cases} \tag{6}$$

其中,$\boldsymbol{R}_{ni} = \dfrac{p(\boldsymbol{t}_n \mid i)\pi_i}{p(\boldsymbol{t}_n)}$,表示第 i 个混合成分生成第 n 个数据点 $\boldsymbol{t}_n$ 的后验概率。

2　算　法

在已知类别数 K 的条件下提出了如下算法:

(1)根据信息量门限用 Power method 计算出待保留主成分的数目[5],一般取信息量门限为 98%,即可满足保留信息分量、消除噪声分量的要求;

(2)用 K – 均值算法对数据进行初始分类,计算混合 PPCA 模型初始参数;

(3)将初始参数代入(6)式的 EM 算法,进行迭代直到收敛;

(4)根据模型参数按照最大后验概率(MAP)准则对数据进行分类。

当类别数 K 未知时,同样可以根据图像的复杂程度设定最大类别数 $K_{\max}$,从计算得到的多个模型中采用 Bayes 信息量准则(BIC)进行模型选择以确定合适的类别数[1]。

与应用全局 PCA 再进行分类的方法比较,本文的方法具有如下优点:(1)在数据为高维时通常没有足够的数据使样本协方差阵为满秩,应用全局 PCA 会存在困难,而 PPCA 并不直接计算协方差阵,应用 EM 算法可以在数据缺失的情况下进行降维和分类;(2)可以适应实际应用中数据分布更复杂的情况,有效克服了 PCA 作为线性投影的局限性,保留了数据间更多的可分性,可得到更高的正确分类率;(3)利用得到的精确的降维模型可以方便地对新数据点进行降维和分类。

将本文算法与另外两种算法进行比较,一种是先用 PCA 进行光谱维降维,再用 K – 均值进行分类的方法;另一种是文献[1]中算法的简化算法,算法首先应用 PCA 线性投影对数据进行光谱维降维,然后对变换后数据用 EM 算法进行分类[1],将其简化以在相同条件下进行对比验证:指定待保留主成分的维数并用 PCA 对数据降维;用 K – 均值算法对降维后数据初始分类,计算高斯混合模型初始参数;用降维后的数据代入求解高斯混合的 EM 算法进行迭代,直到收敛;根据模型参数,按照 MAP 原则对降维数据进行分类。

3　实验结果

针对仿真数据和 PHI 图像,将本文算法与另外两种算法进行了比较。先将后两种算法应用于仿真数据,数据包含两类,每类含 1000 个样本,共 2000 个数据点,为直观起见,数据维数取为 2,降至 1 维,如图 1 所示。每类数据都服从二维高斯分布,均值分别为:$\boldsymbol{\mu}_1 = [0,0]$,$\boldsymbol{\mu}_2 = [8,0]$,协方差阵分别为 $\boldsymbol{C}_1 = [1,2]$,$\boldsymbol{C}_2 = [2,1]$,并将第一个分布沿均值顺时针旋转 45°,这样数据总体的主轴方向与每个分布各自的主轴方向都不一致,应用全局 PCA 投影至一维时会有部分重合,损失了一定的可分性;而本文的算法通过估计概率密度,有针对性地根据各类的分布对数据进行投影,保留了更多的可分性。

将三种算法针对 PHI 高光谱图像数据进行了比较。PHI 是推帚式 CCD 面阵成像光谱仪,光谱范围为可见光到近红外波段(波长:0.40 ~ 0.95μm),可选波段为 244 个。图像尺寸为 330 × 311,共选取了 80 个波段,数据进行了定标处理,为 DN 值。图 2 是将第 56、71 和 67 波段图像分别作为红、绿、蓝三波段的

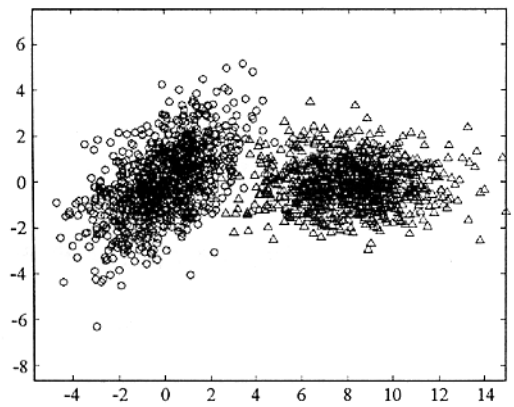

图1 仿真数据
Fig.1 Simulated data

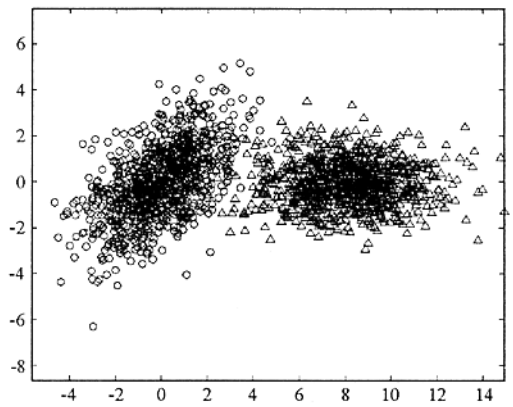

图2 PHI真实数据
Fig.2 PHI real data

合成图像。图中1、2、3为水域,4、5为草地,6、7为机场跑道。取图中所示的不同类型的地物分别进行了三次实验,类别数均为3,以比较本文算法与先进行PCA降维再做分类的算法的分类效果。为便于比较,三种方法保留的主成分维数都取为4。

表1给出了各次实验的分类错误率和错误的点数,其中"混合PPCA"代表本文的算法,"PCA + K – Mean"代表先进行PCA再进行K – 均值分类的方法,"PCA + EM"代表文献[1]中算法的简化算法。显然,无论对仿真数据还是真实高光谱数据,本文算法的分类错误率均低于应用全局PCA降维再进行分类的方法。

表1 分类算法分类错误率比较

Tab.1 Classification error rate comparison of the two methods

实验数据	混合PPCA	PCA + K – Mean	PCA + EM
仿真数据(2类,2000点)	0.70%(14)	1.70%(34)	1.25%(25)
PHI – 1(区域1、4、6,5700点)	0.19%(11)	0.32%(18)	0.30%(17)
PHI – 2(区域2、4、7,5351点)	0.11%(6)	0.22%(12)	0.21%(11)
PHI – 3(区域3、5、7,4604点)	0.09%(4)	0.22%(10)	0.15%(7)

4 结论

将混合概率主成分分析模型用于高光谱图像的非监督分类,并采用期望最大化(EM)算法求解模型参数,同时实现了降维和聚类,根据估计出的数据分布尽可能多地保留了数据可分性,分类结果较精确。该方法还可以输出数据的协方差模型,与协方差阵相关的参数的最大似然估计可以高效地由数据主成分模型计算得到,具有更多潜在的应用。

致谢 对提供数据的中科院上海技术物理研究所表示感谢。

参考文献:

[1] Wu H, et al. An Unsupervised Classification Method for Hyperspectral Image Combining PCA and Gaussian Mixture Model[A]. Proc. of the 3rd Inter. Symposium on MIPPR[C], SPIE, Beijing, 2003:729 – 734.

[2] Tipping M E, Bishop C M. Probabilistic Principal Component Analysis[J]. Journal of the Royal Statistical Society, 1999, B, 61(3):611 – 622.

[3] Tipping M E, Bishop C M. Mixtures of Probabilistic Principal Component Analysis[J]. Neural Computation, 1999(11):443 – 482.

[4] Dempster A P, Laird N M, Rubin D B. Maximum-likelihood from Incomplete Data via the EM Algorithm [J]. J. Royal Stat. Soc. Ser. B, 1977, 39:1 – 38.

[5] Kaewpijit S, Moigne J L, El-Ghazawi T. Finding the Dimensionality of Hyperspectral Data[A]. SPIE[C], 2001, 4381:339 – 347.