

基于非负稀疏图的高光谱数据降维

高 阳 王雪松* 程玉虎

(中国矿业大学信息与电气工程学院 徐州 221116)

摘要: 为减少因大量的光谱信息带来的计算复杂及数据冗余带来的高光谱数据分类性能降低, 该文提出一种非负稀疏图降维算法。首先, 构建超完备块字典对高维高光谱数据进行非负稀疏表示。然后, 根据块非负稀疏表示, 分别构建内部非负稀疏图和惩罚非负稀疏图, 基于单调递减函数定义边的权重以体现样本间的相似程度。最后, 通过同时最大化异类和最小化同类非负稀疏重构样本间的距离, 得到从高维到低维的最优映射关系, 从而实现对高维高光谱数据的降维。AVIRIS 92AV3C 高光谱数据上的实验结果表明, 所提算法能以较少的训练样本获得较高的整体分类精度和 Kappa 系数。

关键词: 高光谱; 降维; 非负稀疏图; 整体分类精度; Kappa 系数

中图分类号: TP751.1

文献标识码: A

文章编号: 1009-5896(2013)05-1177-08

DOI: 10.3724/SP.J.1146.2012.01082

Dimensionality Reduction of Hyperspectral Data Using Non-negative Sparsity Graph

Gao Yang Wang Xue-song Cheng Yu-hu

(School of Information and Electrical Engineering, China University of Mining and Technology, Xuzhou 221116, China)

Abstract: In order to reduce the computation complexity resulted from large number of spectral information and to reduce the decline of classification performance resulted from data redundancy, a dimensionality reduction algorithm called non-negative sparse graph is proposed. At first, an over-complete block dictionary is constructed to realize the non-negative sparse representation of high-dimensional hyperspectral data. Then, according to the non-negative sparse representation, an inner non-negative sparsity graph and a penalty non-negative sparsity graph are built where the weights of edges are defined by a monotone decreasing function to embody the similarity degree of samples. At last, an optimal mapping from the high-dimensional space to a low-dimensional subspace can be obtained by simultaneously maximizing the distance between non-negative sparsity reconstruction samples of different classes and minimizing the distance between non-negative sparsity reconstruction samples of the same class, which makes the dimensionality reduction of high-dimensional hyperspectral data realized. Experimental results on AVIRIS 92AV3C hyperspectral data show that the proposed algorithm can obtain higher overall accuracy and Kappa coefficient with few training samples.

Key words: Hyperspectral; Dimensionality reduction; Non-negative sparsity graph; Overall accuracy; Kappa coefficient

1 引言

随着高光谱传感器的发展, 高光谱图像获得广泛的应用, 每个像素点能够同时获取数百个密集且连续的光谱波段^[1]。每个像素点的丰富的信息增加了区分材料和地物的能力, 从而可以从新的方向处理图像分类问题。然而, 大量的光谱信息不仅增加了计算的复杂度, 而且可能会降低分类器的精度^[2], 同

时也增加了光谱波段间的重叠及 Hughes 现象^[3]。这些表明, 高光谱数据中的判别信息不是均匀分布在整个频谱, 其中有一些波段包含比其它波段更多的有用可分性信息。为了得到更高的分类性能指标, 通常选取更有价值信息的波段来分类。目前, 有两种不同的策略来解决这个问题: 直接选择有效光谱波段的波段选择以及将高维样本投影到低维子空间的波段提取^[4]。

目前, 已有大量基于图的降维方法被成功应用于机器学习和模式识别领域。其最大的特点是用图来表示数据之间的关系, 样本被看作高维空间中的点。经典流形学习方法有局部线性嵌入^[5]、等距特征

2012-08-24 收到, 2013-01-10 改回

国家自然科学基金(60974050, 61072094, 61273143), 教育部新世纪优秀人才支持计划(NCET-08-0836, NCET-10-0765)和教育部博士点基金(20110095110016, 20120095110025)资助课题

*通信作者: 王雪松 wangxuesongcumt@163.com

映射 Isomap^[6]、拉普拉斯特征映射^[7], 这些方法从不同角度利用图的思想。此外, 大量数据降维算法, 如独立主成分分析^[8,9]、线性判别分析 (Linear Discriminant Analysis, LDA)^[10]和局部保持投影 (Locality Preserving Projections, LPP)^[11]都可以归为图嵌入框架。现在存在两种主流的图构建方法: k 近邻和 ϵ 邻域, 主要为每个样本选择距离最近的 k 个样本点和距离小于 ϵ 的球面内样本, 一般通过二进制、高斯核^[7]、逆欧氏距离^[12]和局部线性重构^[5]进一步计算图中边对应的权重。但是, 这两种方法均面临近邻参数选择这一开放问题。因此, 构建一个合理、高效的图在基于图的降维学习算法中至关重要。从机器学习的角度考虑, 图的构建需要满足 4 个要求^[13,14]: (1)高判别力; (2)稀疏性; (3)自适应近邻选择; (4)数据噪声的鲁棒性。

传统的 k 近邻和 ϵ 邻域方法并不能满足上述 4 个要求, 首先, 它们都是通过一个参数为样本选择近邻, 不能够自适应地为每个样本选择近邻; 其次, 它们都是利用欧氏距离构建近邻图, 而欧氏距离对噪声数据非常敏感, 即当不利噪声进来时, 图的结构很容易改变。为了能够自适应选择近邻, Yang 等人^[15]提出一种样本依赖的局部保持投影 (Sample-dependent Locality Preserving Projection, SLPP) 和 Raducanu 等人^[16]提出一种监督拉普拉斯特征 (Supervised Laplacian Eigenmaps, S-LE) 映射, 它们很好地避免了近邻参数的选择, 但是它们同 k 近邻和 ϵ 邻域方法一样, 对数据噪声的鲁棒性小, 且当样本数据量非常大及样本点密集时, 如高光谱数据, 自适应选择近邻时会选到大量的样本作为近邻, 而这些样本中存在相似度高的冗余样本, 不利于对样本的特征提取。稀疏表示 (Sparse Representation, SR)^[17,18]是近年来在信号处理、图像处理和模式识别等领域得到广泛应用的新方法。受 SR 的启发, Wright 等构建了 l^1 -graph^[14], 其中每个样本可以通过其它样本来稀疏表示且为每个样本自动选择信息量大的近邻样本。通过 l^1 最小化稀疏表示可以自然地满足稀疏性和自适应性, 而且通过 l^1 最小化可以描述样本间的线性关系, 这有助于增强基于图方法的学习性能。类似于 l^1 -graph, 文献[19]提出稀疏保持投影 (Sparsity Preserving Projections, SPP)。鉴于负的权重系数无法合理进行样本间的信息传递, Wong^[20]提出非负稀疏保持嵌入 (Non-negative Sparseness Preserving Embedding, NSPE)。这些方法满足了构图的 4 个要求并得到了广泛的应用, 然而它们需要对整体样本进行稀疏重构, 耗时多, 而且它们均为无监督学习方法, 没有利用样本的先验

类别知识。为了充分利用样本的类别信息, 在 SPP 基础上, 结合最大间隔准则 (Maximum Margin Criterion, MMC)^[21]方法, 出现了 DSNPE1^[22]和 DSNPE2^[23]两种监督型稀疏近邻保持嵌入方法, 并获得较好的效果。但是, 它们仍存在一些不足: DSNPE1 仅考虑了局部的类内稀疏重构, 却忽略了其它类的稀疏重构, 无法找到全局最优解; DSNPE2 同时考虑了类内稀疏重构和类间稀疏重构, 然而需要进行两次稀疏重构, 增加了时间复杂度。这些基于稀疏表示的方法, 在降维时能够保持样本局部稀疏几何结构不变, 即投影样本保持相同的重构权重, 但是无法保持样本近邻关系不变。除了 NSPE, 其它几种算法在建图过程中还可能会出现负权重系数。

为解决上述问题, 本文提出一种新的监督降维方法——非负稀疏图 (Non-negative Sparsity Graph, NSG), 并将其应用于高光谱数据降维问题。NSG 的主要贡献是: (1)构建的超完备字典能够减少字典中的原子数量, 进而提高计算效率。(2)定义的基于同类非负稀疏重构样本的内部非负稀疏图和基于异类非负稀疏重构样本的惩罚非负稀疏图, 能够使 NSG 包含判别信息。(3)通过单调递减函数自适应表示内部非负稀疏图和惩罚非负稀疏图中稀疏重构样本的重要程度。(4)最大间隔准则使得同类非负稀疏重构样本在输出子空间靠近, 同时使异类非负稀疏重构样本在输出子空间远离, 从而可以改善高光谱数据的分类效果。

2 基于非负稀疏图的高光谱数据降维

2.1 问题描述

给定一个高维高光谱样本数据集 $O_m = \{(\mathbf{x}_1, \mathbf{y}_1), \dots, (\mathbf{x}_i, \mathbf{y}_i), \dots, (\mathbf{x}_n, \mathbf{y}_n)\} \in \mathbf{X} \times \mathbf{Y}$, 其中 $i = 1, 2, \dots, n$, $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_n] \in R^{m \times n}$ 是输入样本的数据矩阵, n 是样本的个数, m 是样本的维数, $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_i, \dots, \mathbf{y}_n] \in R^n$ 是输入样本的类别标签集, 类别个数为 c 。目标是找到一个低维子空间表示高维样本数据, 即

$$\mathbf{z} = \mathbf{W}^T \mathbf{x} \in R^d \quad (1)$$

其中 \mathbf{z} 为高维样本数据 \mathbf{x} 的低维表示, $d \ll m$, 投影矩阵 $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_d] \in R^{m \times d}$, $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_d$ 是特征向量。

2.2 块非负稀疏表示

给定一个高维样本数据 \mathbf{x}_i , 超完备字典 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in R^{m \times n}$, 且 $m < n$, \mathbf{x}_i 是超完备字典 \mathbf{X} 中的原子, 非负稀疏表示的目的是用尽可能少的超完备字典 \mathbf{X} 中的列元素(原子)表示 \mathbf{x}_i :

$$\left. \begin{array}{l} \min_{\mathbf{h}_i} \|\mathbf{h}_i\|_0 \\ \text{s.t. } \mathbf{x}_i = \mathbf{X}\mathbf{h}_i; \mathbf{h}_i \geq 0 \end{array} \right\} \quad (2)$$

其中 $\|\mathbf{h}_i\|_0$ 表示 \mathbf{h}_i 的 l_0 范数, 其值为 \mathbf{h}_i 中非零元素的个数, $\mathbf{h}_i = [h_{i1}, \dots, h_{ii-1}, 0, h_{ii+1}, \dots, h_{in}]^T \in R^n$, h_{ij} 表示第 j 个样本 \mathbf{x}_j 对重构样本 \mathbf{x}_i 所做出的贡献。因为将 \mathbf{x}_i 从 \mathbf{X} 中去除, 所以 \mathbf{h}_i 中第 i 个元素为 0, 即 $\mathbf{x}_i = h_{i1}\mathbf{x}_1 + \dots + h_{ii-1}\mathbf{x}_{i-1} + h_{ii+1}\mathbf{x}_{i+1} + \dots + h_{in}\mathbf{x}_n$ 。

基于超完备字典的非负稀疏表示及稀疏表示方法计算复杂, 重构精度低。为解决这些问题, 受文献[24]启发, 将超完备字典 \mathbf{X} 均匀划分成 v 块, 给出一种新的用于重构样本的块非负稀疏表示方法。

$$\mathbf{X} = \left[\begin{array}{c} \mathbf{x}_1^1, \dots, \mathbf{x}_b^1, \dots, \mathbf{x}_B^1, \dots, \mathbf{x}_1^v, \dots, \mathbf{x}_b^v, \dots, \mathbf{x}_B^v, \dots, \\ \mathbf{x}_1^V, \dots, \mathbf{x}_b^V, \dots, \mathbf{x}_B^V \end{array} \right] \quad (3)$$

其中 $n = VB$, \mathbf{X}^v 为一个超完备块字典, $\mathbf{x}_b^v = \mathbf{x}_{(v-1)B+b}$, 则第 v 子块中高维样本 \mathbf{x}_b^v 可用超完备块字典 \mathbf{X}^v 表示为

$$\left. \begin{array}{l} \min_{\mathbf{h}_b^v} \|\mathbf{h}_b^v\|_0 \\ \text{s.t. } \mathbf{x}_b^v = \mathbf{X}^v \mathbf{h}_b^v; \mathbf{h}_b^v \geq 0 \end{array} \right\} \quad (4)$$

其中 $\mathbf{h}_b^v \in R^B$ 。不幸的是, 式(4)是一个 NP 难的非凸组合优化问题。但是, 在解足够稀疏的条件下, 可采用求解 l_1 范数的方法代替求解 l_0 范数^[25]。 l_1 最小化问题可以通过 LASSO^[26]或者 Elastic Net^[27]来求解。

$$\left. \begin{array}{l} \min_{\mathbf{h}_b^v} \|\mathbf{h}_b^v\|_1 \\ \text{s.t. } \mathbf{x}_b^v = \mathbf{X}^v \mathbf{h}_b^v; \mathbf{1} = \mathbf{1}^T \mathbf{h}_b^v; \mathbf{h}_b^v \geq 0 \end{array} \right\} \quad (5)$$

其中 $\mathbf{1} \in R^B$ 是全 1 向量。利用非负最小二乘法求解式(5), 得重构误差^[28]为

$$\left. \begin{array}{l} \min_{\mathbf{h}_b^v} E(\mathbf{H}^v) = \min_{\mathbf{h}_b^v} \|\mathbf{X}^v - \mathbf{X}^v \mathbf{H}^v\|_2^2 + \gamma \|\mathbf{h}_b^v\|_1 \\ \text{s.t. } \mathbf{h}_b^v \geq 0; \mathbf{1} = \mathbf{1}^T \mathbf{h}_b^v \end{array} \right\} \quad (6)$$

其中 γ 是平衡稀疏和重构误差的非负系数。

根据式(6), 计算第 v 子块中每一个高维样本 \mathbf{x}_b^v 最优的非负稀疏重构权重向量 $\tilde{\mathbf{h}}_b^v \in R^B$, 则第 v 子块的块非负稀疏重构权重矩阵 $\mathbf{H}^v = [\tilde{\mathbf{h}}_b^v]_{B \times B}$ 。高维样本的整体非负稀疏重构权重矩阵 \mathbf{H} 表示为

$$\mathbf{H} = \text{diag}[\mathbf{H}^1, \dots, \mathbf{H}^v, \dots, \mathbf{H}^V] = [\tilde{\mathbf{h}}_i]_{n \times n} \quad (7)$$

2.3 非负稀疏图构建

将原始样本空间投影到低维子空间时, 需同时保持样本的稀疏几何结构和样本间的近邻关系不变, 即应同时缩减同类稀疏重构样本间的距离和扩

大异类稀疏重构样本间的距离, 以避免在输出子空间出现大量稀疏样本重叠。为此, 根据整体非负稀疏重构权重矩阵 \mathbf{H} , 我们尝试去构建一个内部非负稀疏图 $\mathbf{G}^q = \{\mathbf{X}, \Psi^q\}$ 和一个惩罚非负稀疏图 $\mathbf{G}^p = \{\mathbf{X}, \Psi^p\}$, 其中 $\Psi^q \in R^{n \times n}$ 和 $\Psi^p \in R^{n \times n}$ 是邻接矩阵。为了分类, 根据类别信息, 定义两个邻接矩阵中第 i 个样本和第 j 个样本间的权值分别为 Ψ_{ij}^q 和 Ψ_{ij}^p 。

$$\Psi_{ij}^q = \begin{cases} \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/t), & y_i = y_j, \tilde{h}_{ij} > 0 \\ 0, & \text{其它} \end{cases} \quad (8)$$

$$\Psi_{ij}^p = \begin{cases} \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/t), & y_i \neq y_j, \tilde{h}_{ij} > 0 \\ 0, & \text{其它} \end{cases} \quad (9)$$

其中, $\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/t)$ 是严格单调递减函数, 表示样本 \mathbf{x}_i 和 \mathbf{x}_j 间的距离。 Ψ_{ij}^q 表示在内部非负稀疏图中同类样本 \mathbf{x}_i 和 \mathbf{x}_j 之间的相似程度, 距离 $\|\mathbf{x}_i - \mathbf{x}_j\|^2$ 越小, 样本间的相似程度越大, 其它样本间的相似程度为 0。类似地, Ψ_{ij}^p 表示在惩罚非负稀疏图中异类样本 \mathbf{x}_i 和 \mathbf{x}_j 之间的相似程度, 距离 $\|\mathbf{x}_i - \mathbf{x}_j\|^2$ 越小, 样本间的相似程度越大, 其它样本间的相似程度为 0。

现在, 我们考虑将内部非负稀疏图和惩罚非负稀疏图投影到低维子空间时, 使 \mathbf{G}^q 中连接的点尽可能靠近在一起, 而 \mathbf{G}^p 中连接的点尽可能远离。设 $\mathbf{Z} = [z_1, z_2, \dots, z_n] \in R^{d \times n}$ 是这个低维子空间, 为了获得好的投影图, 需优化下述两个目标函数:

$$\min \sum_{i=1}^n \sum_{j=1}^n \|\mathbf{z}_i - \mathbf{z}_j\|^2 \Psi_{ij}^q \quad (10)$$

$$\max \sum_{i=1}^n \sum_{j=1}^n \|\mathbf{z}_i - \mathbf{z}_j\|^2 \Psi_{ij}^p \quad (11)$$

式(10)最小化内部非负稀疏图, 是为了保证如果 \mathbf{x}_i 和 \mathbf{x}_j 是同类且相近的, 那么 z_i 和 z_j 也会相近。式(11)最大化惩罚非负稀疏图, 是为了确保如果 \mathbf{x}_i 和 \mathbf{x}_j 相近但类别相异, 那么 z_i 和 z_j 将会远远分开。此学习过程如图 1 所示。

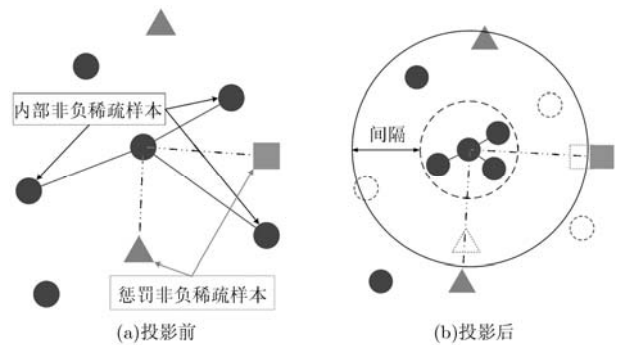


图1 内部非负稀疏图和惩罚非负稀疏图的投影过程

图1中,与中心点实线相连且同形状的3个点属于内部非负稀疏图中连接的同类点,在投影过程中,3个同类点向中心点靠近。与中心点点划线相连且不同形状的2个点属于惩罚非负稀疏图中连接的异类点,在投影过程中,2个异类点远离中心点。有2个点与中心点同类,但没有连接,虽然它们与中心点相距较近,应该在投影过程中向中心点靠近,但它们在重构中心点时没有做出贡献,故它们在投影过程中保持位置不变,相当于给内部非负稀疏图一个惩罚。同样,有1个与中心点没有连接的异类点,虽然它与中心点相距相对其它异类点较近,应该在投影过程中远离中心点,但它们在重构中心点时没有做出贡献,故它在投影过程中保持位置不变,相当于给惩罚非负稀疏图一个惩罚。这样使得同类非负稀疏重构样本在输出子空间靠近,而使异类非负稀疏重构样本在输出子空间远离,从而最大化它们之间的间隔。

2.4 目标函数

NSG的目标是同时最大化式(11)和最小化式(10)。因 $\mathbf{z} = \mathbf{W}^T \mathbf{x}$,则目标函数式(10)变成:

$$\begin{aligned} & \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \|z_i - z_j\|^2 \Psi_{ij}^q \\ &= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \|\mathbf{W}^T \mathbf{x}_i - \mathbf{W}^T \mathbf{x}_j\|^2 \Psi_{ij}^q \\ &= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \text{tr} \{ (\mathbf{W}^T \mathbf{x}_i - \mathbf{W}^T \mathbf{x}_j) \\ & \quad \cdot (\mathbf{W}^T \mathbf{x}_i - \mathbf{W}^T \mathbf{x}_j)^T \} \Psi_{ij}^q \end{aligned} \quad (12)$$

由于矩阵迹的运算是线性的且 Ψ_{ij}^q 是一个标量,式(12)可以化简为

$$\begin{aligned} & \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \text{tr} \{ \mathbf{W}^T (\mathbf{x}_i - \mathbf{x}_j) \Psi_{ij}^q (\mathbf{x}_i - \mathbf{x}_j) \mathbf{W}^T \} \\ &= \frac{1}{2} \text{tr} \{ \mathbf{W}^T (2\mathbf{X}\mathbf{D}^q \mathbf{X}^T - 2\mathbf{X}\Psi^q \mathbf{X}^T) \mathbf{W} \} \\ &= \text{tr} \{ \mathbf{W}^T \mathbf{X} (\mathbf{D}^q - \Psi^q) \mathbf{X}^T \mathbf{W} \} \\ &= \text{tr} \{ \mathbf{W}^T \mathbf{X} \mathbf{L}^q \mathbf{X}^T \mathbf{W} \} \end{aligned} \quad (13)$$

其中 $\mathbf{D}^q = \text{diag}(D_{11}^q, \dots, D_{mm}^q)$ 是对角矩阵,其对角线上的元素为邻接矩阵 Ψ^q 中所有对应行元素的和,即 $D_{ii}^q = \sum_{j=1}^n \Psi_{ij}^q$, $\mathbf{L}^q = \mathbf{D}^q - \Psi^q$ 为拉普拉斯矩阵。

类似于式(12)和式(13),式(11)可以化简为

$$\begin{aligned} & \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \|z_i - z_j\|^2 \Psi_{ij}^p = \text{tr} \{ \mathbf{W}^T \mathbf{X} (\mathbf{D}^p - \Psi^p) \mathbf{X}^T \mathbf{W} \} \\ &= \text{tr} \{ \mathbf{W}^T \mathbf{X} \mathbf{L}^p \mathbf{X}^T \mathbf{W} \} \end{aligned} \quad (14)$$

其中 $\mathbf{D}^p = \text{diag}(D_{11}^p, \dots, D_{mm}^p)$ 是对角矩阵,其对角线上的元素为邻接矩阵 Ψ^p 中所有对应行元素的和,即 $D_{ii}^p = \sum_{j=1}^n \Psi_{ij}^p$, $\mathbf{L}^p = \mathbf{D}^p - \Psi^p$ 为拉普拉斯矩阵。

利用MMC的优势,即最大化式(14)和最小化式(13),定义NSG的目标函数:

$$\begin{aligned} & \max_{\mathbf{W}} \text{tr} \{ \mathbf{W}^T \mathbf{X} \mathbf{L}^p \mathbf{X}^T \mathbf{W} \} - \text{tr} \{ \mathbf{W}^T \mathbf{X} \mathbf{L}^q \mathbf{X}^T \mathbf{W} \} \\ &= \max_{\mathbf{W}} \text{tr} \{ \mathbf{W}^T (\alpha \mathbf{X} \mathbf{L}^p \mathbf{X}^T - (1-\alpha) \mathbf{X} \mathbf{L}^q \mathbf{X}^T) \mathbf{W} \} \end{aligned} \quad (15)$$

其中 $\alpha \in [0, 1]$ 是平衡两个目标函数的常数。显然, $\mathbf{X} \mathbf{L}^q \mathbf{X}^T$ 和 $\mathbf{X} \mathbf{L}^p \mathbf{X}^T$ 是对称和半正定的。为了消除 \mathbf{W} 与一些非0标量相乘的任意性, \mathbf{W} 必须是由单位向量构成的,即 $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_k, \dots, \mathbf{w}_d]$ 和 $\mathbf{w}_k^T \mathbf{w}_k = 1$ 。则式(15)转化为求解下述约束优化问题:

$$\begin{aligned} & \max_{\mathbf{w}} \sum_{k=1}^d \mathbf{w}_k^T (\alpha \mathbf{X} \mathbf{L}^p \mathbf{X}^T - (1-\alpha) \mathbf{X} \mathbf{L}^q \mathbf{X}^T) \mathbf{w}_k \\ & \text{s.t.} \quad \mathbf{w}_k^T \mathbf{w}_k = 1, \quad k = 1, \dots, d \end{aligned} \quad (16)$$

利用拉格朗日乘子方法求解式(16),得

$$(\alpha \mathbf{X} \mathbf{L}^p \mathbf{X}^T - (1-\alpha) \mathbf{X} \mathbf{L}^q \mathbf{X}^T) \mathbf{w}_k = \lambda_k \mathbf{w}_k \quad (17)$$

其中, λ_k 是 $\alpha \mathbf{X} \mathbf{L}^p \mathbf{X}^T - (1-\alpha) \mathbf{X} \mathbf{L}^q \mathbf{X}^T$ 的第 k 个特征值, \mathbf{w}_k 是对应的正交特征向量。我们需要求出 d 个最优投影向量 $\mathbf{w}_1, \dots, \mathbf{w}_k, \dots, \mathbf{w}_d$,其对应的特征值为 $\lambda_1, \dots, \lambda_k, \dots, \lambda_d$ 且满足 $\lambda_1 \geq \dots \geq \lambda_k \geq \dots \geq \lambda_d$ 。DSNPE1, S-LE, NSPE, l -Graph, SLPP和LDA均需要求解广义特征向量,与此不同,NSG不需要进行矩阵的逆运算,可以成功地避免矩阵的奇异问题,也可以避免小样本问题。

3 算法步骤

非负稀疏图的高光谱数据降维算法如表1所示。

表1 非负稀疏图的高光谱数据降维算法

输入: 高维高光谱样本数据 $\mathbf{O}_m = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)\}$, 低维子空间维数 d , 常数 α, γ 。
输出: 低维高光谱样本数据 $\mathbf{O}_d = \{(z_1, y_1), (z_2, y_2), \dots, (z_n, y_n)\}$ 。
步骤1 根据式(3), 划分超完备字典 \mathbf{X} ;
步骤2 求解式(6), 得最优的块非负稀疏重构权重矩阵 \mathbf{H}^n ;
步骤3 根据式(7), 得整体非负稀疏重构权重矩阵 \mathbf{H} ;
步骤4 根据式(8)和式(9), 构建邻接矩阵 Ψ^q 和 Ψ^p ;
步骤5 求解对角阵 \mathbf{D}^q 和 \mathbf{D}^p , 拉普拉斯矩阵 \mathbf{L}^q 和 \mathbf{L}^p ;
步骤6 通过拉格朗日乘子方法求解式(16), 得最优投影矩阵 $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_k, \dots, \mathbf{w}_d]$;
步骤7 求高维样本 \mathbf{x} 的低维表示 $\mathbf{z} = \mathbf{W}^T \mathbf{x}$ 。

4 实验与分析

采用真实高光谱数据 AVIRIS 92AV3C 上的分类实验,来评估NSG的性能。AVIRIS 92AV3C是1992年由Airborne Visible/Infrared Imaging Spectrometer sensor获取的印第安纳州西北部农业

区影像，光谱范围 0.4~2.45 μm ，大小为 145 \times 145 像素，220 个波段，20 m 光谱分辨率，实验采集样本个数为 10194。该数据集主要用于农业研究，其中 Corn, Soybean 和 Crass 共 3 大类地物中包含不同量上季度作物残留物。图 2 为 92AV3C 高光谱数据的假彩色合成图及类别标记图。



图2 高光谱数据的假彩色合成图及类别标记图

实验过程中，选用 DSNPE1, DSNPE2, S-LE, SLPP, NSPE, SPP, MMC 和 LDA 作为对比算法。为比较公平性，统一采用支持向量机(Support Vector Machine, SVM)进行监督分类，SVM 的核函数为高斯核函数且核函数的宽度和惩罚因子用 5 重交叉验证得到，各算法中 $\alpha = 0.6$ ，稀疏类算法中 $\gamma = 1$ 。DSNPE1 和 DSNPE2 是按类别进行重构样本的，因此 NSG 中块字典个数 $v = c$ ，仿真中，每个实验均做 20 次，取平均值。首先，我们随机从各类地物中选取 10% 高光谱数据作为训练样本进行特征提取，并根据得到的 d 个特征向量将剩下的 90% 高光谱数据投影到低维子空间。分别采用 NSG, DSNPE1, DSNPE2, S-LE, SLPP, NSPE, SPP, MMC 和 LDA 共 9 种算法对 92AV3C 高光谱数据进行降维，然后利用 SVM 进行分类，图 3 和图 4 给

出了 92AV3C 高光谱数据上的整体分类精度和 Kappa 系数与低维子空间维数间的关系图。

由图可以看出，随着低维子空间维数的增加，各算法的整体分类精度和 Kappa 系数基本上呈稳步增长趋势直至趋于稳定。另外，在所有算法中，NSG 拥有最高的整体分类精度和 Kappa 系数，这是因为：NSG 利用了高光谱数据的先验类别信息，所以要比无监督的 SLPP, NSPE 和 SPP 的最高整体分类精度和 Kappa 系数高；不同于监督的 S-LE, MMC 和 LDA，监督的 NSG 充分利用高光谱数据的非负稀疏结构，可以有效地降低冗余样本的影响，从而获得较高的整体分类精度和 Kappa 系数；与只考虑同类稀疏结构的 DSNPE1 相比，NSG 同时考虑了同类和异类稀疏结构，能够找到高光谱数据降维的全局最优解；DSNPE2 同时最小化类内稀疏重构和最大化类间稀疏重构，投影到低维时能够保持重构关系不变且获得较好的全局最优解，但是它不能保持样本在投影过程中近邻关系不变，而 NSG 在投影过程中能够同时使同类稀疏重构样本相互靠近，异类稀疏重构样本相互远离，从而最大限度地分开不同类样本并获得更好的全局最优解。

然后，评价各种降维算法将高光谱数据统一降维到 30 维时的各项性能指标，如表 2 所示，其中包括生产者精度(Producer's Accuracy, PA(%))，使用者精度(User's Accuracy, UA(%))，平均精度(Average Accuracy, AA(%))，整体分类精度(Overall Accuracy, OA(%))，Kappa 系数和运算时间(Time(s))。由表 2 可知：一些地物的 PA 为 NaN，这表示没有其它类地物(包括同类地物)被分成该类。一些地物的 UA 为 0，这表示该类地物中样本都被误分成其它地物了。出现这种现象的主要是那些样本采集量较少的地物，如 Alfalfa(46 像素)，Grass/pasture-mowed(28 像素)和 Oats(20 像素)。

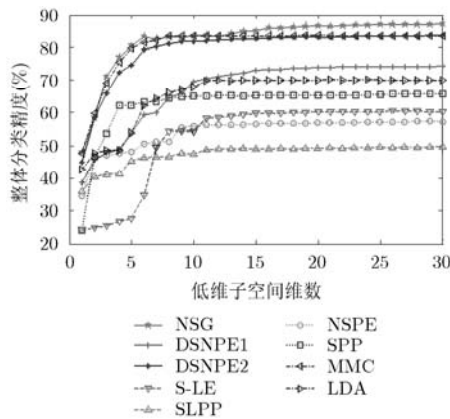


图3 整体分类精度与低维子空间维数关系

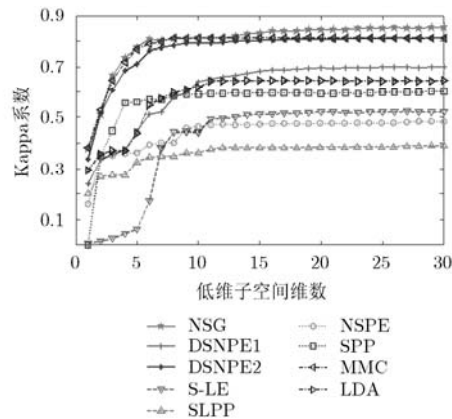


图4 Kappa系数与低维子空间维数关系

表 2 高光谱数据分类性能对比

	NSG		DSNFE1		DSNPE2		S-LE		SLPP		NSPE		SPP		MMC		LDA	
	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA
Corn																		
-min (751)	83.22	79.89	79.30	27.03	77.56	73.64	50.00	0.67	NaN	0	81.82	2.40	66.84	50.47	82.81	73.77	65.17	7.72
Corn- notill (1428)	81.67	82.70	66.59	76.75	81.74	76.82	67.77	37.25	50.98	18.14	64.15	55.39	57.80	44.61	82.72	72.06	65.47	75.14
Corn (237)	73.14	74.68	72.73	3.38	75.74	64.56	NaN	0	NaN	0	93.75	6.33	32.50	5.49	73.76	81.86	NaN	0
Alfalfa (46)	97.56	86.96	NaN	0	100.00	73.91	NaN	0	NaN	0	NaN	0	NaN	0	86.49	69.57	NaN	0
Grass/ Pasture (483)	96.53	92.13	81.97	79.09	94.31	89.23	92.75	74.12	NaN	0	100.00	0.21	66.67	22.36	91.55	91.93	92.99	82.40
Grass/ Trees (736)	92.77	97.69	82.99	98.78	93.60	97.42	84.24	96.60	47.72	74.05	61.81	95.65	74.13	98.51	92.42	96.06	86.55	97.96
Grass/ pasture- mowed (28)	89.29	89.29	NaN	0	85.71	85.71	NaN	0	NaN	0	NaN	0	NaN	0	85.71	85.71	NaN	0
Hay- win- drowed (479)	98.15	100.00	87.25	100.00	97.14	99.16	86.78	100.00	86.46	100.00	85.84	100.00	88.25	95.62	96.92	98.54	86.46	100.00
Oats(20)	87.50	70.00	NaN	0	50.00	40.00	NaN	0	NaN	0	NaN	0	NaN	0	46.15	60.00	NaN	0
Soybeans notill (977)	81.22	85.88	94.13	36.13	79.65	78.10	72.73	1.64	NaN	0	0	0	68.69	56.81	83.91	71.55	84.31	4.40
Soybeans- min (2455)	85.45	85.87	58.87	90.79	79.36	85.50	42.71	97.92	40.38	97.92	44.60	93.12	59.97	87.98	75.78	88.84	50.22	92.46
Soybeans- clean (593)	87.64	81.28	96.78	60.88	83.83	84.82	100.00	14.33	NaN	0	81.82	1.52	72.05	39.12	80.90	75.72	100.00	29.68
Wheat (205)	97.14	99.51	98.54	99.02	98.99	95.61	97.55	97.07	NaN	0	90.78	96.10	80.49	96.59	93.95	99.00	99.51	100.00
Woods (1265)	92.87	96.76	90.46	98.89	91.46	98.26	87.00	98.42	64.37	99.68	72.20	98.97	74.80	98.10	93.53	96.05	89.53	98.02
Bldg- Grass- Tree- Drive (386)	77.38	61.14	93.06	41.71	77.66	58.55	90.91	36.27	NaN	0	78.13	6.48	58.06	13.99	73.10	59.84	88.53	50.00
Stone- stell tower (105)	100.00	89.52	98.94	88.57	97.92	89.52	98.91	86.67	97.44	72.38	99.01	95.24	100.00	91.43	100.00	89.52	96.88	88.57
AA	88.85	85.82	84.74	56.31	85.29	80.68	80.95	46.31	64.56	28.89	73.38	40.71	69.25	50.07	83.73	81.85	83.80	51.62
OA	0.87		0.74		0.85		0.61		0.49		0.58		0.67		0.84		0.68	
Kappa	0.8512		0.6949		0.8227		0.5338		0.3799		0.4893		0.6170		0.8136		0.6209	
Time	4.35		4.91		10.12		1.27		1.11		5.20		7.24		1.69		1.09	

这是因为这些地物的训练样本个数很少，如果处理不当的话，其在分类过程中极有可能被误认为是噪声样本，进而影响分类精度。但是 NSG, DSNPE2 和 MMC 没有出现这种现象，主要是因为这 3 种算法不用进行矩阵的逆运算，不仅可以避免矩阵的奇异问题，还可以避免小样本问题；尽管 NSG 在部分地物上的 PA 和 UA 低于其它算法，但在主要性能指标方面(AA, OA 和 Kappa 系数)，NSG 体现出了优越性；与同为基于稀疏表示的降维算法(DSNPE1, DSNPE2, NSPE 和 SPP)相比，NSG 的运算时间最

少；NSG 的运算时间要高于 S-LE, SLPP, MMC 和 LDA 方法，但是，NSG 的分类效果要优于上述方法，因此，增加出来的少量计算开销还是可以接受的。

最后，给出各算法在 ACER 高光谱数据集上的分类效果图，如图 5 所示。由图可以看出，NSG 能很好识别地物 Hay-windrowed, Wheat 和 Grass/Pasture, 但是将大量的 Bldg-Grass-Tree-Drive 误分成了 Woods, 将部分的 Corn-notill 误分成了 Soybeans-min, 其它几类地物也有部分的误分类。

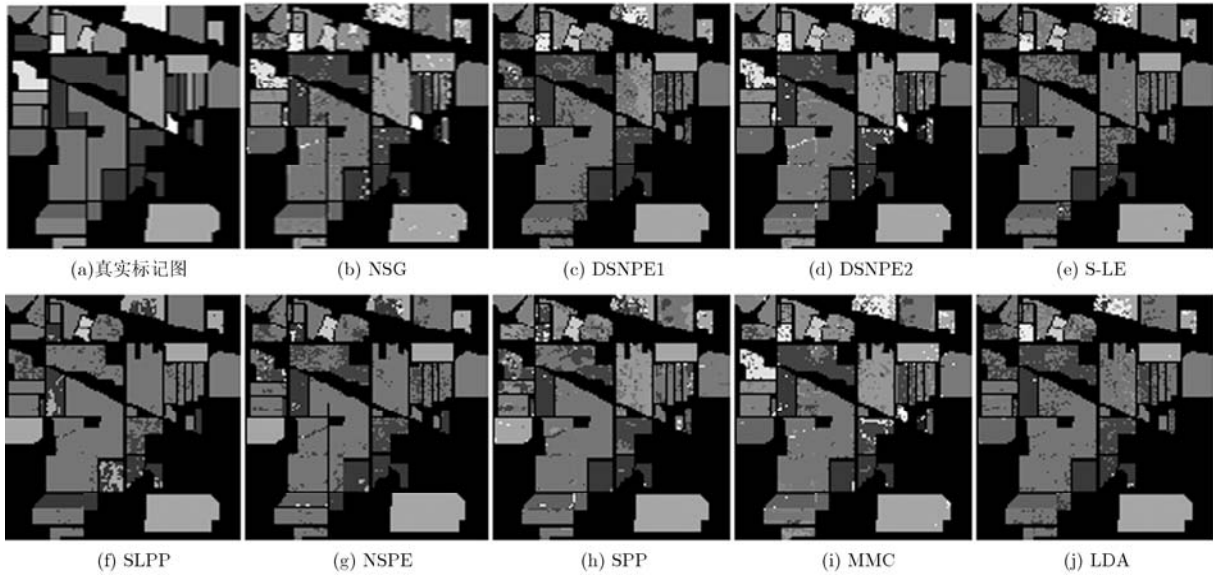


图 5 分类结果

其它 8 种算法都有类似的现象，且大部分算法误分地物的情况更为严重。Corn, Soybean 和 Grass 3 大类地物中样本量少的地物难以分类的主要原因在于：它们生长的地面含有大量的或中等程度的上季度残留物，且含量不一样；样本的含量较少很容易被误分成样本量大的地物。总的来说，NSG 能够较好地从高维高光谱数据中提取信息量大的光谱波段，并剔除冗余度高及重叠率高的光谱波段。因此，通过 NSG 降维后得到的图像更清晰、平滑，细节更丰富。

5 结束语

光谱遥感技术已经成为人们获取地球表面信息的一种非常重要的手段。然而，由于高光谱数据具有信息量大、信息相关性强、信息冗余度高和 Hughes 现象等特性，使得目前遥感信息的提取和利用水平大大滞后于遥感技术的发展。为提高高光谱数据的分类性能，本文提出一种非负稀疏图高光谱数据降维算法。算法的主要思想是：构建超完备块字典，提高非负稀疏表示重构精度的同时降低时间复杂度；同时最大化异类和最小化同类非负稀疏重构样本间的距离，能够使同类非负稀疏重构样本在低维子空间中相互靠近、异类非负稀疏重构样本在低维子空间中远离，从而提高算法的分类性能。尤其在数据量大且训练样本数目有限的情况下，效果更为明显。但是，算法仍有进一步改进的空间，比如高光谱数据实际上是张量形式的，而文中是将其转化成向量形式进行分类处理的，因此，如何将算法推广到张量空间是我们的下一步工作。

参 考 文 献

- [1] Li S J and Qi H R. Sparse representation based band selection for hyperspectral images[C]. Proceedings of 18th IEEE International Conference on Image Processing, Brussels, 2011: 11-14.
- [2] Hasanlou M and Samadzadegan F. Comparative study of intrinsic dimensionality estimation and dimension reduction techniques on hyperspectral images using K-NN classifier[J]. *IEEE Geoscience and Remote Sensing Letters*, 2012, 9(6): 1046-1050.
- [3] Hughes G. On the mean accuracy of statistical pattern recognizers[J]. *IEEE Transactions on Information Theory*, 1968, IT-14(1): 55-63.
- [4] Chang C I and Safavi H. Progressive dimensionality reduction by transform for hyperspectral imagery[J]. *Pattern Recognition*, 2011, 44(10/11): 2760-2773.
- [5] Roweis S T and Saul L K. Nonlinear dimensionality reduction by locally linear embedding[J]. *Science*, 2000, 290(5500): 2323-2326.
- [6] Tenenbaum J B, Silva V, and Langford J C. A global geometric framework for nonlinear dimensionality reduction[J]. *Science*, 2000, 290(5500): 2319-2323.
- [7] Belkin M and Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation[J]. *Neural Computation*, 2002, 15(6): 1373-1396.
- [8] 冯燕, 何明一, 宋江红, 等. 基于独立成分分析的高光谱图像数据降维及压缩[J]. *电子与信息学报*, 2007, 29(12): 2871-2875.
Feng Yan, He Ming-yi, and Song Jiang-hong, *et al.* ICA-based dimensionality reduction and compression of

- hyperspectral images[J]. *Journal of Electronics & Information Technology*, 2007, 29(12): 2871-2875.
- [9] 赵春晖, 张隼, 王玉磊. 基于小波核主成分分析的相关向量机高光谱图像分类[J]. 电子与信息学报, 2012, 34(8): 1905-1910. Zhao Chun-hui, Zhang Yi, and Wang Yu-lei. Relevant vector machine classification of hyperspectral image based on wavelet kernel principle component analysis[J]. *Journal of Electronics & Information Technology*, 2012, 34(8): 1905-1910.
- [10] Bellhumeur P, Hespanha J, and Kiregeman D. Eigenfaces vs. fisherfaces: recognition using class specific linear projection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, 19(7): 711-720.
- [11] He X F and Partha N. Locality preserving projections[C]. Proceedings of Advances in Neural Information Processing Systems, Canada, 2004, 16: 153-160.
- [12] Cortes C and Mohri M. On transductive regression[C]. Proceedings of Advances in Neural Information Processing Systems, Canada, 2007: 305-312.
- [13] Cheng Bin, Yang Jian-chao, Yan Shui-cheng, et al. Learning with l^1 -graph for image analysis[J]. *IEEE Transactions on Image Processing*, 2010, 19(4): 858-866.
- [14] Wright J, Ma Y, Mairal J, et al. Sparse representation for computer vision and pattern recognition[J]. *Proceedings of the IEEE*, 2009, 98(6): 1031-1044.
- [15] Yang Bo and Chen Song-can. Sample-dependent graph construction with application to dimensionality reduction[J]. *Neurocomputing*, 2010, 74(1-3): 301-314.
- [16] Raducanu B and Dornaika F. A supervised non-linear dimensionality reduction approach for manifold learning[J]. *Pattern Recognition*, 2012, 45(6): 2432-2444.
- [17] 宋相法, 焦李成. 基于稀疏表示及光谱信息的高光谱遥感图像分类[J]. 电子与信息学报, 2012, 34(2): 268-272. Song Xiang-fa and Jiao Li-cheng. Classification of hyperspectral remote sensing image based on sparse representation and spectral information[J]. *Journal of Electronics & Information Technology*, 2012, 34(2): 268-272.
- [18] 王伟伟, 廖桂生, 朱圣棋, 等. 一种基于压缩感知的地面运动目标检测方法[J]. 电子与信息学报, 2012, 34(8): 1872-1878. Wang Wei-wei, Liao Gui-sheng, Zhu Sheng-qi, et al. A ground moving target indication method based on compressive sensing[J]. *Journal of Electronics & Information Technology*, 2012, 34(8): 1872-1878.
- [19] Qiao Li-shan, Chen Song-can, and Tan Xiao-yang. Sparsity preserving projections with applications to face recognition[J]. *Pattern Recognition*, 2010, 43(1): 331-341.
- [20] Wong W K. Discover latent discriminant information for dimensionality reduction: non-negative sparseness preserving embedding[J]. *Pattern Recognition*, 2012, 45(4): 1511-1523.
- [21] Li H, Jiang T, and Zhang K. Efficient and robust feature extraction by maximum margin criterion[J]. *IEEE Transactions on Neural Networks*, 2006, 17(1): 1157-1165.
- [22] Gui Jie, Sun Zhe-nan, Jia Wei, et al. Discriminant sparse neighborhood preserving embedding for face recognition[J]. *Pattern Recognition*, 2012, 45(8): 2884-2893.
- [23] Lu Gui-fu, Jin Zhong, and Zou Jian. Face recognition using discriminant sparsity neighborhood preserving embedding[J]. *Knowledge-based Systems*, 2012, 31(2): 119-127.
- [24] Eldar Y C, Kuppinger P, and Bolcskei H. Block-sparse signals: uncertainty relations and efficient recovery[J]. *IEEE Transactions on Signal Processing*, 2010, 58(6): 3042-3054.
- [25] Wright J, Yang A Y, Ganesh A, et al. Robust face recognition via sparse representation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(2): 210-227.
- [26] Wang H S, Li G D, and Tsai C L. Regression coefficient and autoregressive order shrinkage and selection via the lasso[J]. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 2007, 69(1): 63-78.
- [27] Zou H and Hastie T. Regularization and variable selection via the elastic net[J]. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 2005, 67(2): 301-320.
- [28] He Ran, Hu Bao-gang, Zheng Wei-shi, et al. Two-stage sparse representation for robust recognition on large-scale database[C]. Proceedings of the National Conference on Artificial Intelligence, United States, 2010: 475-480.
- 高 阳: 男, 1985 年生, 博士生, 研究方向为遥感信息处理。
- 王雪松: 女, 1974 年生, 教授, 博士生导师, 主要从事遥感信息处理、机器学习方面的研究。
- 程玉虎: 男, 1973 年生, 教授, 博士生导师, 主要从事遥感信息处理、机器学习方面的研究。