

基于互信息波段选择和经验模态分解的高精度高光谱数据分类

沈毅 张敏 张淼

(哈尔滨工业大学控制科学与工程系, 黑龙江 哈尔滨 150001)

摘要 在遥感数据处理研究中, 高维高光谱数据的冗余信息和噪声严重影响高光谱数据的分类精度, 针对此问题提出基于互信息波段选择和经验模态分解的高精度高光谱数据分类算法(MI-EMD-SVM)。分别采用基于互信息波段选择方法和经验模态分解实现对高光谱数据的冗余信息处理和特征提取, 并获得处理后的高光谱数据 X'' 。采用支持向量机分类算法对处理后的高光谱数据 X'' 进行分类实验。仿真实验结果证实 MI-EMD-SVM 算法不仅提高高光谱数据分类精度, 同时还减少支持向量数目, 提高高光谱数据分类速度。

关键词 图像处理; 高光谱数据; 分类; 互信息; 波段选择; 经验模态分解; 分类精度

中图分类号 TP751 **文献标识码** A **doi**: 10.3788/LOP48.091001

Mutual Information Bands Selection and Empirical Mode Decomposition Based Support Vector Machines for Hyperspectral Data High-Accuracy Classification

Shen Yi Zhang Min Zhang Miao

(Department of Control Science and Engineering, Harbin Institute of Technology, Harbin, Heilongjiang 150001, China)

Abstract In remote-sensing data processing research, redundant information and noise of high-dimensional hyperspectral data affect the classification accuracy of hyperspectral data seriously. To solve this problem, we propose an algorithm of hyperspectral data classification based on band selection with mutual information and empirical mode decomposition (MI-EMD-SVM). Band selection based on mutual information is used to achieve redundant information processing, and empirical mode decomposition (EMD) is used to achieve feature extraction. And the obtained hyperspectral data X'' has been processed. The support vector machines (SVM) classification of the data is classified, which has been processed. Experimental results of the AVIRIS data indicate that the proposed approach improves the classification accuracy of hyperspectral data, significantly reduces the number of support vector, and improves the speed of hyperspectral data classification.

Key words image processing; hyperspectral data; classification; mutual information; band selection; empirical mode decomposition; classification accuracy

OCIS codes 100.5010; 110.4234

1 引言

在遥感图像处理研究中, 高光谱传感器通过同时采样数百个高分辨率、连续光谱波段图像来观测地表。例如, AVIRIS 高光谱传感器具有 224 个光谱波段, 即从 $0.4 \mu\text{m}$ 到 $2.5 \mu\text{m}$ 每隔 10 nm 采样一次。高光谱传感器这一特点决定了高光谱数据是一个富含冗余信息的高维数据, 这影响了高光谱数据分类精度和分类速

收稿日期: 2011-02-16; **收到修改稿日期**: 2011-04-13; **网络出版日期**: 2011-07-25

基金项目: 国家自然科学基金(60975009)资助课题。

作者简介: 沈毅(1965—), 男, 博士, 教授, 主要从事控制系统故障诊断、飞行器控制及超声成像技术等方面的研究。

E-mail: shen@hit.edu.cn

度^[1]。另外,虽然高光谱传感器数据信噪比高,但是高光谱数据中信息的有效性在很大程度上还受到数据中噪声的影响,这严重影响了高光谱数据的分类^[2~4]。因此研究高光谱数据的冗余信息处理和高光谱数据特征提取是十分必要的。

高光谱数据冗余信息处理的常用方法是波段选择。波段选择是指在一定准则下,挑选出那些能够代表原始数据信息波段光谱图像,约减掉那些影响分类性能和效率的冗余特征。波段选择的早期准则有 Bayes 误差^[5]、Bhattacharyya 距离^[6]、Chernoff 距离^[7]、Jeffreys-Matusita 距离^[8]等。这些准则都是基于两分类问题,并且输入数据满足高斯分布。对于非高斯分布、多分类问题的高光谱数据,这些准则很难作为波段选择的准则。对此,Battiti 等^[9,10]提出了基于互信息(MI)的波段选择。互信息无需假设数据服从任何分布,因而在高光谱数据波段选择上面具有明显的潜力和优势。

经验模态分解(EMD)是在 1998 年由 N. E. Huang^[11]提出的一种处理非平稳、非线性信号的有效算法,它是将信号分解成有限个内固模态函数(IMF)和一个残差信号。经验模态分解(EMD)相对小波变换,有更好的时频特性,能够自适应地提取非线性非平稳信号的本质特征。对于高光谱数据的特征提取,经验模态分解具有明显的潜力和优势。

本文结合互信息波段选择和经验模态分解的优点,提出基于互信息波段选择和经验模态分解的高精度高光谱数据分类算法。对输入高光谱数据 X 进行互信息波段选择,保留那些与输出类别标定相关性大的波段光谱图像 X' ,降低高光谱数据维数,减少高光谱数据的冗余信息。对已选波段数据 X' 进行 EMD 分解,保留含有高光谱数据本质特征的内固模态函数,完成对高光谱数据特征提取。采用富含高光谱数据本质特征的内固模态函数重构高光谱数据 X'' ,并对 X'' 进行支持向量机(SVM)分类。仿真实验结果证明该方法能够获得高精度的分类结果。

2 基于互信息波段选择方法

2.1 互信息的物理意义

互信息是信息论中一个用来测量两个随机变量间的统计相关性的数值量,可表示为^[12]

$$I(x, y) = \iint p(x, y) \ln \frac{p(x, y)}{p(x)p(y)} dx dy, \quad (1)$$

式中 $p(x, y)$ 是连续随机变量 x 和 y 的联合概率密度函数, $p(x)$ 和 $p(y)$ 分别是 x 和 y 边际概率密度函数。

对于给定的高光谱两个波段光谱图像 A 和 B ,假设图像 A 和 B 对应的元素分别为变量 a 和变量 b , $p_i(a)$ 和 $p_j(b)$ 分别是变量 a 和变量 b 对应的第 i 个状态和第 j 个状态的概率密度, $p_{ij}(a, b)$ 为联合概率密度。则由互信息定义得到图像 A 和 B 的互信息为

$$I(A, B) = \sum_{i=1}^{L_a} \sum_{j=1}^{L_b} p_{ij}(a, b) \ln \frac{p_{ij}(a, b)}{p_i(a)p_j(b)}. \quad (2)$$

信息论中熵描述信号的信息量,信号的信息量越丰富时,它对应的熵就越大。假设图像 A ,变量 a 为图像 A 里的元素,则图像 A 的熵定义为

$$H(X) = - \sum_x p(x) \ln p(x), \quad (3)$$

由(2)式和(3)式可得,采用信息熵描述的互信息为

$$I(A, B) = H(A) + H(B) - H(A, B) = H(A) - H(A|B) = H(B) - H(B|A), \quad (4)$$

式中 $H(A)$ 和 $H(B)$ 分别是图像 A 和图像 B 的熵; $H(A, B)$ 是图像 A 和图像 B 的联合熵; $H(A|B)$ 和 $H(B|A)$ 分别为给定图像 B 时图像 A 的条件熵和给定图像 A 时图像 B 的条件熵。

由(4)式可以看出互信息实质上是在给定某一随机变量情况下另一变量不确定性(熵)的减少,比如给定图像 A 情况下,图像 B 的不确定性的减少值。在高光谱数据波段选择中,可以通过比较各个波段高光谱数据与输出类别标定 C 的互信息,保留互信息大的波段高光谱数据,进而实现对高光谱数据的降维、减少高光谱数据的冗余信息,并提高高光谱数据的可分性。

2.2 基于互信息的波段选择

定义 1[基于互信息波段选择准则 $J(X)$]:根据最大相关性(max-relevance),选出在给定输出类别标定 C 时不确定性减少大的波段图像,可表示为^[13]

$$J(X) = \max_{X' \subset X} \frac{1}{|X'|} \sum_{x_i \in X'} I(x_i; C), \quad (5)$$

式中 $I(x_i; C)$ 是波段图像 x_i 与输出类别标定 C 的互信息, X' 是已选波段数据。

对于 N 维原始高光谱数据集合 X ,波段选择就是根据互信息波段选择准则,寻找高光谱数据集合 X' ,维数为 $M(M \leq N)$,使得 X' 与输出类别标定 C 的互信息指标达到最大^[14]。此时的最佳选择波段集合记为 X_0 ,即

$$X_0 = X', \quad (6)$$

3 基于经验模态分解的特征提取

经验模态分解是一种适用于非线性非平稳信号的处理方法,它能将信号分解成有限个内固模态函数(IMF)和一个残差之和,具有完全自适应性。通过 EMD 方法可以将任意信号分解为从高频到低频的具有物理可实现性的 IMF 信号的叠加。对于非线性非平稳的高光谱数据,EMD 算法能够自适应地提取出高光谱数据中的本质属性。

3.1 高光谱数据的非线性非平稳性

图 1 是机载可见光/红外成像光谱仪(AVIRIS)传感器的 92AV3C 数据集合中的 hay-windrowed 和 bldg-grass-tree-drivers 两种地物的光谱反射曲线。从光谱的响应曲线可以看出,同类地物的光谱响应曲线有明显的差异,具有非平稳性,而且响应值不跟波段序列成正比,具有非线性。所以高光谱数据是一个典型的非线性非平稳信号。

3.2 内固模态函数

满足下面两个条件的信号称为内固模态函数:1)极值点和过零点的数目必须相等或者至多只相差 1。2)在任意时刻,由极大值点定义的上包络线和由极小值点定义的下包络线的平均值为零,也就是说信号的上下包络线对称于时间轴。

3.3 经验模态分解特征提取

经验模态分解本质是借着不断重复的筛选程序来逐步找出内固模态函数 IMF,将原信号分解成各个 IMF。以原始图像 X 作为输入信号,假设 $x(m, n)$ ($m = 1, 2, \dots, M; n = 1, 2, \dots, N$) 是 X 在像素点 (m, n) 的值,筛选程序的过程如下:

1) 找出 X 中的所有局部极大值以及局部极小值,并利用样条插值算法,分别将局部极大值拟合成上包络线 $e_{\max}(m, n)$ 、局部极小值拟合成下包络线 $e_{\min}(m, n)$ 。

2) 求出上下包络线的均值,得到均值包络线 $m_1(m, n)$

$$m_1(m, n) = \frac{e_{\max}(m, n) + e_{\min}(m, n)}{2}, \quad (7)$$

3) 原始信号 $x(m, n)$ 与均值包络线 $m_1(m, n)$ 相减,得到第一个分量 $h_1(m, n)$

$$h_1(m, n) = x(m, n) - m_1(m, n), \quad (8)$$

4) 检查 $h_1(m, n)$ 是否符合 IMF 的条件。如果不符合,则回到步骤 1 并且 $h_1(m, n)$ 将当作原始信号,进行第二次的筛选,即

$$h_2(m, n) = h_1(m, n) - m_2(m, n), \quad (9)$$

重复筛选 k 次

$$h_k(m, n) = h_{k-1}(m, n) - m_k(m, n), \quad (10)$$

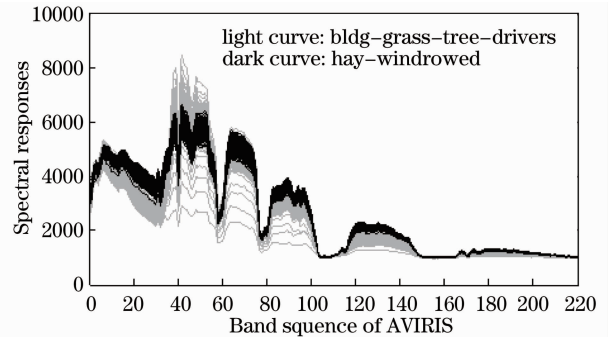


图 1 92AV3C 数据集合中两类地物的光谱响应曲线

Fig. 1 Spectral responses of two classes in 92AV3C

直到 $h_k(m, n)$ 符合 IMF 的条件, 即得到第一个 IMF 分量 $c_1(m, n)$

$$f_{\text{IMF1}} = c_1(m, n) = h_k(m, n), \quad (11)$$

5) 原始信号 $x(m, n)$ 减去 $c_1(m, n)$ 可得到剩余量 $r_1(m, n)$

$$r_1(m, n) = x(m, n) - c_1(m, n), \quad (12)$$

6) 将 $r_1(m, n)$ 当作新的信号, 重新执行步骤 1) 至步骤 5), 得到第二个 IMF 分量 $c_2(m, n)$ 和新的剩余量 $r_2(m, n)$ 。如此重复 q 次

$$\begin{aligned} f_{\text{IMF2}} &= c_2(m, n) \\ r_2(m, n) &= r_1(m, n) - c_2(m, n) \\ f_{\text{IMF3}} &= c_3(m, n) \\ r_3(m, n) &= r_2(m, n) - c_3(m, n) \\ &\vdots \\ f_{\text{IMF}q} &= c_q(m, n) \\ r_q(m, n) &= r_{q-1}(m, n) - c_q(m, n) \end{aligned}$$

当第 q 个剩余量 $r_q(m, n)$ 满足单调时, 整个 EMD 的分解过程完成。原始信号 $x(m, n)$ 可以表示成 q 个 IMF 分量与一个平均趋势分量(残差) $r_q(m, n)$ 的组合, 即

$$x(m, n) = \sum_{k=1}^q c_k(m, n) + r_q(m, n), \quad (13)$$

内固模态函数的本质是信号在不同频率范围内原信号的本质特征, 得到的残差就是原信号噪声的集合^[15]。可以通过内固模态函数的选择和残差的去除, 实现对信号的特征提取。

4 高光谱数据分类算法

互信息实质上是在给定某一随机变量情况下另一变量不确定性(熵)的减少程度。在高光谱数据预处理中, 通过比较高光谱各个波段与输出类别标定 C 的互信息, 保留互信息大的波段图像, 从而降低高光谱数据的维数, 提高高光谱数据的可分性。对于高光谱数据, EMD 算法能够自适应地将高光谱数据分解为富含高光谱数据本质特征的内固模态函数和富含噪声的残差^[15], 进而通过内固模态函数的选择和残差的去除实现对高光谱数据的特征提取。

高光谱数据的互信息波段选择算法如图 2 所示。首先是初始化, 设置高光谱数据波段选择子集 X' 的期望维数(波段数量) D , 并将波段选择子集 X' 设为空集, 波段选择子集 X' 的波段数量 $d = 0$ 。然后计算出高光谱各个波段图像与输出类别标定 C 的互信息 $I(x_i; C)$, 根据最大相关性准则把相应的波段 x_i 从 X 中移出并加入波段选择子集 X' , $d = d + 1$ 。接着判断 d 是否等于期望波段数量 D , 如果等于则波段选择结束, 最终的波段子集为 X' , 否则重复第三步, 继续从 X 中选出一个满足最大相关性的波段图像, 并将它从 X 中移出并加入波段选择子集 X' , 直到波段选择子集的数量等于期望波段数量 D 时, 波段选择结束。

波段选择之后, 对高光谱数据子集 X' 进行 EMD 分解, 算法如图 3 所示。首先求输入信号 X' 的局部极值, 并采用三次样条求出上包络、下包络和均值包络。然后把输入信号减去均值包络, 获得第一个分量 $h_1(m, n)$, 判断 $h_1(m, n)$ 是否满足 IMF 条件。如果满足则 $h_1(m, n)$ 是第一个 IMF, 如果不满足则把 $h_1(m, n)$ 作为输入进行第二次筛选, 直到 $h_k(m, n)$ 满足 IMF 条件, 则获得第一个 IMF, 并求得余量 $r_1(m, n)$ 。判断余量是否是单调函数, 如果余量是单调函数则 EMD 分解结

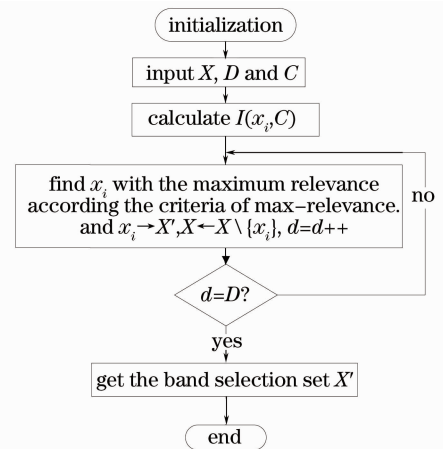


图 2 基于互信息的波段选择流程图
Fig. 2 Flow chart of bands selection based on mutual information

束,如果不是则把余量作为输入,继续求第二个 IMF 和第二个余量,直到余量是单调时,保存所有 IMF,EMD 分解结束。

EMD 分解获得内固模态函数 IMF1, IMF2, ..., IMF_n。选择内固模态函数构成预处理之后的高光谱数据 X'' 。最后,采用支持向量机分类算法(SVM)对 X'' 进行分类。

5 仿真实验结果与分析

实验所用的数据是 AVIRIS 传感器于 1992 年在美国印地安那州西北部所采集的高光谱数据中的一个片段。采用该数据集的优势是它附带了通过实地测绘而得到的地物真实参考图,利用该图可检验分类精度,也可作为波段选择的输出类别标定 C 。虽然 AVIRIS 传感器收集了 224 个波段数据,但 4 个波段都是 0 值而被丢弃,剩下的 220 个波段构成了高光谱数据集 92AV3C^[16,17]。采用最大相关性指标对高光谱数据 X 进行互信息波段选择,获得与 C 最大相关的高光谱数据子集 X' ;对波段选择出的图像 X' 进行经验模态分解,实现数据的特征提取;最后对处理后的数据进行 SVM 分类。SVM 的核函数采用径向基函数(RBF)

$$K(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right) \quad (14)$$

式中 σ 是核参数, c 是惩罚因子。

对所有 16 类像素的地物作为实验样本(表 1),总的训练样本和测试样本数分别为 5183 和 5183,进行 5 折分类。将得到的分类结果与测试集所包含的真实类别相比较,可以得出各类别的分类精度;将所有类别的类别精度平均得到平均精度(AA);将所有正确分类的像素数除以总像素数则得到总体精度(OA)。

表 1 各类地物所对应的训练样本数和测试样本数

Table 1 Numbers of training and testing pixels in each class

Class	Training sample	Testing sample
Alfalfa	27	27
Corn-no till	717	717
Corn-min till	417	417
Corn	117	117
Grass/Pasture	249	248
Grass/Trees	373	374
Grass/Pasture-mowed	13	13
Hay-windrowed	245	244
Oats	10	10
Soybeans-no till	484	484
Soybean-min till	1234	1234
Soybean-clean till	307	307
Wheat	106	106
Woods	647	647
Bldg-Grass-Tree-Drivers	190	190
Stone-steel towers	47	48
Total	5183	5183

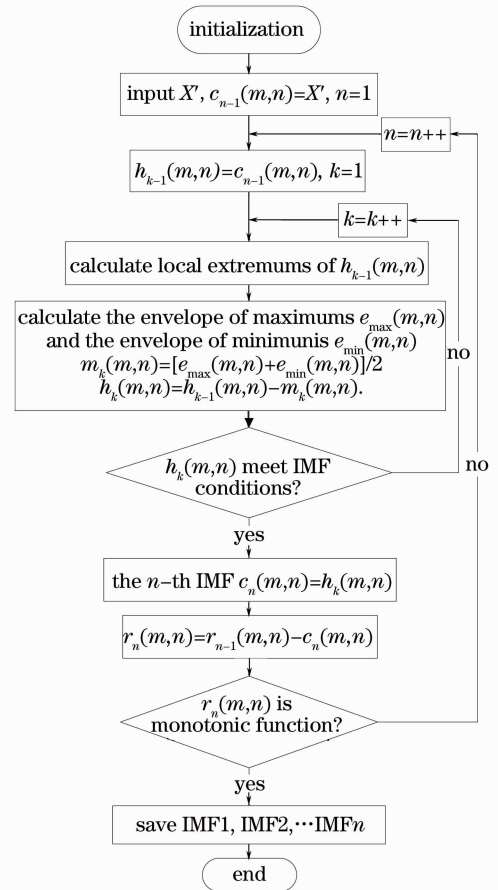


图 3 高光谱数据的经验模态分解的流程图

Fig. 3 Flow chart of empirical mode decomposition for hyperspectral data

5.1 SVM 核函数参数选取

对高光谱数据集 92AV3C 进行不同核函数参数下的 SVM 分类, SVM 的核函数采用 RBF, 且 RBF 惩罚因子 c 的取值范围为 $0.1 \sim 1000$, RBF 核参数 σ 的取值范围为 $0.2 \sim 2$ 。通过对不同 RBF 参数下的 SVM 分类性能的比较, 选择最合适的 RBF 参数。这里选取分类的平均精度作为分类性能评价指标, 如表 2 所示。从实验结果可以看出, 最佳参数配置仍处在 $\sigma \in (0.2, 1)$ 及 $c \in (10, 100)$ 这一广泛认同的典型区间。进一步实验确认最佳参数为 $\sigma = 0.4, c = 60$ 。

表 2 SVM 在不同 RBF 参数下的平均精度

Table 2 Average accuracies of SVM in different parameters

Kernel parameter	Penalty coefficient c					
	0.1	1	10	100	1000	
Parameter σ	0.2	30.9024	68.9628	77.6524	77.6138	77.6138
	0.4	37.8538	73.0270	85.6678	85.9757	85.9757
	0.6	36.1079	68.8580	84.7857	85.4527	85.4165
	0.8	34.2665	62.9232	84.2792	85.4114	85.3829
	1	29.1405	57.8920	81.8855	85.1203	85.0937
	2	27.9695	41.4004	69.5650	84.9994	84.4304

5.2 高光谱数据分类实验

分别采用 SVM、EMD-SVM^[16] 和本文提出的 MI-EMD-SVM 分类算法对高光谱数据进行分类。在 MI-EMD-SVM 分类中, 利用 AVIRIS 数据中真实的地物参考图作为波段选择的输出类别标定 C , 分别进行了 5, 10, 15, \dots , 220 波段数量的波段选择, 即波段数量 $D = 5N, N = 1, 2, \dots, 44$; 然后再对波段选择出的高光谱数据进行经验模态分解和 SVM 分类, 得到了不同波段数量的高光谱数据分类平均精度和总体精度。把 MI-EMD-SVM 算法的分类精度与 EMD-SVM、SVM 算法的分类精度进行比较, 如图 4 和图 5 所示。

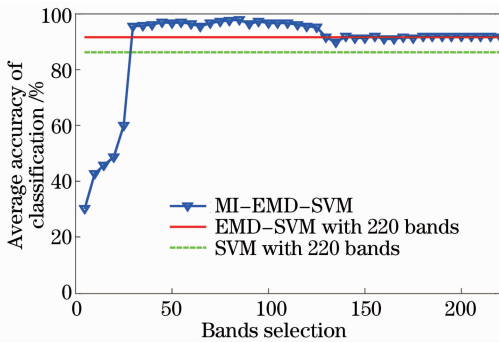


图 4 MI-EMD-SVM 的 16 类地物平均精度

Fig. 4 16-class classification average accuracy based on MI-EMD-SVM

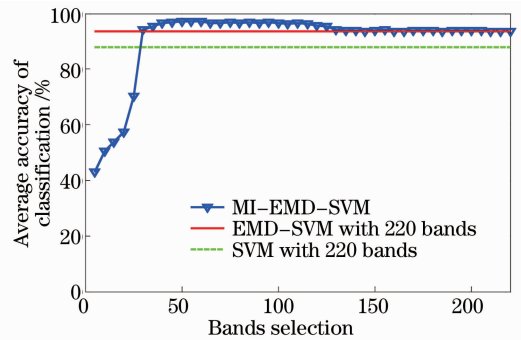


图 5 MI-EMD-SVM 的 16 类地物总体精度

Fig. 5 16-class classification overall accuracy based on MI-EMD-SVM

从图 4 的 16 类地物平均精度可以看出, 在 MI-EMD-SVM 算法分类实验中, 在波段数量少于 30 时, 由于选择的波段过少, 导致数据信息的不全, 因此 MI-EMD-SVM 算法的平均精度比较低; 当基于互信息波段选择数量取 30 开始, 所获分类平均精度要高于 EMD-SVM 算法和 SVM 算法; 而且在波段选择数量取 40~130 时, MI-EMD-SVM 算法的平均精度优势更是明显。从图 5 的 16 类地物整体精度可以看出, 在波段数量少于 30 时, 由于选择的波段过少, 导致数据信息的不全, 因此 MI-EMD-SVM 算法的整体精度比较低; 当 MI-EMD-SVM 算法的波段选择数量取 30 时, 所获得的整体精度都高于 SVM 算法和 EMD-SVM 算法; 而且在波段选择数量取 40~130 时, MI-EMD-SVM 算法的整体精度优势更是明显。

为了更全面分析 MI-EMD-SVM 算法, 根据波段少精度高的原则, 参照图 4 和图 5 的分类精度, 选取了波段选择数量为 45 的 MI-EMD-SVM 的分类算法进行全面实验分析。在分类精度、支持向量数量、数据处理时间等方面, 将 MI-EMD-SVM 分类算法与 EMD-SVM、SVM 方法进行比较, 如图 6~8 和表 3 所示。

从图 6 可以看出, 在分类的平均精度方面, MI-EMD-SVM 方法最高 (96.76%), 它比 SVM 要高

10.72%，比 EMD-SVM 要高 5.15%；EMD-SVM 方法比 SVM 高 5.57%。从图 7 可以看出，在分类的整体精度方面，MI-EMD-SVM 方法最高(96.92%)，它比 SVM 要高 9.15%，比 EMD-SVM 要高 3.34%；EMD-SVM 方法只比 SVM 高 5.81%。EMD-SVM 算法的分类精度明显高于 SVM 算法，说明了高光谱数据在 EMD 处理过程中，通过内固模态函数的选择和残差的去除实现了高光谱数据的特征提取。MI-EMD-SVM 算法的分类精度高于 EMD-SVM 算法，说明了高光谱数据经过互信息波段选择后，有效保留了与 C 最大相关波段图像，去除了含有冗余信息的波段图像，降低了高光谱数据的数据量，同时进一步提高了高光谱数据可分性。MI-EMD-SVM 算法的分类精度大大高于 SVM 算法，说明了 MI-EMD-SVM 算法具备较好的滤波效果。

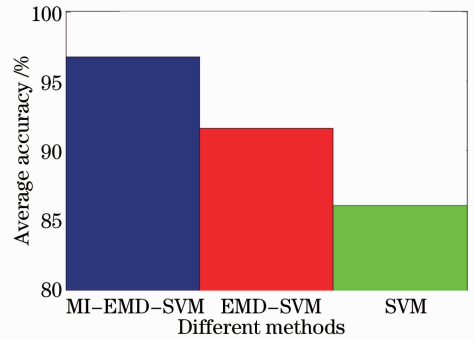


图 6 MI-EMD-SVM、EMD-SVM、SVM 平均精度对比
Fig. 6 Comparison on average accuracy in MI-EMD-SVM, EMD-SVM and SVM

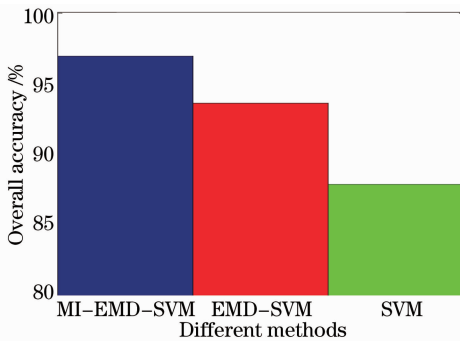


图 7 MI-EMD-SVM、EMD-SVM、SVM 整体精度对比
Fig. 7 Comparison on overall accuracy in MI-EMD-SVM, EMD-SVM and SVM

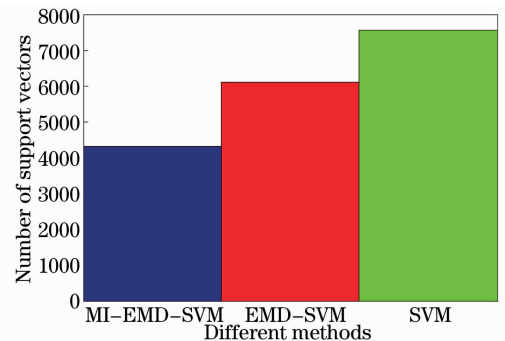


图 8 MI-EMD-SVM 与 EMD-SVM、SVM 支持向量数目对比
Fig. 8 Comparison on number of support vectors in MI-EMD-SVM, EMD-SVM and SVM

表 3 MI-EMD-SVM 与 EMD-SVM、SVM 分类时间对比

Table 3 Comparison on time of classification in MI-EMD-SVM, EMD-SVM and SVM

Time of classification	MI-EMD-SVM	EMD-SVM	SVM
Training time /s	0.4300	2.0645	2.5176
Testing time /s	7.7156	36.0474	43.4566
Overall time /s	8.1456	38.1119	45.9742

在支持向量数方面，如图 8 所示，MI-EMD-SVM 的支持向量数为 4324，比 SVM 少 3234，比 EMD-SVM 少 1785。MI-EMD-SVM 的支持向量数是 SVM 的 57.21%、EMD-SVM 的 70.78%。在训练测试时间方面，实验电脑的 CPU 为 Pentium (R) Dual-Core E5300、内存为 2GB、软件为 Matlab R2010a，实验结果如表 3 所示；MI-EMD-SVM 的图像分类速度最快，分类时间仅用了 8.1456s，比 EMD-SVM 要少 29.9663s，比 SVM 要少 37.8286s。支持向量数目的减少和图像分类时间的缩短，说明 MI-EMD-SVM 算法在波段选择过程中减少了冗余信息，进而提高了高光谱数据的分类速度。

仿真实验结果表明，MI-EMD-SVM 算法实现了对高光谱数据的冗余信息的处理和本质特征提取，不仅获得高精度的分类效果，而且还减少分类过程中的支持向量数目，缩短图像分类的训练时间和测试时间。

6 结 论

结合互信息波段选择和经验模态分解的优点，提出了基于互信息波段选择和经验模态分解的高精度高光谱数据分类算法(MI-EMD-SVM)。为了进一步验证 MI-EMD-SVM 算法的有效性，采用 MI-EMD-SVM 和传统 SVM、EMD-SVM 分类算法分别对高光谱数据集 92AV3C 进行分类实验和比较分析。仿真实验

结果证实,MI-EMD-SVM 分类算法不仅提高分类精度并获得高精度的分类效果,而且还减少分类过程中的支持向量数目,缩短高光谱数据的分类时间。

参 考 文 献

- 1 Baofeng Guo, Steve R. Gunn, R. I. Damper *et al.*. Band selection for hyperspectral image classification using mutual information[J]. *IEEE Geoscience and Remote Sensing Letters*, 2006, **3**(4): 522~526
- 2 Han Ling, Dong Lianfeng, Zhang Min *et al.*. Destriping hyperspectral image based on an improved moment matching method[J]. *Acta Optica Sinica*, 2009, **29**(12): 3333~3338
韩 玲,董连凤,张 敏 等. 基于改进的矩匹配方法高光谱影像条带噪声滤波技术[J]. *光学学报*, 2009, **29**(12): 3333~3338
- 3 Liu Xiaogang, Zhao Huijie, Li Na. Feature extraction based on multifractal spectrum for hyperspectral data[J]. *Acta Optica Sinica*, 2009, **29**(3): 844~848
刘小刚,赵慧洁,李 娜. 基于多重分形谱的高光谱数据特征提取[J]. *光学学报*, 2009, **29**(3): 844~848
- 4 Li Shanshan, Zhang Bing, Gao Lianru *et al.*. Research of hyperspectral target detection algorithms based on variance minimum[J]. *Acta Optica Sinica*, 2010, **30**(7): 2116~2122
李山山,张 兵,高连如 等. 基于方差最小的高光谱目标探测算法研究[J]. *光学学报*, 2010, **30**(7): 2116~2122
- 5 J. A. Hoeting, D. Madigan, A. E. Raftery *et al.*. Bayesian model averaging: a tutorial[J]. *Statistical Science*, 1999, **14**(4): 382~417
- 6 G. R. Xuan, X. M. Zhu, P. Q. Chai *et al.*. Feature selection based on the bhattacharyya distance[C]. Proceedings of 18th International Conference on Pattern Recognition, Hong Kong, 2006. 1232~1235
- 7 R. P. W. Duin, M. Loog. Linear dimensionality reduction via a heteroscedastic extension of LDA; the Chernoff criterion [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, **26**(6): 732~739
- 8 A. Ifarraguerri, M. W. Prairie. Visual method for spectral band selection[J]. *IEEE Geoscience and Remote Sensing Letters*, 2004, **1**(2): 101~106
- 9 R. Battiti. Using mutual information for selecting features in supervised neural net learning[J]. *IEEE Transactions on Neural Networks*, 1994, **5**(4): 537~550
- 10 Nojun Kwak, Chong-Ho Choi. Improved mutual information feature selector for neural networks in supervised learning[C]. Proceeding of 1999 International Joint Conference on Neural Networks, Washington, 1999, **2**: 1313~1318
- 11 N. E. Huang, Z. Shen, S. R. Long. The empirical mode decomposition and the hilbert spectrum for nonlinear and nonstationary time series analysis[C]. Proceeding of Royal Society, London, 1998, **A454**: 903~995
- 12 He Guanglin, Peng Linke. FPGA implement of SVD for dimensionality reduction in hyperspectral images[J]. *Chinese J. Lasers*, 2009, **36**(11): 2983~2988
何光林,彭林科. 基于 FPGA 的高光谱图像奇异值分解降维技术[J]. *中国激光*, 2009, **36**(11): 2983~2988
- 13 H. Peng, F. Long, C. Ding. Feature selection based on mutual information; criteria of max-dependency, max-relevance, and min-redundancy[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, **27**(8): 1226~1238
- 14 A. Jain, D. Zongker. Feature selection; valuation, application, and small sample performance[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, **19**(2): 153~158
- 15 B. Demir, S. Erturk. Empirical mode decomposition of hyperspectral images for support vector machine classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2010, **48**(11): 4071~4084
- 16 AVIRIS Northwest Indiana's Indian Pines 1992 Dataset [OL]. Available: ftp://ftp.ecn.purdue.edu/biehl/PC_MultiSpec/ThyFiles.zip (ground truth map) and ftp://ftp.ecn.purdue.edu/biehl/MultiSpec/92AV3C
- 17 Zhang Miao, Shen Yi, Wang Qiang. Nonlinear correlation coefficient based kernel method for hyperspectral data classification[J]. *Acta Optica Sinica*, 2009, **29**(9): 2607~2614
张 淼,沈 毅,王 强. 基于非线性相关系数核方法的超光谱数据分类[J]. *光学学报*, 2009, **29**(9): 2607~2614