

doi: 10. 6046/gtzyyg. 2017. 04. 07

引用格式: 董安国,龚文娟,韩雪. 基于线性表示的高光谱影像波段选择算法[J]. 国土资源遥感,2017,29(4):39-42. (Dong A G,Gong W J,Han X. Band selection method for hyperspectral image based on linear representation[J]. Remote Sensing for Land and Resources,2017,29(4):39-42.)

基于线性表示的高光谱影像波段选择算法

董安国,龚文娟,韩 雪

(长安大学理学院,西安 710064)

摘要: 为了去除高光谱影像的数据冗余,提高高光谱影像处理的精度和效率,提出了一种基于线性表示的高光谱影像波段选择算法。针对每一个波段,建立与其他波段的线性表示关系,依据复相关系数确定相关程度最高的波段,将其作为冗余波段去除;对剩余波段重复上述过程,得到最小波段集;并证明了利用该波段集和全波段所选的端元是一致的,在不影响端元提取的前提下,最大程度地去除了冗余波段。通过 2 组实验结果证明了该波段选择算法的可行性和有效性。

关键词: 高光谱影像;线性表示;波段选择;复相关系数算法

中图法分类号: TP 79 **文献标志码:** A **文章编号:** 1001-070X(2017)04-0039-04

0 引言

随着高光谱遥感技术的发展,高光谱传感器可以同时捕捉从可见光到红外范围的数百个波段,高光谱图像丰富的波段信息为地物分类和目标识别提供了数据支持,同时,其海量数据和信息冗余也给计算带来了困难。为了提高计算效率和精度,需要在不损失信息的情况下尽可能降低数据的维数。其主要方法有特征提取^[1]和波段选择 2 种。特征提取是利用线性或非线性变换来降低数据的维数,将原始的高维数据变换到一个新的低维空间,如主成分分析法^[2]和 Fisher 线性判别分析^[3]等,但是由于数据变换,使得降维后的数据在物理属性上发生了变化;而波段选择作为一种高光谱数据的预处理技术,从高光谱图像的上百个波段中,去除部分信息冗余波段,选择出能充分代表原始图像信息的波段子集,在保持数据信息不受损失的前提下去除部分波段,降维后的数据仍保持了原始数据的物理属性。

常见的波段选择算法有:①基于信息量的算法,如基于加强的快速密度峰值的聚类算法^[4]、基于三元互信息的波段选择算法^[5]、基于互信息的子空间检测算法^[6]、典型性成分分析^[7]和粗糙集理论

算法^[8]等,这类算法通常先根据各个波段的信息进行排序,选出排序靠前的一些波段,虽然简单易行,但容易选出连续波段;②基于搜索的算法,如灰狼优化算法^[9]、快速体积梯度算法^[10]、改进的萤火虫算法^[11]、列子集选择算法^[12]、克隆选择算法^[13]和基于正交投影散度的算法^[14]等,这类算法虽然可以解决波段选择过程中波段组合数目多、遍历难的问题,但算法相对复杂,通常需要的搜索时间较长,而且会存在收敛速度问题。

目前大部分的波段选择算法都是基于聚类的思想,根据两两波段的相似关系进行聚类,进而去除每个类中的部分波段,但并未对多个波段间的线性表示关系进行有效地刻画和利用。鉴于这一思想,提出一种新的基于线性表示的波段选择算法。该算法针对每一个波段,建立与其他波段的线性表示关系,依据复相关系数的大小剔除冗余波段,得到最小波段集。

1 算法理论

1.1 理论基础

假设高光谱图像中的像元数为 N ,端元数目为 s ,波段数为 L ,正整数集合 $\Psi_1 = \{1, 2, \cdots, N\}$, $\Psi_2 =$

收稿日期: 2016-05-25; 修订日期: 2016-08-08

基金项目: 国家自然科学基金项目“高分辨率遥感影像信息提取的特征结构化多尺度分析方法研究”(编号: 41571346)、“基于高维马尔可夫网结构统计方法的高光谱图像分割研究”(编号: 40971217)和“无限时滞脉冲泛函微分方程及其在经济中的应用”(编号: 11201038)共同资助。

第一作者: 董安国(1964-),男,教授,硕士生导师,主要从事数值代数及数字图像处理研究。Email: donganguo@chd.edu.cn。

通信作者: 韩雪(1987-),女,硕士研究生,主要从事数字图像处理方向研究。Email: gngwenjuan0919@163.com。

$\{1, 2, \dots, s\}, \Psi_3 = \{1, 2, \dots, L\}$ 。第 i 个像元的高光谱向量为 $\alpha_i (i \in \Psi_1)$, 记向量组 $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)$ 。在理想情形下, 根据线性表示理论, α 中一定存在若干个线性无关的向量 $\alpha_{k_1}, \alpha_{k_2}, \dots, \alpha_{k_s}$, 使得 α 中的每一个向量均可由 $\alpha_{k_1}, \alpha_{k_2}, \dots, \alpha_{k_s}$ 线性表示。

设向量 $\alpha_i = (\alpha_{1i}, \alpha_{2i}, \dots, \alpha_{Li})^T$, 其中 $i \in \Psi_1$, 则

$$\alpha = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1N} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_{L1} & \alpha_{L2} & \cdots & \alpha_{LN} \end{pmatrix}, \quad (1)$$

记其行向量组为 $X_1^T, X_2^T, \dots, X_L^T$, 列向量组为 $\alpha_1, \alpha_2, \dots, \alpha_N$, 若 X_t 可由 $\{X_k\}_{k \in \Psi_3 - \{t\}} (t \in \Psi_3)$ 线性表示, 且

$$\alpha_k = \lambda_1 \alpha_{k_1} + \lambda_2 \alpha_{k_2} + \cdots + \lambda_s \alpha_{k_s}, \quad (2)$$

则线性方程组

$$x_1 \alpha_{k_1} + x_2 \alpha_{k_2} + \cdots + x_s \alpha_{k_s} = \alpha_k, \quad (3)$$

有唯一解, 该方程组增广矩阵的第 t 行可由其他行线性表示, 所以从方程组中去掉第 t 个方程后与原方程是同解方程, 故可得式 (2) ($\alpha_{k_1}, \alpha_{k_2}, \dots, \alpha_{k_s}$ 线性无关) 的充分必要条件是

$$\tilde{\alpha}_k = \lambda_1 \tilde{\alpha}_{k_1} + \lambda_2 \tilde{\alpha}_{k_2} + \cdots + \lambda_s \tilde{\alpha}_{k_s}, \quad (4)$$

式中 $\tilde{\alpha}_k$ 是 $\alpha_k (k \in \Psi_1)$ 去掉第 t 个分量得到的向量。

由此可进一步得出, 若 $X_i (i \in \Psi_4 \subset \Psi_3)$ 为 $X_1, X_2, \dots, X_L$ 的极大线性无关组, 则式 (2) 的充分必要条件是式 (4), 其中 $\tilde{\alpha}_k$ 是 $\alpha_k (k \in \Psi_1)$ 只保留下标集 Ψ_4 中的分量得到的向量。

由上述推论可知, 当一像元向量可以由若干个线性不相关的其他像元向量线性表示时, 删除可被其他波段向量线性表示的波段后, 每个像元与其他像元之间的线性表示关系不变, 故波段选择可以转化为求矩阵行的极大线性无关组。

1.2 评价指标

复相关系数是反映一个因变量与一组自变量之间相关程度的指标, 其值越大, 变量之间的线性相关程度越密切, 可以用来判断某向量是否能被其他向量近似线性表示。设线性回归模型为

$$y = x_1 \beta_1 + x_2 \beta_2 + \cdots + x_p \beta_p + \varepsilon, \quad (5)$$

式中: 响应变量 y 的观测值为 $(y_1, y_2, \dots, y_n)^T$; x_i 的观测值为 $(x_{i1}, x_{i2}, \dots, x_{in})^T (i = 1, 2, \dots, p)$; ε 为随机误差向量。利用最小二乘法可得

$$\beta = (\beta_1, \beta_2, \dots, \beta_p)^T = (x^T x)^{-1} x^T y. \quad (6)$$

y 与 x 之间的复相关系数 $R (0 \leq R \leq 1)$ 的计算公式为 **万方数据**

$$R = \sqrt{1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (7)$$

$$\hat{y}_i = x_{1i} \beta_1 + x_{2i} \beta_2 + \cdots + x_{pi} \beta_p, \quad (8)$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i. \quad (9)$$

由 1.1 节理论基础可知, 采用复相关系数进行高光谱波段选择, 在去掉若干波段后, 剩余的波段仍能代表原始高光谱波段的信息。

2 波段选择算法

2.1 算法流程

计算 X_i 与 $\{X_k\}_{k \in \Psi_3 - \{i\}}$ 之间的 $R_i (i \in \Psi_3)$; 若 $R_{i0} = \max_{i \in \Psi_3} (R_i)$, 且 $R_{i0} > \alpha$ (α 为阈值), 则表明 X_{i0} 可以被其他向量近似线性表示, 将该向量对应的波段去除; 依次对剩余波段重复上述过程, 直到不能去除为止。

算法的具体步骤如下:

1) 输入高光谱图像数据 A (其中每个行向量代表一个对应波段)、集合 $S = \Psi_3$ 和阈值 α 。

2) 对所有 $i \in S$, 计算 X_i 与 $\{X_k\}_{k \in S - \{i\}}$ 的 R_i 。

3) 获取 $R_{i0} = \max_{i \in S} (R_i)$, 如果 $R_{i0} > \alpha, S = S - \{i0\}$, 转第 2 步; 否则, 输出 S 作为波段选择结果集, 算法结束。

2.2 算法优化

设 X_i 与 $\{X_k\}_{k \in \Psi_3 - \{i\}}$ 的线性表达系数为 $\beta^{(i)}$, 则

$$\beta^{(i)} = (X_{(i)}^T X_{(i)})^{-1} X_{(i)}^T X_i, \quad (10)$$

式中 $X_{(i)}$ 表示从 $X = (X_1, X_2, \dots, X_L)$ 中删除第 i 列得到的矩阵。对不同的 $i, j \in \Psi_3$, 矩阵 $X_{(i)}$ 和 $X_{(j)}$ 只有一列不同, 导致求 $R_i (i \in \Psi_3)$ 时, 存在大量重复计算, 为了简化计算, 进行算法优化。

设矩阵 $B = (b_{kj})_{L \times L} = X^T X$, 则对任意 $i \in \Psi_3$, $C^{(i)} = (c_{kj}^{(i)})_{(L-1) \times (L-1)} = X_{(i)}^T X_{(i)}, b_{kj} = X_k^T X_j, c_{kj}^{(i)} = (X_{(i)})_k^T (X_{(i)})_j$, 其中, $(X_{(i)})_k$ 表示矩阵 $X_{(i)}$ 的第 k 列, 由 $X_{(i)}$ 的定义, 当 $k < i$ 时, $(X_{(i)})_k = X_k$, 当 $k \geq i$ 时, $(X_{(i)})_k = X_{k+1}$ 。从而有

$$c_{kj}^{(i)} = \begin{cases} b_{kj} & k, j < i \\ b_{k+1, j} & k \geq i, j < i \\ b_{k, j+1} & k < i, j \geq i \\ b_{k+1, j+1} & k \geq i, j \geq i \end{cases}, \quad (11)$$

进而可以得出, 矩阵 $X_{(i)}^T X_{(i)}$ 是 B 划去第 i 行第 i 列

得到的子矩阵。由 $\mathbf{B} = (b_{k,j})_{L \times L} = \mathbf{X}^T \mathbf{X}$, 可知 $b_{k,j} = \mathbf{X}_k^T \mathbf{X}_j$, $\mathbf{B}$ 的第 i 列为 $\mathbf{X}^T \mathbf{X}_i$, 即

$$\mathbf{X}^T \mathbf{X}_i = (b_{1,i}, b_{2,i}, \dots, b_{i,i}, \dots, b_{L,i})^T, \quad (12)$$

$$\mathbf{X}_{(i)}^T \mathbf{X}_i = (b_{1,i}, b_{2,i}, \dots, b_{i-1,i}, b_{i+1,i}, \dots, b_{L,i})^T. \quad (13)$$

依据以上结论,原算法第1步后先进行优化,即:

- 1) $\mathbf{X} = \mathbf{A}^T, \mathbf{B} = \mathbf{X}^T \mathbf{X}$.
- 2) 对 $i \in S$, 划去 $\mathbf{B}$ 的第 i 行第 i 列得到的子矩阵 $\mathbf{C}^{(i)}$, 并记

$$\mathbf{D}^{(i)} = (b_{1,i}, b_{2,i}, \dots, b_{i-1,i}, b_{i+1,i}, \dots, b_{L,i})^T, \quad (14)$$

计算

$$\boldsymbol{\beta}^{(i)} = (\mathbf{C}^{(i)})^{-1} \mathbf{D}^{(i)}, \quad (15)$$

由公式(7)计算 $\mathbf{X}_i$ 与 $\{\mathbf{X}_k\}_{k \in S - \{i\}}$ 的 R_i , 转原算法第3步。

通过优化可知,计算工作量主要集中在第1步,整个迭代过程没有出现重复计算。

3 实验结果与分析

为了验证本文算法的有效性,分别采用一组具体数字和美国内华达州 Cuprite 地区的 AVIRIS 数据进行实验,并在全波段和利用本文算法选出的波段下分别提取端元,分析提取效果。

3.1 实验1

给定5个互不相关的100维向量作为端元,分别用 $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5$ 来表示,随机产生10 000组归一化的非负系数(每组5个),每组系数作为 $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5$ 的组合系数合成10 000个向量。将包括 $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5$ 在内的10 005个向量视作具有100个波段的高光谱向量(属理想情形),用本文算

法进行波段选择,取 $\alpha = 0.995$ 阈值时,所选择的波段数目为5。采用文献[15]中的端元提取方法对本文算法所选的5个波段进行端元提取实验,得到的端元恰好是 $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5$ 。

在上述10 000个向量中,附加高斯噪声(属非理想情形),采用本文算法进行波段选择,阈值 $\alpha = 0.95$,所选择的波段数目仍为5。采用文献[15]中方法提取得到的端元也仍为 $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5$ 。

为了使实验更具有普适性,再分别给定8和12个互不相关的100维向量作为端元,并分别用 $m_1, m_2, \dots, m_8$ 和 $n_1, n_2, \dots, n_{12}$ 来表示,重复进行上述实验,实验结果如表1所示。

表1 端元提取结果

Tab.1 Results of endmember extraction

情况	端元个数	全波段下的端元	本文算法所选波段下的端元
理想情况	5	a_1, a_2, a_3, a_4, a_5	a_1, a_2, a_3, a_4, a_5
	8	$m_1, m_2, \dots, m_8$	$m_1, m_2, \dots, m_8$
	12	$n_1, n_2, \dots, n_{12}$	$n_1, n_2, \dots, n_{12}$
非理想情况	5	a_1, a_2, a_3, a_4, a_5	a_1, a_2, a_3, a_4, a_5
	8	$m_1, m_2, \dots, m_8$	$m_1, m_2, \dots, m_8$
	12	$n_1, n_2, \dots, n_{12}$	$n_1, n_2, \dots, n_{12}$

由表1可以看出,在理想和 non-ideal 情况下,用文献[15]中的端元提取方法在全波段下和采用本文算法所选波段下提取出的端元是一致的,从而验证了本文算法的准确性和有效性。

3.2 实验2

Cuprite 地区 AVIRIS 原始数据共有224个光谱波段,光谱分辨率为10 nm。本文删除了噪声较大和光谱吸收较大的波段,选择172—221光谱区间的50个波段用于算法测试。在实验中,取阈值 $\alpha = 0.995$,所选择的波段数目为20,并采用文献[15]中端元提取方法,分别基于全波段及本文算法所选波段提取端元,最终确定对应的矿物类型。

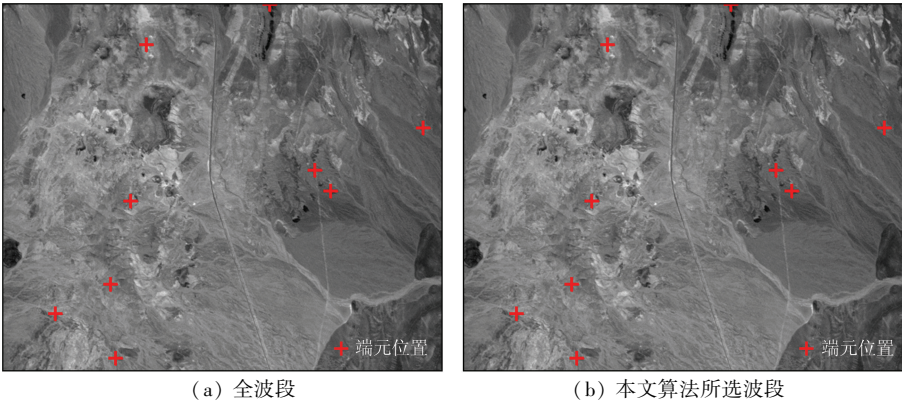


图1 端元位置
Fig.1 Location of the endelement

由图1对比可以看出,2种波段范围选择的端元位置是一致的,即在本文算法所选波段与全波段

下提取出的端元是相同的。图中的9个端元位置坐标分别为: (2,374), (58,202), (174,588), (233,

437), (262, 459), (276, 180), (392, 152), (433, 75) 和 (495, 159)。将所选择端元进行丰度分解, 便可确定出各端元所对应的矿物类型。由本文算法选出的少量波段与全波段提取的端元相同, 不仅减少了计算量, 还达到了较好的应用效果, 证明了该算法的可行性与有效性, 为后续工作的开展提供了帮助。

4 结论

基于线性表示提出了一种新的高光谱影像波段选择算法, 该算法利用复相关系数进行波段选择, 并对算法进行了优化, 克服了以往算法仅仅计算两两波段间相关性的不足, 而从整体上考虑波段间的相关性, 简化了计算量的同时, 起到了降噪的目的。实验结果表明了该算法在波段选择中的可行性与有效性, 并且具有普适性。该算法为后续高光谱影像的地物识别提供了一种技术支持。

参考文献 (References):

- [1] Jia X P, Kuo B C, Crawford M M. Feature mining for hyperspectral image classification[J]. Proceedings of the IEEE, 2013, 101(3): 676 – 697.
- [2] Agarwal A, El – Ghazawi T, El – Askary H, et al. Efficient hierarchical – PCA dimension reduction for hyperspectral imagery[C]// Proceedings of 2007 IEEE international symposium on signal processing and information technology. Giza: IEEE, 2007: 353 – 356.
- [3] Lei W, Prasad S, Fowler J E, et al. Locality – preserving dimensionality reduction and classification for hyperspectral image analysis[J]. IEEE Transactions on Geoscience and Remote Sensing, 2012, 50(4): 1185 – 1198.
- [4] Jia S, Tang G H, Zhu J S, et al. A novel ranking – based clustering approach for hyperspectral band selection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(1): 88 – 102.
- [5] Feng J, Jiao L C, Zhang X R, et al. Hyperspectral band selection based on trivariate mutual information and clonal selection[J].

IEEE Transactions on Geoscience and Remote Sensing, 2014, 52(7): 4092 – 4105.

- [6] Hossain M A, Jia X P, Pickering M. Subspace detection using a mutual information measure for hyperspectral image classification[J]. IEEE Geoscience Remote Sensing Letters, 2014, 11(2): 424 – 428.
- [7] Sun K, Geng X R, Ji L Y. Exemplar component analysis: A fast band selection method for hyperspectral imagery[J]. IEEE Geoscience Remote Sensing Letters, 2015, 12(5): 998 – 1002.
- [8] Patra S, Modi P, Bruzzone L. Hyperspectral band selection based on rough set[J]. IEEE Transactions on Geoscience and Remote Sensing, 2015, 53(10): 5495 – 5503.
- [9] Medjahed S A, Saadi T A, Benyettou A, et al. Gray Wolf Optimizer for hyperspectral band selection[J]. Applied Soft Computing, 2016, 40: 178 – 186.
- [10] Geng X R, Sun K, Ji L Y, et al. A fast volume – gradient – based band selection method for hyperspectral image[J]. IEEE Transactions on Geoscience and Remote Sensing, 2014, 52(11): 7111 – 7119.
- [11] Su H J, Yong B, Du Q. Hyperspectral band selection using improved firefly algorithm[J]. IEEE Geoscience Remote Sensing Letters, 2016, 13(1): 68 – 72.
- [12] Wang C, Gong M G, Zhang M Y, et al. Unsupervised hyperspectral image band selection via column subset selection[J]. IEEE Geoscience Remote Sensing Letters, 2015, 12(7): 1411 – 1415.
- [13] Feng J, Jiao L C, Liu F, et al. Mutual – information – based semi – supervised hyperspectral band selection with high discrimination, high information, and low redundancy[J]. IEEE Transactions on Geoscience and Remote Sensing, 2015, 53(5): 2956 – 2969.
- [14] 苏红军, 盛业华, He Y, 等. 基于正交投影散度的高光谱遥感波段选择算法[J]. 光谱学与光谱分析, 2011, 31(5): 1309 – 1313.
Su H J, Sheng Y H, He Y, et al. Orthogonal projection divergence – based hyperspectral band selection[J]. Spectroscopy and Spectral Analysis, 2011, 31(5): 1309 – 1313.
- [15] Gao L R, Gao J W, Li J, et al. Multiple algorithm integration based on ant colony optimization for endmember extraction from hyperspectral imagery[J]. IEEE Journal of Selected Topics in Applied Earth Observations in Remote Sensing, 2015, 8(6): 2569 – 2582.

Band selection method for hyperspectral image based on linear representation

DONG Anguo, GONG Wenjuan, HAN Xue

(School of Science, Chang'an University, Xi'an 710064, China)

Abstract: In order to remove the data redundancy of hyperspectral image and improve the accuracy and efficiency of hyperspectral image processing, this paper proposes a band selection method based on linear representation of hyperspectral image. A linear relationship is established for a band with the other bands, and the most relevant band is removed as a redundant band which is determined based on the multiple correlation coefficient. The set of minimum bands is finally obtained by repeating the above process for the remaining bands. It is proved that the set of selected endmembers by using the above bands is consistent with the set selected by using all bands, and the redundancy bands are removed to the greatest extent without affecting the endmember extraction. The experimental results show that the band selection algorithm in the paper is feasible and effective.

Keywords: hyperspectral images; linear representation; band selection; multiple correlation coefficient algorithm
万方数据 (责任编辑: 陈 理)