

非线性变换和信息相邻相关的高光谱自适应波段选择

张爱武^{1,2}, 杜楠^{1,2}, 康孝岩^{1,2}, 郭超凡^{1,2}

(1. 首都师范大学 三维信息获取与应用教育部重点实验室, 北京 100048;

2. 首都师范大学 空间信息技术教育部工程研究中心, 北京 100048)

摘要: 通过非线性函数变换改进后的谱间 Pearson 相关分析可同时获取高光谱影像光谱间的综合相关系数(r_{cl})、相关类型和统计显著性水平; 研究发现, 非线性是高光谱影像的谱间相关性的主要类型。基于相关系数的波段相邻相关系数(r_{ac}) 在自适应波段选择算法(ABS)中是为了表达波段的独立性, 然而发现 ABS 算法中 r_{ac} 并不能有效表达波段独立性。鉴于此, 提出了一种信息相邻相关系数(r_{iac}) 和基于此指数改进的自适应波段选择算法(MABS)。使用公共数据和实验室采集数据, 对 ABS、基于线性相关系数(r_l) 的 MABS(r_l) 和基于 r_{cl} 的 MABS(r_{cl}) 等三种算法进行实验。结果表明: 在波段范围和算法有效性及精度方面, MABS 均优于 ABS; MABS 较好地兼顾了大信息量和强独立性原则, 其波段选择结果的光谱范围明显大于 ABS; MABS(r_{cl}) 的光谱范围略大于 MABS(r_l); 三种算法的总体分类精度(OA)和 Kappa 系数的大小顺序均为: MABS(r_{cl})>MABS(r_l)>ABS。

关键词: 非线性相关; 谱间相关系数; 信息相邻相关; 波段选择; 高光谱图像

中图分类号: TP753 **文献标志码:** A **DOI:** 10.3788/IRLA201746.0538001

Hyperspectral adaptive band selection method through nonlinear transform and information adjacency correlation

Zhang Aiwu^{1,2}, Du Nan^{1,2}, Kang Xiaoyan^{1,2}, Guo Chaofan^{1,2}

(1. Key Laboratory of 3D Information Acquisition and Application, Ministry of Education, Capital Normal University,

Beijing 100048, China; 2. Engineering Research Center of Spatial Information Technology, Ministry of Education,

Capital Normal University, Beijing 100048, China)

Abstract: Through nonlinear functional transform of hyperspectral remote sensing data, the modified Pearson correlation analysis can effectively identify comprehensive correlation coefficient (r_{cl}), correlation type, and statistical significance level between spectrums. In this paper, nonlinear correlation the main correlation relationship type between hyperspectral bands was proved. Based on correlation coefficient, the adjacent bands' correlation coefficient (r_{ac}) of adaptive band selection (ABS) is to express band independence, but r_{ac} of ABS algorithm cannot effectively express such independence. Herein, a kind of information adjacency/equivalent bands' correlation coefficient (r_{iac}), and via this index, the modified ABS (MABS) were proposed. Using public data and collected private data, the experiments of ABS, MABS(r_l)

收稿日期: 2016-09-05; 修订日期: 2016-10-15

基金项目: 国家自然科学基金(41571369); 教育部博士点基金(20131108110005); 北京市长城学者(20150323)

作者简介: 张爱武(1972-), 女, 教授, 博士生导师, 博士, 主要从事空间信息获取与处理、三维激光成像、高光谱成像等方面的研究。

Email: zhangaw98@163.com

based on linear correlation coefficient(r_l), and MABS(r_{cl}) based on rcl were carried out. These two case studies demonstrate that MABS is superior to ABS on spectral range, algorithm validity and accuracy. MABS can take both large amount of information and strong independence into consideration effectively. The spectral range of MABS's bands selection result is more than ABS's obviously, and MABS (r_{cl})'s is a little more than MABS (r_l)'s. The ranking both overall classification accuracy and Kappa coefficient of those three kinds of algorithms are MABS(r_{cl})>MABS(r_l)>ABS.

Key words: nonlinear correlation; spectral correlation coefficient; information adjacency correlation; band selection; hyperspectral imagery

0 引言

图谱合一是高光谱影像有别于宽波段遥感数据的突出特征,使得几百个窄波段包含有光谱、辐射和图像等丰富信息^[1]。高光谱遥感的光谱具有连续性,且谱间波长差异在纳米数量级,故而其光谱的谱间相关程度较高^[2]。度量谱间相关性的经典指标有协方差和 Pearson 积矩相关系数 (Pearson's product-moment correlation coefficient),后者用于描述两个随机变量的线性相关程度,简称“Pearson 相关系数”,被广泛应用于高光谱影像的压缩、降维和分类等数据处理中。针对上述两种经典指标只能表征谱间的线性相关性,而对非线性相关性缺乏把握,范雪莉等^[3]提出用互信息 (Mutual Information, MI)来衡量谱间相关程度,但该方法并不能明确谱间共变关系属于线性或是非线性,并且 MI 对真实地物类型的依赖性较大^[4]。

鉴于光谱的高维度特征,高光谱数据直接参与图像分类时会引起 Hughes 现象,这使得高光谱降维处理变得尤为重要。波段选择是高光谱降维的一类有效手段^[5],在保证特征空间得到优化的同时,又较好地保留光谱信息^[6]。波段选择的模型多样,而模型评价准则通常包括信息量度量、独立性度量和类间可分性度量等^[7]三个方面;这其中,非监督方法着重于前两者的把握,而后者则需要有监督的算法来实现^[8]。相较于基于结合蚁群算法和遗传算法等复杂模型的波段选择方法^[9],最佳指数因子法 (Optimal Index Factor, OIF) 作为一种简单的无监督算法综合表达了评价准则的前两项,但其庞大的计算量使其不具有实践性^[10]。针对 OIF 的缺陷,刘春红等^[11,10]提出一种改进的 OIF,即自适应波段选择 (Adaptive

Band Selection, ABS)模型;ABS 大大降低了计算量,但容易选取相连波段,并且仅通过相邻波段相关性来体现波段的独立性并不全面。

综上所述,无论从谱间相关性评价还是波段选择方法上看都有需要提升和改进的空间。提出了一种非线性变换和信息相邻相关的高光谱自适应波段选择算法:首先,范修斌等^[11]从数理角度论证了互信息(可揭示谱间非线性共变特征^[13])与 Pearson 相关系数之间存在一定的关系,因此可以证明“通过对波段信息进行经典非线性函数变换后, Pearson 综合相关系数可以揭示波谱间的部分非线性关联关系”的假设成立;其次,采用按信息量排序后的波段“信息相邻”相关系数来表达波段的独立性,并通过实验验证;接着,结合上述两个方面内容,提出非线性变换和信息相邻相关结合的 ABS 波段选择方法。最后,将波段选择结果用于地物分类,并对比不同波段选择结果对地物分类的总体分类精度 (Overall Accuracy, OA)和 Kappa 系数的影响,进一步证明该研究提出的波段选择方法的正确性和可行性。

1 非线性函数变换的谱间综合相关系数

假设通过对波段信息进行经典非线性函数变换后的 Pearson 综合相关系数可以揭示波谱间的部分非线性关联关系,下文是具体论证假设成立。

1.1 Pearson 相关系数概述

综观高光谱影像降维的诸类理论方法,不管特征提取或是波段选择,不少理论中均涉及有 Pearson (积矩)相关系数的求解并作为模型的一部分或一个参量,如 OIF、ABS、ASP (auto-subspace partition)、PCA (principal components analysis)、MNF (minimum

noise fraction)等。以光谱间 Pearson 相关系数为例,其计算公式如下:

$$r(A, B) = \frac{\sum_{i=1}^n (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_{i=1}^n (a_i - \bar{a})^2} \sqrt{\sum_{i=1}^n (b_i - \bar{b})^2}} \quad (1)$$

式中: n 是光谱矢量 A 和 B 的维数; a_i 和 b_i 分别表示光谱矢量 A 和 B 的第 i 个维度的值; \bar{a} 和 \bar{b} 分别是两个光谱矢量的均值; r 取值为 $[-1, 1]$, $|r|$ 越大表示波段之间的线性相关性越大; 反之亦然。

Pearson 相关系数描述的是谱间的线性共变程度, 缺乏对谱间可能存在的非线性相关特征的把握, 针对这个问题, 有学者提出用互信息 MI 来衡量谱间线性/非线性相关程度^[3], 但该方法并不能明确谱间共变关系属于线性或是非线性, 并且 MI 对真实地物类型的依赖性较大^[4]。范修斌等^[1]从数理角度论证了互信息(可揭示谱间非线性共变特征^[3])与 Pearson 相关系数之间存在一定的关系, 因而该研究假设通过对波段信息进行经典非线性函数变换后, Pearson 相关系数可以揭示波谱间的部分非线性关联关系。

1.2 非线性函数变换的 Pearson 相关系数

该研究拟采用非线性函数变换后的相关性测度方法来更为全面地表达“曲线相关性 r_c (curve correlation)”的概念, 现有的五种典型非线性函数中, 三角函数和反三角函数均是关于角度的函数, 而幂函数和指数函数变换的时间复杂度相对较大, 且对计算机软硬件的性能要求较高, 故而, 研究选用对数函数来明确定义曲线相关性, 其公式如下:

$$r_c(A, B) = \pm \max[|r(\ln A, B)|, |r(A, \ln B)|, |r(\ln A, \ln B)|] \quad (2)$$

式中: $r(\ln A, B)$ 、 $r(A, \ln B)$ 、 $r(\ln A, \ln B)$ 分别表示波段 A 和 B 的三种非线性相关系数; $r_c(A, B)$ 取它们中绝对值最大的值, 符号取绝对值最大的相关系数的符号。

同理, 比较曲线相关系数与线性相关系数 r_l (linear correlation), 得到改进的 Pearson 综合相关系数 r_{cl} :

$$r_{cl}(A, B) = \pm \max[|r_l(A, B)|, |r_c(A, B)|] = \pm \max[|r(A, B)|, |r(\ln A, B)|, |r(A, \ln B)|, |r(\ln A, \ln B)|] \quad (3)$$

式中各参数与上式相同。

由此, 可以明确得出任意光谱间的综合相关系数(绝对值最大的相关系数)、综合相关类型(绝对值最大的相关系数的类型, 包括非线性相关类型(三种)和线性相关类型)以及统计意义上的显著性水平高

低。通过光谱之间的相似相关程度及类型的可视化显示, 可以更好的发现规律, 以便下一步的研究工作。

1.3 实证分析

该研究选用如下两组数据对基于非线性函数变换改进的谱间综合相关性的有效性进行验证, 并作为后一步在波段选择应用的数据基础。

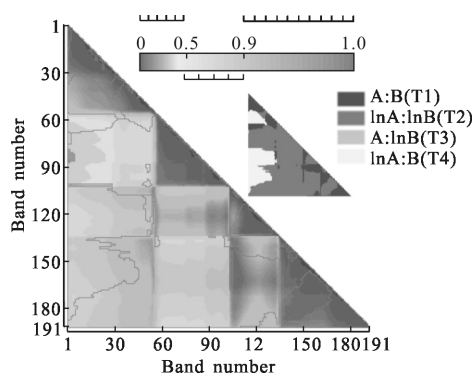
1.3.1 实验 1

实验数据为 HYDICE (hyperspectral digital imagery collection experiment) 于 1995 年获取的华盛顿特区(Washington DC Mall)的航空高光谱影像。截取图像大小为 $400 \text{ pixel} \times 300 \text{ pixel}$, 光谱范围 $400 \sim 2400 \text{ nm}$, 共包含 210 个波段(光谱分辨率为 10.022 nm), 覆盖了建筑、道路、阴影、草地、树木和小径等 6 种确定分类。去除所有无效波段后, 保留 191 个有效波段。取第 59, 26, 16 波段分别作为 R, G, B 分量合成假彩色图(图 1(a))。



(a) 研究区及数据示意

(a) Schematic plot of study area and data



(b) 综合相关系数及其类型

(b) Correlation coefficients and their types

图 1 华盛顿市区数据及其相关分析

Fig.1 Data and correlation analysis of Washington DC Mall

首先,按照公式(2)中的三种非线性模型对初始化后的数据进行非线性变换;然后,根据公式(1),解算非线性变换后的数据和初始化数据,得到四种相关系数;最后,通过公式(3)计算所有波段两两之间的综合相关系数及其类型,如图 1(b)所示。

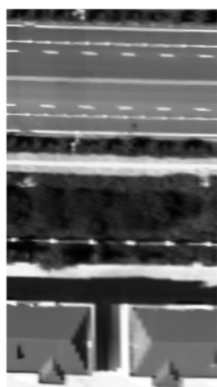
从图 1(b)可见,首先,两两波段之间的综合相关系数类型不尽相同,并且非线性相关类型(超过总体的 80%)远多于线性相关,四种类型大小依次为 T2(63.65%)>T4(18.66%)>T1(17.60%)>T3(0.09%);其次,总体而言,波段与相邻相近的波段之间的相关性类型多为线性相关或 T2 型非线性相关,以相邻相近 30 个波段为例,这两种类型的占比多达 94.79%;最后,图中还验证了前人就谱间线性相关性研究的一些结论^[1],如相距越近的波段间的相关性越大,相关阵可划分为一系列明显的小方块阵等。

1.3.2 实验 2

实验数据为基于飞艇的高光谱相机于 2014 年采集的青海湖附近某城区的图像。截取样区图像大小为 900 pixel×500 pixel,光谱范围为 400~1 000 nm(可见光和近红外),共包含 840 个波段(光谱分辨率为 0.727 nm),去除噪声和饱和波段后,共剩余 800 个波段,覆盖了建筑、植被、阴影、柏油路面、水泥路面、标识线 1 和标识线 2 等 7 种确定类型。取 360, 189, 69 波段分别作为 R, G, B 分量合成假彩色图(图 2(a))。

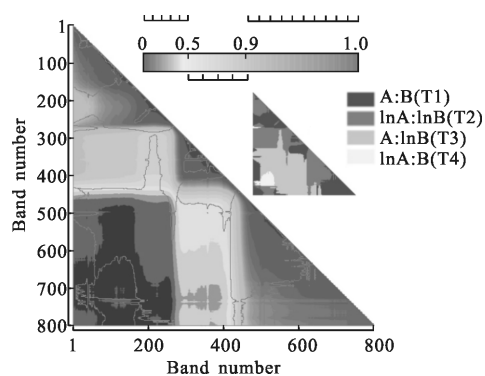
与实验 1 同样的数据处理方法和步骤,得到该实验数据所有波段两两之间的综合相关系数及其类型的可视化如图 2(b)所示。

图 2(b)与图 1(b)对比可见,图 2(b)反映出相似的结果:(1)非线性相关类型占总体的多数,占比



(a) 研究区及数据示意

(a) Schematic plot of study area and data



(b) 综合相关系数及其类型

(b) Correlation coefficients and their types

图 2 海晏县数据及其相关分析

Fig.2 Data and correlation analysis of Haiyan County

73.85%;(2)以相邻相近 100 个波段为例,线性相关和 T2 型非线性相关类型共占九成以上(T1:50.63%; T2:48.74%),故这两种类型对于近距离波段间的相关性有着较好的表征性;(3)此处与前人基于线性相关研究谱间关系的一些结论相一致^[1],如呈现出分块子空间等特征,并且这些特征与光谱范围的大小无关。

2 信息相邻相关的 ABS 非监督波段选择

经典的 ABS 模型分别选用标准差和线性相关系数来表达波段信息量和独立性大小,本节将选用上文提出的非线性变换的综合相关系数对 ABS 进行改进:首先,厘清 ABS 的模型构建基础;其次,指出 ABS 波段相邻相关系数的不足,并提出一种信息相邻相关系数及其改进算法;最后进行对比实证分析。

2.1 OIF 和 ABS 概述

波段选择模型需要满足如下 3 个准则^[7]:

- (1) 信息量度量:所选波段的波段信息要大;
- (2) 独立性度量:所选波段与其他波段的相关性要小;
- (3) 类间可分性度量:研究区内类别地物的可分性要强。

这其中,非监督方法着重于前两者的把握,而类间可分性则需要有监督的算法来实现。作为一种简单的无监督算法,OIF(公式(4))综合表达了评价准则的前两项,但其庞大的计算量使其不具有实践性^[8]。

$$OIF = \frac{\sum_{i=1}^n \sigma_i}{\sum_{i=1}^n \sum_{j=i+1}^n |r_{ij}|} \quad (4)$$

式中： σ_i 为第 i 个波段的标准差； r_{ij} 为第 i, j 波段间的 Pearson 线性相关系数； n 为波段组合的波段数目。

针对 OIF 的缺陷，刘春红等^[9]提出一种改进的 OIF，即 ABS 模型(公式(5))；ABS 大大降低了计算量，但容易选出连续波段。

$$ABS = \frac{\sigma_i}{(r_{i-1,i} + r_{i,i+1})/2} \quad (5)$$

式中： $r_{i-1,i}$ 和 $r_{i,i+1}$ 为第 i 波段和前后两个波段间的 Pearson 线性相关系数；其他参数与上文相同。

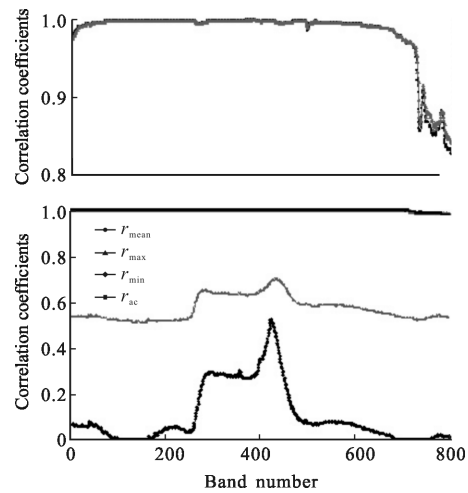
2.2 波段相邻相关和信息相邻相关

Pearson 相关性可表达波段独立性，相关系数越大，则波段间独立性越小；反之亦然。从图 1(b)和图 2(b)可见，波段与其相邻的前后两个波段的相关性极强，趋近于 1，而与相距越远的波段的相关性呈现减小趋势，最小值趋近于 0。因此，理论上以波段相邻相关系数 r_{ac} (adjacent correlation, 公式(6))来近似表达波段与其他所有波段的独立性，缺乏一定的说服力。

$$r_{ac}(i) = (r_{cl}(i-1, i) + r_{cl}(i, i+1))/2 \quad (6)$$

式中： $r_{ac}(i)$ 为第 i 波段的波段相邻相关系数，其他参量参考公式(3)的定义。

图 3 展示的是实验 1 和 2 中每个波段与其他波段的相关系数均值 r_{mean} (公式(7))、最大值 r_{max} (公式(8))、最小值 r_{min} (公式(9))和相邻相关系数 r_{ac} 。可见， r_{ac} 与 r_{max} 不仅具有显著的一致性(实验 1: 两者的线性相关系数 $r=0.7683, P=0.0000$; 实验 2: 两者的线性相关系数 $r=0.9995, P=0.0000$)，且相差不大(实



(b) 实验 2

(b) Test 2

图 3 单个波段与其他所有波段间的多种相关系数

Fig.3 Correlation coefficients between a band and other bands

验 1: 两者差值的均值为 0.001 855; 实验 2: 两者差值的均值为 0.000 122)。而 r_{ac} 与 r_{mean} 则不具有共变性质(实验 1: 两者不相关: $r=-0.0412, P=0.5712$; 实验 2: 两者弱相关: $r=0.2822, P=0.0000$)。

$$r_{mean}(i) = \frac{\sum_{j=1}^n r_{cl}(i, j) - 1}{n - 1} \quad (7)$$

$$r_{max}(i) = \max\{r_{cl}(i, j) | j \in [1, n], j \neq i\} \quad (8)$$

$$r_{min}(i) = \min\{r_{cl}(i, j) | j \in [1, n], j \neq i\} \quad (9)$$

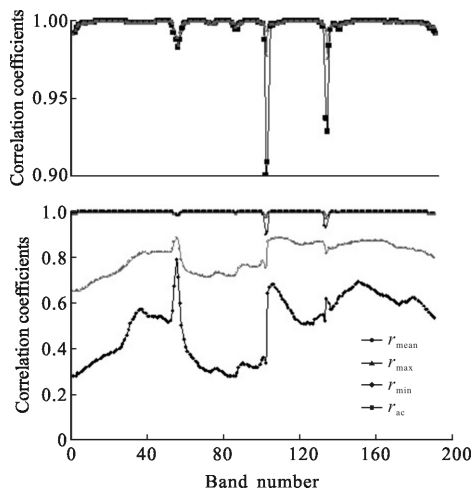
式中： n 为全波段总数，其他各参量参考公式(3)的定义。

实验结果表明，在表征波段与其他所有波段的相关性上，与 r_{max} 等同的 r_{ac} 不具有代表性，因此，用 r_{ac} 来表达波段的独立性的方法并不全面。

为此，该研究假设：设 b_k^σ 为按信息排序后的排列 $B_n^\sigma: b_1^\sigma, b_2^\sigma, \dots, b_k^\sigma, \dots, b_n^\sigma$ 在 k 位置上的波段，那么在与 b_k^σ 信息量相当(相近)的波段排列中， b_k^σ 与相邻波段的相关性越小，则其独立性越强。因此，定义信息相邻相关系数 r_{iac} ：首先，计算全波段的标准差 σ ，并根据 σ 的大小对波段进行排序；然后，计算新顺序下的 r_{ac} 即为该波段的 r_{iac} ：

$$r_{iac}(i) = (r_{cl}^{sort\sigma}(k-1, k) + r_{cl}^{sort\sigma}(k, k+1))/2 \quad (10)$$

式中： r_{iac} 为第 i 波段(波段顺序)的信息相邻相关系数； k 为按照波段信息量排序(sort)后的第 i 波段(波



(a) 实验 1

(a) Test 1

段顺序)对应的位置; $k-1$ 和 $k+1$ 分别为第 i 波段(波段顺序)在新序列中的前后波段号。那么,定义信息相邻相关结合的 MABS(modified ABS):

$$\text{MABS} = \frac{\sigma_i}{r_{\text{iac}}(i)} \quad (11)$$

在计算量方面,MABS 仅比 ABS 多了一次排序;在满足算法准则方面,MABS 能更好的满足信息量度量和独立性度量的要求。

2.3 实证分析

2.3.1 波段选择结果

表 1 显示为两组实验的波段选择结果(依次前 40 个波段),其中, σ 为标准差排序;ABS(r_i)和 ABS(r_d) 分别为基于 r_i 和 r_d 的 ABS 选择结果;MABS(r_i)和 MABS(r_d) 分别为基于 r_i 和 r_d 的 MABS 选择结果。可见,ABS(r_i)和 ABS(r_d)的结果相同,ABS 与 σ 选择的结果近乎相同(实验 1:除 6 个波段对调外,前

40 个波段完全对应相同,全波段(191)中仅有 16 个波段未对应相同,且大部分位于前 40 以外;实验 2:前 40 波段号完全对应相同,全波段(800)仅有 46 个波段未对应相同,且大部分位于前 200 以外),故而,基于 r_i 和 r_d 的波段相邻相关性均无法有效表达出波段的独立性。而 MABS(r_i)与 MABS(r_d)的波段对应差别较大(实验 1:前 40 波段中有 28 个不同,全波段中有 82 个不同;实验 2:前 200 波段有 187 个不同,全波段中 703 个不对应),对比 ABS 可见,改进的综合相关系数对 ABS 无作用,而对 MABS 有着显著的影响。此外,MABS(r_i)和 MABS(r_d)分别与 σ 的结果差别极大(实验 1:前者在前 40 波段完全不对应,在全波段中 129 个不对应;后者在前 40 波段仅有一个对应,在全波段中 116 个不对应。实验 2:两者的前 200 波段与 σ 均不相同,全波段中 703 个不对应),对比 ABS 可见,在表达波段独立性方面,MABS 对相

表 1 波段选择结果

Tab.1 Band selection results

		Former 40 bands	
		Test 1	Test 2
σ		62, 43, 63, 48, 47, 44, 51, 50, 45, 42, 46, 59, 61, 49, 40, 52, 41, 58, 68, 67, 69, 39, 64, 37, 36, 38, 35, 66, 55, 34, 54, 70, 53, 33, 32, 31, 65, 57, 28, 27	375, 374, 373, 376, 372, 371, 377, 380, 382, 379, 381, 370, 378, 383, 384, 369, 368, 385, 367, 386, 366, 387, 365, 388, 364, 389, 363, 390, 362, 391, 361, 392, 393, 360, 352, 351, 353, 394, 350, 354
	ABS(r_i)	62, 43, 63, 48, 47, 44, 51, 50, 45, 42, 46, 59, 61, 49, 40, 52, 58, 41, 68, 67, 69, 39, 64, 37, 36, 38, 35, 55, 54, 66, 34, 70, 53, 33, 32, 31, 65, 57, 28, 27	375, 374, 373, 376, 372, 371, 377, 380, 382, 379, 381, 370, 378, 383, 384, 369, 368, 385, 367, 386, 366, 387, 365, 388, 364, 389, 363, 390, 362, 391, 361, 392, 393, 360, 352, 351, 353, 394, 350, 354
ABS(r_d)	62, 43, 63, 48, 47, 44, 51, 50, 45, 42, 46, 59, 61, 49, 40, 52, 58, 41, 68, 67, 69, 39, 64, 37, 36, 38, 35, 55, 54, 66, 34, 70, 53, 33, 32, 31, 65, 57, 28, 27	375, 374, 373, 376, 372, 371, 377, 380, 382, 379, 381, 370, 378, 383, 384, 369, 368, 385, 367, 386, 366, 387, 365, 388, 364, 389, 363, 390, 362, 391, 361, 392, 393, 360, 352, 351, 353, 394, 350, 354	
MABS(r_i)	83, 84, 14, 77, 72, 62, 63, 43, 39, 13, 64, 70, 76, 12, 66, 11, 48, 49, 61, 69, 10, 90, 46, 59, 53, 35, 37, 54, 58, 65, 41, 81, 29, 22, 31, 82, 55, 60, 26, 71	144, 492, 491, 101, 100, 495, 516, 99, 482, 98, 517, 485, 97, 96, 519, 95, 520, 522, 521, 87, 481, 477, 86, 479, 170, 526, 85, 474, 84, 145, 532, 83, 82, 81, 533, 80, 534, 217, 79, 164	
MABS(r_d)	83, 84, 62, 14, 63, 43, 72, 77, 39, 13, 64, 76, 70, 12, 48, 66, 11, 49, 61, 46, 69, 59, 53, 10, 90, 35, 37, 41, 58, 54, 65, 31, 81, 22, 26, 29, 60, 82, 71, 21	144, 492, 100, 491, 101, 516, 99, 98, 517, 97, 522, 519, 96, 95, 520, 521, 87, 495, 86, 526, 85, 84, 83, 532, 482, 81, 533, 80, 534, 485, 82, 79, 535, 78, 536, 477, 77, 170, 145, 481	

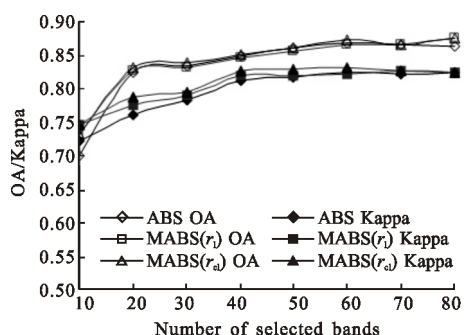
关系数的响应更明显和有效。

2.3.2 解译结果

为了验证 MABS 选择结果的有效性,分别对实验 1(依次选择排序结果的前 10、20、30、40、50、60、70、80 等 8 种波段组合,选取建筑)、道路、阴影、草

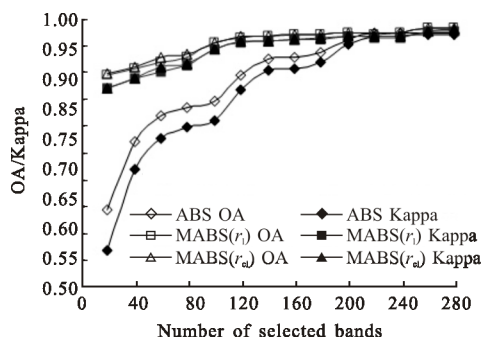
地、树木和小径等 6 类地物)和实验 2(依次选择排序结果的前 20、40、60、80、100、120、140、160、180、200、220、240、260、280 等 14 种波段组合,选取建筑、植被、阴影、柏油路面、标识线 1、标识线 2 和水泥路面等 7 类地物)进行了监督分类(马哈利诺贝斯距离分

类 MDC(mahalanobis distance classification), 并计算统计了各波段组合维数下的 OA 和 Kappa 系数(图4)。



(a) 实验 1

(a) Test 1



(b) 实验 2

(b) Test 2

图 4 不同波段组合分类结果的总体精度和 Kappa 系数

Fig.4 OA and Kappa coefficient of classification with different band combinations

图 4 中,ABS(r₁)、ABS(r_{cd})和 σ 算法在两个实验中的诸多波段组合中,选取的波段均是一样的,故统称为 ABS。

可见,实验 1(图 4(a))中,ABS、MABS(r₁)和 MABS(r_{cd})三种算法的 OA 和 Kappa 系数均有相对稳定的略微差别,总体而言,三者 OA 均值大小依次为 MABS(r_{cd})>MABS(r₁)>ABS(8 个组合:84.15%>83.75%>83.32%;前 4 个组合:81.37%>80.93%>80.22%);三者 Kappa 系数均值关系为 MABS(r_{cd})>MABS(r₁)>ABS(8 个组合:0.808 1>0.802 8>0.795 9;前 4 个组合:0.788 9>0.783 0>0.769 9)。当波段维数达到 60 后,OA 和 Kappa 系数不再随波段增加而出现明显变化,此时,MABS(r_{cd}) OA 比 MABS(r₁)高 0.53%,比 ABS 高 0.90%。图 5 为实验 1 前 60 波段组合的分类效

果,MABS(r_{cd})OA(87.43%)高于 MABS(r₁)(86.52%)和 ABS(86.37%)。整体上,树木、草地和小径的分类效果较好;道路多位于建筑物之间而与其发生部分混分;而阴影区则受建筑物的直接影响,多毗邻建筑物。

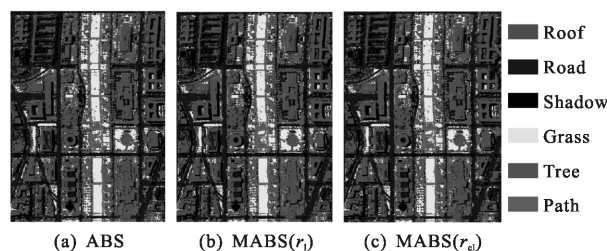


图 5 实验 1 的分类结果(前 60 波段)

Fig.5 Classification results of test 1(selected 60 bands)

实验 2(图 4(b))中,MABS(r₁)和 MABS(r_{cd})两种算法均明显优于 ABS,MABS(r_{cd})略优于 MABS(r₁),总体而言,三者 OA 均值的排序为 MABS(r_{cd})>MABS(r₁)>ABS(14 个组合:95.59%>95.41%>88.95%;前 7 个组合:93.69%>93.39%>81.94%);三者 Kappa 系数均值关系为 MABS(r_{cd})>MABS(r₁)>ABS(14 个组合:0.944 5>0.942 7>0.864 2;前 7 个组合:0.920 6>0.917 9>0.778 0)。当波段维数达到 120、140、160 时,MABS(r₁)和 MABS(r_{cd})两种算法的 OA 和 Kappa 系数均具有较高的稳定值;而 ABS 则在 220 波段之后才达到与前两者在 120 波段相当的 OA 和 Kappa 系数。图 6 为实验 2 前 160 波段组合的分类效果,MABS(r₁)和 MABS(r_{cd})的分类结果明显优于 ABS,在 OA 上前两者分别比后者(93.63%)高 3.41%和 3.44%,在 Kappa 系数上前两者比后者(0.9194)大 0.042 9 和 0.043 4。从分类结果可见,ABS 对两类标识线的识别度和区分度均较差,建筑物和水泥路发生了部分混分情况,这些问题在 MABS(r₁)和 MABS(r_{cd})算法中得到较好解决。

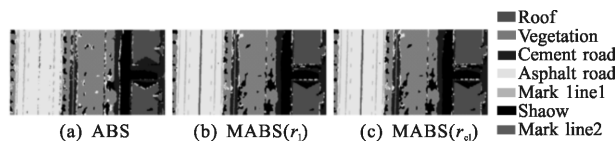


图 6 实验 2 的分类结果(前 160 波段)

Fig.6 Classification results of test 2(selected 160 bands)

3 结果分析

通过第三方公共数据集和课题组采集的两组不

同分辨率的高光谱数据进行实验, 该研究验证了非线性函数变换改进后的谱间综合相关系数(r_{ci})的有效性; 两组实验的谱间相关性类型中, 非线性类型系数占比均大于总体的三分之二(图 1、2)。

该研究将 r_{ci} 应用于 ABS 算法, 发现 ABS(r_{ci}) 的波段选择结果与 ABS(r_l) 无差别, 进而发现 ABS 算法对两组数据的波段选择结果与其波段信息(σ) 排序具有显著的一致性(表 1)。因此, 研究进行了波段相邻相关系数 r_{ac} 与 r_{mean} 、 r_{max} 、 r_{min} 等三种指数的比较, 发现 r_{ac} 与 r_{mean} 并无统计意义上的相关性, 反而与 r_{max} 有着极为明显的关系(图 3); 从相关系数矩阵(图 1、2)亦可直观看出 r_{ac} 与 r_{max} 的显著联系。那么, ABS 算法通过 r_{ac} 来表达波段的独立性, 其代表性较差。

基于 r_{ac} 的不足, 研究提出一种信息量相邻相关系数 r_{iac} , 实验发现, 基于 r_{iac} 的 MABS 能够较好地兼顾信息量和独立性原则, 所选波段组合的分类 OA 和 Kappa 系数优于基于 r_{ac} 的 ABS。从波段选择结果的波长范围可见, 实验 1 选取的前 40 波段中, ABS 和 σ 结果均位于 503~804 nm 之间(绿、黄、橙、红光和近红外), 而 MABS 结果遍及 432~1 175 nm(除紫光外的所有可见光谱和近红外); 实验 2 的前 40 波段中, ABS 和 σ 结果均位于 636~669 nm 之间(红光), 且近乎完全连续, 而 MABS 结果遍及 442~775 nm(除紫光外的所有可见光谱和近红外), MABS(r_{ci}) 结果范围略大于 MABS(r_l)。从分类结果曲线可见, MABS 的 OA 和 Kappa 系数的起点值较大, 随波段组合维数的增加而呈现稳定增大, 直到平稳状态; 而 ABS 的起点值较小, 实验 2 中的结果曲线在 100~120 和 180~200 位置突增, 研究发现, 这两个区间波段的加入扩展了波段组合的波谱范围, 进而为波段组合大大提高了信息量和类间区分度, 这也说明, ABS 的波段选择有效性低于 MABS。

4 结 论

通过对波段信息进行非线性函数变换, Pearson 综合相关系数可以表达出波谱间的非线性相关信息, 且实验证明非线性是波谱间相关性的主要类型。ABS 算法中的波段相邻相关系数(r_{ac}) 不能有效表达出波段与其他波段之间的整体相关性, 因此, r_{ac} 对波段独立性的表征缺乏有效性。而该研究提出的信息

相邻相关系数(r_{iac}) 和基于此指数改进的自适应波段选择算法(MABS) 则较好地兼顾了“大信息量”和“强独立性”的非监督算法原则, 与 ABS 相比, MABS 具有更高的有效性。实验结果也佐证了非线性变换后的 Pearson 综合相关系数(r_{ci}) 优于线性相关系数(r_l)。后期将对研究 MABS 与其他波段选择算法的优劣, 探讨 r_{ci} 对其他降维算法的影响。

参 考 文 献:

- [1] Zhao Huijie, Li Mingkang, Li Na, et al. A band selection method based on improved subspace partition [J]. *Infrared and Laser Engineering*, 2015, 44 (10): 3155–3160. (in Chinese)
赵慧洁, 李明康, 李娜, 等. 一种基于改进子空间划分的波段选择方法[J]. 红外与激光工程, 2015, 44(10): 3155–3160.
- [2] Wang Yiting, Huang Shiqi, Liu Daizhi, et al. Novel band selection method based on target detection [J]. *Infrared and Laser Engineering*, 2013, 42(8): 2294–2298. (in Chinese)
王艺婷, 黄世奇, 刘代志, 等. 一种新的基于目标检测的波段选择方法[J]. 红外与激光工程, 2013, 42(8): 2294–2298.
- [3] Fan Xueli, Feng Haihong, Yuan Meng. PCA based on mutual information for feature selection [J]. *Control and Decision*, 2013, 28(6): 915–919. (in Chinese)
范雪莉, 冯海泓, 原猛. 基于互信息的主成分分析特征选择算法[J]. 控制与决策, 2013, 28(6): 915–919.
- [4] Martinez-Uso A, Pla F, Sotoca J M, et al. Clustering-based hyperspectral band selection using information measures [J]. *IEEE Transactions on Geoscience & Remote Sensing*, 2008, 45(12): 4158–4171.
- [5] Zhao Chunhui, Chen Wanhai, Yang Lei. Research advances and analysis of hyperspectral remote sensing image band selection [J]. *Journal of Natural Science of Heilongjiang University*, 2007, 24(5): 592–602. (in Chinese)
赵春晖, 陈万海, 杨雷. 高光谱遥感图像最优波段选择方法的研究进展与分析[J]. 黑龙江大学自然科学学报, 2007, 24(5): 592–602.
- [6] Qin Fangpu, Zhang Aiwu, Wang Shumin, et al. Hyperspectral band selection based on spectral clustering and inter-class separability factor [J]. *Spectroscopy and Spectral Analysis*, 2015, 35(5): 1357–1364. (in Chinese)
秦方普, 张爱武, 王书民, 等. 基于谱聚类与类间可分性因子的高光谱波段选择[J]. 光谱学与光谱分析, 2015, 35(5): 1357–1364.

- [7] Bajwa S G, Bajcsy P, Groves P, et al. Hyperspectral image data mining for band selection in agricultural applications[J]. *Transactions-American Society of Agricultural Engineers*, 2004, 47(3): 895-908.
- [8] Serpico S B, Bruzzone L. A new search algorithm for feature selection in hyperspectral remote sensing images [J]. *IEEE Transactions on Geoscience & Remote Sensing*, 2001, 39(7):1360-1367.
- [9] Wei Fangjie. Research on band selection for hyperspectral imagery [D]. Harbin: Harbin Engineering University, 2013. (in Chinese)
魏芳洁. 高光谱图像波段选择方法的研究[D]. 哈尔滨: 哈尔滨工程大学, 2013.
- [10] Liu Chunhong, Zhao Chunhui, Zhang Lingyan. A new method of hyperspectral remote sensing image dimensional reduction [J]. *Journal of Image and Graphics*, 2005, 21(6): 676-680. (in Chinese)
刘春红, 赵春晖, 张凌雁. 一种新的高光谱遥感图像降维方法[J]. 中国图象图形学报, 2005, 21(6): 676-680.
- [11] Fan Xiubin, Li Shiqu. Relationship between the mutual information and the correlation [J]. *Journal of Information Engineering University*, 1999, 18(2): 45-48. (in Chinese)
范修斌, 李世取. 互信息与相关性之间的关系 [J]. 信息工程学院学报, 1999, 18(2): 45-48.