

基于 ICA 与 SVM 算法的高光谱遥感影像分类

梁 亮, 杨敏华*, 李英芳

中南大学信息物理工程学院, 湖南 长沙 410083

摘 要 提出了一种利用独立分量分析(ICA)与支撑向量机(SVM)算法进行高光谱遥感影像分类的新方法。采用 ICA 算法对高光谱遥感影像(PHI 传感器获取, 80 波段)进行了特征提取, 并以提取出的影像数据(光谱维数为 20)构建 SVM 分类器。对 SVM 算法进行核函数删选与参数寻优后, 发现采用 RBF 核的 SVM 算法($C=10^3$, $\gamma=0.05$)分类结果最佳, 分类精度与 Kappa 系数分别达 94.5127% 与 0.9351, 优于 BP 神经网络(分类精度 39.4758%, Kappa 系数 0.3155)、波谱角分类(分类精度 80.2826%, Kappa 系数 0.7709)、最小距离分类(分类精度 85.4627%, Kappa 系数 0.8277)以及最大似然分类(分类精度 86.0156%, Kappa 系数 0.8351)4 种方法。针对分类结果常出现的“椒盐”现象, 利用形态学算子对 SVM(RBF 核)分类结果进行了类别集群处理, 将分类精度与 Kappa 系数分别提高至 94.7584% 与 0.9380, 获得了更接近实况的分类图像。结果表明: ICA 结合 SVM 算法准确率高, 是高光谱遥感影像分类的优选方法, 且类别集群是优化影像分类的有效方法之一。

关键词 高光谱; 分类; 支撑向量机(SVM); 独立分量分析(ICA); 类别集群

中图分类号: TP751.1 文献标识码: A DOI: 10.3964/j.issn.1000-0593(2010)10-2724-05

引 言

高光谱遥感技术是 20 世纪 80 年代遥感领域最大的成就之一, 它使空间对地观测时可获取连续波段的地物光谱图像以实现地表物体的准确识别, 从而在环境监测、海洋遥感、生态调查、精准农业、大气研究以及目标分类等众多领域获得了广泛的应用^[1-5]。但由于高光谱影像包含的地物细节信息极为丰富, 而传统的统计模式识别方法是基于大数定理, 在其分类识别中往往难以获得足够的训练样本而影响到各类别先验知识估计的准确性。且高光谱影像的光谱维数很高, 分类时易因 Hughes 现象(维数灾难)而无法获得理想的结果, 从而导致了利用高光谱数据进行地物分类识别存在一系列的问题。

在分类识别中, 支持向量机(support vector machine, SVM)学习方法在解决小样本、非线性及高维模式识别中表现出许多特有的优势, 可有效避免高光谱分类时的 Hughes 现象, 已在高维度光谱的分析中获得了成功的应用^[6-8]。而

在特征信号的提取与光谱维度的压缩方面, 独立分量分析(independent component analysis, ICA)方法被证明可在取得高压缩率的同时能很好地保持数据的特征^[9-11]。因此, 如果将 ICA 与 SVM 算法结合起来设计分类器, 将可成功实现高光谱遥感影像海量信息的特征提取, 并在分类过程中有效避免 Hughes 现象而获得较理想的结果。本文结合 ICA 与 SVM 算法的优点设计分类模型, 并针对分类后图像通常所出现的“椒盐”现象(分类区域存在斑点或未分类的空洞), 采用类别集群的方法, 运用形态学算子对分类图像进行侵蚀操作, 以获得逼近实况的分类影像, 从而为高光谱遥感影像的分类提供一种新的方法。

1 算法原理与实验方法

1.1 独立分量分析原理

ICA 是近年来发展起来的一种新的盲信号分离技术^[12]。盲信号分离系指在源信号与信号混合模型未知的情况下, 从混合信号(观测信号)中分离出源信号的过程。其基本含义是

收稿日期: 2009-11-06, 修订日期: 2010-02-08

基金项目: 国家自然科学基金项目(30570279), 中南大学研究生创新项目(1343-74334000022), 中南大学拔尖博士研究生学位论文创新项目(196071131100007), 中南大学优秀博士论文扶持项目(2008yb024)与中南林业科技大学林业遥感信息工程研究中心开放性研究基金项目(RS2008k03)资助

作者简介: 梁 亮, 1981 年生, 中南大学信息物理工程学院博士研究生 e-mail: liangliang198119@163.com

* 通讯联系人 e-mail: yangminhua@163.com

将多道观测信号按统计独立的原则通过优化算法分解为若干独立分量,而这些独立分量系源信号的一种近似估计。其算法原理如下^[13, 14]:

设 $X = \{x_1, x_2, \dots, x_m\}$ 是源信号 $S = \{s_1, s_2, \dots, s_n\}$ 的观测值,假定第 i 个观测信号系 n 个独立分量线性混合而成,则有

$$x_i = a_{i1}s_1 + a_{i2}s_2 + \dots + a_{in}s_n \quad (i = 1, 2, \dots, m) \quad (1)$$

用 $A(m \times n)$ 表示混合矩阵 a_{ij} , 式(1)可用矢量形式表示为

$$X = AS \quad (2)$$

(2) 式即为独立分量分析模型。其中 $X = [x_1, x_2, \dots, x_m]^T$, $S = [s_1, s_2, \dots, s_n]^T$, A 是未知且可逆的 $m \times n$ 矩阵,表示信号源到接受信号间的传递函数。 A 的可逆性保证了源信号的可分离性。ICA 就是在 S 与 A 均未知的情况下寻找分离矩阵 $W = A^{-1}$, 并从 X 中分离出源信号,其中 A^{-1} 为 A 的逆矩阵,即

$$S = A^{-1}X = W^T X \quad (3)$$

1.2 支持向量机原理

SVM 是 Vapnik 等于 20 世纪 90 年代提出的一种机器学习方法^[15]。与传统算法所采用的经验风险最小化准则不同, SVM 建立在 VC 维理论和结构风险最小化原理的基础上,根据有限的样本信息在模型的复杂性和学习能力之间寻求最佳折衷,以期获得最佳的推广能力。SVM 的核心思想是把输入空间的样本通过非线性变换映射到高维核空间,在高维核空间求取具有较低 VC 维的最优分类超平面。其分类原理如下^[15, 16]:

设有样本集 $\{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$, 其中 $x_i \in R^d$ 表示输入模式, $y_i \in \{\pm 1\}$ 表示目标输出(其中 $i = 1, \dots, N$)。设最优超平面为 $w^T x_i + b = 0$, 则权值向量 w 和偏置 b 必须满足以下约束

$$y_i(w^T x_i + b) \geq 1 - \xi_i \quad (4)$$

其中 ξ_i 为松弛变量,表示模式与理想线性情况的偏离程度。SVM 的目标是找到一个使训练数据平均错误分类误差最小的超平面,从而可推导出以下优化问题

$$Q(w, \xi) = \frac{1}{2} w^T w + C \sum_{i=1}^N \xi_i \quad (5)$$

其中 C 为一需指定的正参数(惩罚系数),表示 SVM 对错样本的惩罚程度。根据拉格朗日乘法,最优分类超平面的求解可转化为以下的约束优化问题

$$Q(a) = \sum_{i=1}^N a_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N a_i a_j y_i y_j K(x_i, x_j) \quad (6)$$

其约束条件为

$$\sum_{i=1}^N a_i y_i = 0, \quad a_i \geq 0, \quad i = 1, \dots, N \quad (7)$$

其中 $\{a_i\}_{i=1}^N$ 为拉格朗日乘子,其中大部分的 a_i 为 0, 而不等于 0 的 a_i 所对应的样本即称为支持向量。 $K(x_i, x_j)$ 为满足 Mercer 定理的核函数,常用的有线性核(Linear)、多项式核(Polynomial)、Sigmoid 核与高斯径向基核(RBF) 4 种。

1.3 实验数据与方法

实验数据为北京昌平地区的 PHI 高光谱影像,其波段

范围为 412~ 833 nm, 波段数为 80, 所截取的大小为 930 行 \times 350 列。结合试验区的实际情况与影像特征,该地区的地物可分为以下类型:旺盛小麦(C1)、稀疏小麦(C2)、空闲耕地(C3)、水体(C4)、疏林地(C5)、硬化地(C6)、居民区(C7)。各类地物的分布概况与光谱特征见图 1。

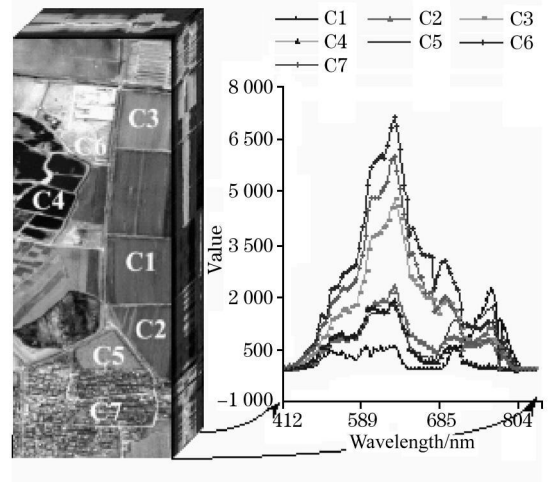


Fig 1 Distribution survey and typical spectra of the ground object in study area

Note: C1: Strong wheat; C2: Sparse wheat; C3: Idle arable land; C4: Water body; C5: Open woodland; C6: Hardening ground; C7: Residential areas

实验过程分为以下 4 步: (1) 采用 ICA 算法进行光谱维度压缩与特征提取; (2) 构建 SVM 分类器,并筛选出 SVM 算法的最佳参数与最佳核函数; (3) 与最小距离分类(minimum distance classification, MDC)、最大似然分类(maximum likelihood classification, MLC)、波谱角分类(spectral angle mapper classification, SAM)以及 BP 神经网络分类(BP ANN)进行比较分析; (4) 运用形态学算子,用类别集群的方法处理分类后的影像,获得接近真实地物的分类影像。

2 结果与分析

2.1 影像光谱特征提取

由于高光谱数据光谱维度过高,直接用于分类,一方面会因计算量过大而导致分类过程耗时较长,增加构建分类器的难度,而且会引入噪声而导致分类精度降低。因此,在进行 SVM 分类前,本文采用 ICA 对光谱信息进行特征提取,实现光谱数据的降维。

由于 ICA 无法保证处理后的光谱波段按信息量的大小从大到小排列,因此,利用二维空间相干性的方法进行 ICA 光谱波段的排序。二维空间相干性的量度是两个相关系数的均值,具体的计算方法见文献[13]。在 ENVI4.5 中进行 ICA 转化后,检查各波段的信息量,发现从 20 维以后所包含的绝大部分是噪声信息,因此,本文选用 ICA 转化后的前 20 维作为有效信息进行分类。

2.2 SVM 分类器的构建与优化

影像经 ICA 特征提取后,从中选取各类地物的部分样本以训练 SVM 分类器,同时选取部分样本作为检验集以进行分类结果的精度评估。选取样本的原则为充分考虑各类地物的光谱结构与纹理特征,从而使选择的样本具有代表性。本文所选取的训练样本数据占实验区数据总量的 32.8%, 检验样本数据占 15.0%。

构建分类器的过程中, SVM 核函数的选取与惩罚系数 C 及核函数参数 γ (Polynomial 核函数还包括 d) 的取值对分类精度影响较大。为了使 SVM 分类器达到最佳分类结果,本文参考文献[17]的方法进行参数寻优以确定 C 和 γ 的取值,并根据对分类结果的精度评估进行核函数的筛选。

在进行参数优化的条件下,各核函数的分类结果的精度评估如表 1 所示。在四类核函数所构建的 SVM 分类器分类精度均在 93% 以上,其中 RBF-SVM 的分类精度达 94.5127%, Kappa 系数达 0.9351, 取得了最佳的分类结果,是 SVM 算法进行高光谱分类的首选。而在采用同一核函数(RBF 核)的前提下,经 ICA 预处理后所取得的分类结果比传统的 PCA 方法具有更高的分类精度与 Kappa 系数,表明利用 SVM 算法进行高光谱影像分类时,ICA 具有比 PCA 更

好的预处理效果。

Table 1 Accuracy value and kappa coefficient of different SVM kernel functions and pretreatment methods

核函数	SVM			预处理	分类精度/%	Kappa 系数
	C	γ	d			
RBF	10	0.1	-	不处理	88.4726	0.8640
RBF	10^4	0.2	-	PCA	92.2195	0.9079
RBF	10^3	0.05	-	ICA	94.5127	0.9351
Sigmoid	10^3	0.05	-	ICA	93.2023	0.9197
Polynomial	10^2	0.05	2	ICA	93.0999	0.9184
Linear	10^2	-	-	ICA	94.4513	0.9344

2.3 SVM 与其他分类方法的比较分析

在影像分类中, MDC, MLC, SAM 以及 BP-ANN 算法是较常用的方法。将影像经 ICA 处理后,利用与 SVM 相同的训练集构建各分类器。各方法的分类结果见图 2。由图可知,除 BP-ANN 算法分类结果不理想外,其余方法均对 7 种不同的地物做出了较清晰的区分。

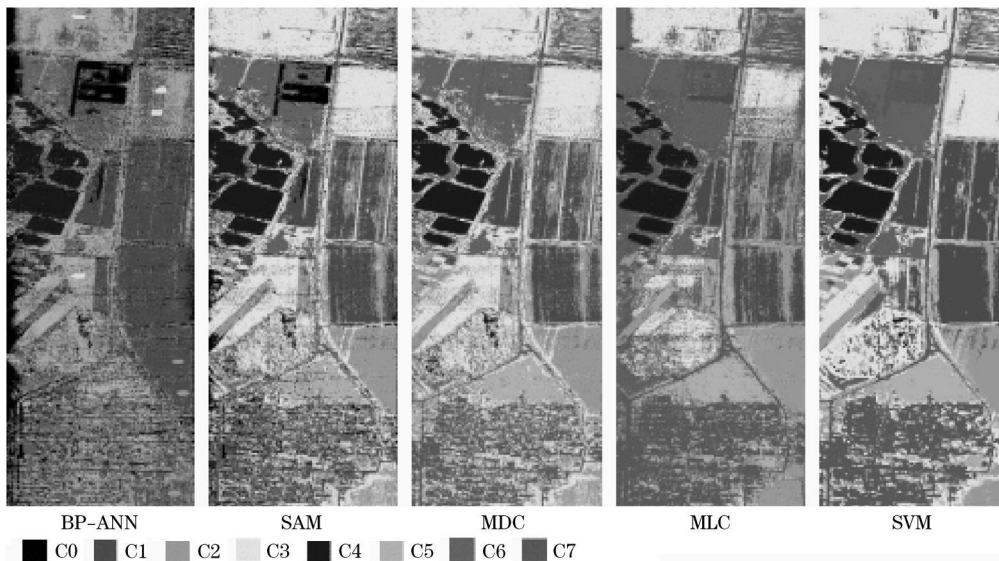


Fig 2 Classification results of different classification algorithms

Note: C0: Unclassified; C1: Strong wheat; C2: Sparse wheat; C3: Idle arable land; C4: Water body; C5: Open woodland; C6: Hardening ground; C7: Residential areas

利用检验样本对分类结果的分析表明, BP-ANN 分类精度仅达 39.4758%, 卡帕系数为 0.3155, 分类结果较差; SAM 是高光谱分类的常用方法,在本文中其分类精度为 80.2826%, Kappa 系数为 0.7709,也未取得精确的分类结果;而最小距离(MDC)与最大似然(MLC)虽取得了较好的分类结果,但两者的分类精度在 87% 以下, Kappa 系数在 0.84 以下,均低于采用 RBF 核函数的 SVM 分类方法(分类精度 94.5127%, Kappa 系数 0.9351)。可见 ICA 结合 SVM 进行高光谱遥感影像分类是一种优选方法(见表 2)。

Table 2 Accuracy value and kappa coefficient of different classification algorithms

分类方法	分类精度/%	卡帕系数
SVM(RBF)	94.5127	0.9351
MDC	85.4627	0.8277
MLC	86.0156	0.8351
SAM	80.2826	0.7709
BP-ANN	39.4758	0.3155

2.4 分类图像“椒盐”现象的去除

分类后的图像往往存在斑点或未分类的空洞,即所谓的

“椒盐”现象,与真实地物不符。本文运用形态学算子,以类别集群的方法,使用 3×3 的变换核对分类图像进行侵蚀操

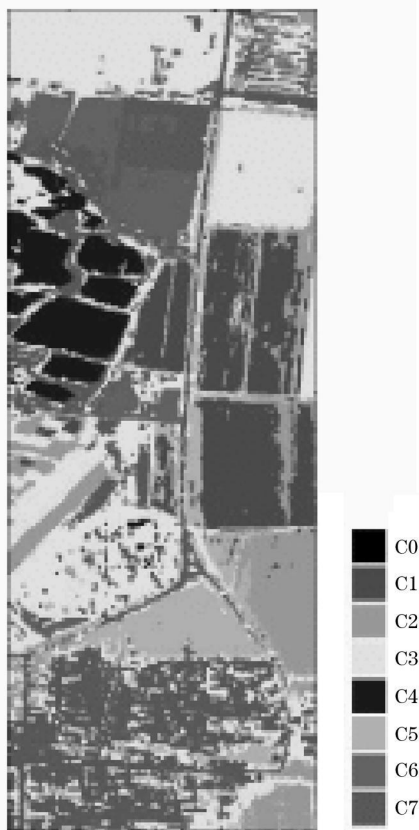


Fig 3 SVM Classification results after cluster treatment

Note: C0: Unclassified; C1: Strong wheat; C2: Sparse wheat; C3: Idle arable land; C4: Water body; C5: Open woodland; C6: Hardening ground; C7: Residential areas

作,从而达到去除“椒盐”现象的目的。图 3 为经类别集群处理后的 ICA-SVM(RBF 核)分类图。相对图 2(SVM),处理后的分类图像更为平滑,且分类精度由原来的 94.5127% 上升至 94.7584%, Kappa 系数由原来的 0.9351 上升至 0.9380,说明类别集群算法可改善分类结果,使其更接近真实地物的分类。

3 结论与讨论

本文针对高光谱数据的特点,设计了一种采用 ICA 与 SVM 算法进行高光谱遥感影像分类的新方法。通过对不同 SVM 核函数的比较分析发现 RBF 核可取得最佳的分类结果,是 ICA-SVM 算法进行高光谱分类的首选。研究发现 ICA 对高光谱影像预处理的结果优于传统的 PCA 算法,与苏令华等^[10]的结论一致,进一步证实了 ICA 算法在高光谱特征提取方面的优越性。相关文献表明, SVM 可在地物光谱分析等一些具有挑战性的领域中获得成功的应用^[6-9]。本文通过比较分析发现,结合 ICA 处理的 SVM 算法在高光谱遥感影像的分类识别中的精度也明显优于 MDC 和 MLC 等传统方法,但存在耗时较长的问题。说明 SVM 算法在高光谱遥感影像分析的应用中也具有巨大的潜力,但如何实现 SVM 算法的快速分类则有待进一步研究。

“椒盐”现象是分类图中常见的问题,使分类图因缺乏空间的连续性而与地物的实际状况不一致。传统的低通滤波虽然可实现分类图像的平滑,但类别信息常常会被临近类别的编码所干扰,且往往因信息损失量较大而导致分类精度下降。而本文运用形态学算子对分类影像进行类别集群处理,可在获得较平滑的分类图像的同时提高分类精度,实现 ICA-SVM 分类算法的优化,从而为高光谱遥感影像分类提供了一种较为精确、完善的新方法。

参 考 文 献

- [1] TONG Qing xi, ZHANG Bing, ZHENG Lar fen(童庆禧,张兵,郑兰芬). Hyperspectral Remote Sensing—Theory, Technology and Application(高光谱遥感—原理、技术与应用). Beijing: Higher Education Press(北京:高等教育出版社), 2006.
- [2] PU Rui liang, GONG Peng(浦瑞良,宫鹏). Hyperspectral Remote Sensing and Its Applications(高光谱遥感及其应用). Beijing: Higher Education Press(北京:高等教育出版社), 2000.
- [3] ZHANG Liang pei, ZHANG Li fu(张良培,张立福). Hyperspectral Remote Sensing(高光谱遥感). Wuhan: Wuhan University Press(武汉:武汉大学出版社), 2005.
- [4] WAN Yu qing, TAN Ke long, ZHOU Ri ping(万余庆,谭克龙,周日平). Application of Hyperspectral Remote Sensing(高光谱遥感应用研究). Beijing: Science Press(北京:科学出版社), 2006.
- [5] LIANG Liang, LIU Zhi xiao, YANG Min hua, et al(梁亮,刘志霄,杨敏华,等). Journal of Infrared and Millimeter Waves(红外与毫米波学报), 2009, 28(5): 353.
- [6] WU Di, HE Yong, FENG Shui juan, et al(吴迪,何勇,冯水娟,等). Journal of Infrared and Millimeter Waves(红外与毫米波学报), 2008, 27(3): 180.
- [7] WANG Li guo, ZHAO Chun hui, QIAO Yu long, et al(王立国,赵春晖,乔玉龙,等). Journal of Infrared and Millimeter Waves(红外与毫米波学报), 2008, 27(6): 442.
- [8] Sun D Y, Li Y M, Wang Q A. IEEE Transaction on Geoscience and Remote Sensing, 2009, 47(8): 2957.
- [9] SU Ling hua, YI Tong sheng, WAN Jiar wei(苏令华,衣同胜,万建伟). Acta Photonica Sinica(光子学报), 2008, 37(5): 973.
- [10] Tu T M, Huang P S, Chen P Y. IEEE Proceedings on Vision, Image and Signal Processing, 2001, 148(4): 217.
- [11] WANG J, CHANG C I. IEEE Transactions on Geoscience and Remote Sensing, 2006, 44(6): 1586.

- [12] Lee T W. Independent Component Analysis: Theory and Applications. Boston: Kluwer Academic Publisher, 1998.
- [13] Hyvarinen A, Oja E. Neural Networks, 2000, 12(45): 411.
- [14] Hyvarinen A. IEEE Transactions on Neural Networks, 1999, 10(3): 626.
- [15] Vapnik V N. Statistical Learning Theory. New York: Wiley, 1998.
- [16] John Shawe Taylor, Nello Cristianini. Kernel Methods for Pattern Analysis. Cambridge University Press, 2004.
- [17] HSU Chir wei, CHANG Chir chung, LIN Chir jen. <http://ntu.csie.org/~cclin/papers/guide/guide.pdf>.

Hyperspectral Remote Sensing Image Classification Based on ICA and SVM Algorithm

LIANG Liang, YANG Mir hua* , LI Ying fang

School of Info Physics and Geomatics Engineering, Central South University, Changsha 410083, China

Abstract A novel method was developed to classify hyperspectral remote sensing image based on independent component analysis (ICA) and support vector machine (SVM) algorithms. The characteristic information of the hyperspectral remote sensing image captured by PHI (made in China, with 80 bands) was extracted by ICA algorithm, and SVM classifier was established with the extracted image data (20 spectral dimensions). After kernel function selecting and parameter optimizing, it was found that the SVM algorithm (RBF kernel function; parameter $C=10^3$, $\gamma=0.05$) with accuracy 94.5127% and kappa coefficient 0.9351 has the best classification result, better than the results of four kinds of conventional algorithms, including neural net classification (accuracy 39.4758% and kappa coefficient 0.3155), spectral angle mapper classification (accuracy 80.2826% and kappa coefficient 0.7709), minimum distance classification (accuracy 85.4627% and kappa coefficient 0.8277) and maximum likelihood classification (accuracy 86.0156% and Kappa coefficient 0.8351). In order to control the "pepper and salt" phenomenon which appeared in classification map frequently, the classification result of SVM (RBF kernel) was operated by the method of clump classes using the morphological operators, and that the classification map closer to actual situation was acquired, with the accuracy and kappa coefficient increasing to 94.7584% and 0.9380, respectively. The study indicated that the ICA combined with SVM was an preferred method for hyperspectral remote sensing image classification, and clump classes was a effective method to optimized the classification result.

Keywords Hyperspectral; Classification; Support vector machine (SVM); Independent component analysis (ICA); Clump classes

(Received Nov. 6, 2009; accepted Feb. 8, 2010)

* Corresponding author