基于 SVM 遥感矿化蚀变信息提取研究

傅文杰[1,2],洪金益[1],朱谷昌[3]
(1. 中南大学地学与环境工程学院,长沙　410083;2. 莆田学院,莆田　351100;3. 有色金属矿产地质调查中心,北京　100814)

摘要:提出一种基于支持向量机(SVM)遥感数据矿化蚀变信息提取的新方法。该方法首先根据蚀变岩及矿体围岩的实测光谱数据,利用光谱角度制图法(SAM)提取训练样本,应用交叉比对(cross - validation)算法确定最优 SVM 模型参数,选择径向基(RBF)核函数,训练 SVM 分类器模型;然后,用训练好的 SVM 模型进行遥感矿化蚀变信息提取;最后,选择青海芒崖地区的 ETM 数据进行遥感矿化蚀变信息提取试验。试验结果经野外检查和验证,效果良好。

关键词:SAM;SVM;矿化蚀变信息;提取;遥感数据
中图分类号:TP 79　文献标识码:A　文章编号:1001 - 070X(2006)02 - 0016 - 04

0 引言

矿化蚀变信息是重要的找矿标志。在遥感图像中,矿化蚀变信息仅仅作为一种弱的信号存在于图像的背景中,因此,如何从遥感图像中提取或从背景中分割出矿化蚀变信息,一直是资源遥感应用与评价研究的重点和难点[1]。比值法与主成分分析法是目前遥感矿化蚀变信息提取的常用方法[2]。但这些方法主要着眼于增强处理,以求目视效果,在背景与噪声的干扰下,效果并不理想。近期新出现的一些方法,如小波分析、神经网络及分数维几何(分形或分几)等,也由于受样本数目以及维数的限制等原因,得不到令人满意的结果。支持向量机(SVM)是一种基于统计学习理论的机器学习算法,在解决有限样本、非线性及高维模式识别问题中表现出许多特有的性能,并且具有强大的泛化能力。但支持向量机算法本身是一种监督学习方法,需要一定数量标注过类型的训练样本。对遥感找矿信息而言,训练样本选取的方法主要通过野外实地调查,然后,在遥感影像上,利用准确的界线建立感兴趣区,选取训练样本,这种地面调查花费的人力、物力和财力较大,且有些地理区域,地面难以抵达,这些地区的样本信息就难以获取,实际应用效率较低。因此,本文首先利用光谱角度制图法(SAM)提取训练样本,然后,利用支持向量机(SVM)提取遥感矿化蚀变信息。通过对青海芒崖地区的应用试验,证明了该方法在遥感矿化蚀变信息提取应用中的可行性,同时得到了令人满意的效果。

1 SVM 基本原理及算法实现

1.1 SVM 基本原理

支持向量机完整的数学描述见文献[3,4],这里我们只对这个算法进行简单描述。支持向量机的基本思想如图 1 所示。

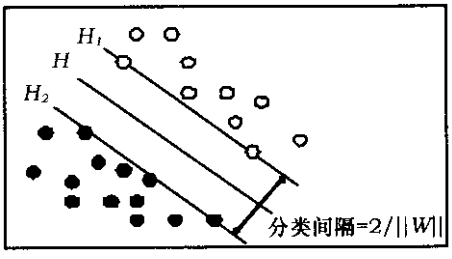


图 1　支持向量机的基本思想

图 1 中实心点和空心点代表两类样本,H 为分类线,H_1、H_2 分别为过各类中离分类线最近的样本且平行分类线的直线,它们之间的距离叫做分类间隔(Margin)。假设训练样本集为(x_i, y_i),$i = 1, \cdots, n$;$x \in R^d$。其中,x 为输入向量,i 为样本数,y 是输入向量所属的类别,对于两类的分类问题,$y \in \{+1, -1\}$。

收稿日期:2006 - 02 - 20;修订日期:2006 - 03 - 30
基金项目:国家“十五”科技攻关计划项目(2003BA612A - 04)中的“SVM 遥感数据矿化信息提取技术研究”资助。

在 D 维空间中,线性判别函数的一般形式为 $g(\boldsymbol{x}) = \boldsymbol{w} \cdot \boldsymbol{x} + b$,分类面方程为

$$\boldsymbol{w} \cdot \boldsymbol{x} + b = 0 \tag{1}$$

将判别函数进行归一化,使两类所有样本都满足 $|g(\boldsymbol{x})| \geqslant 1$,即,使离分类面最近的样本 $|g(\boldsymbol{x})| = 1$,这样分类间隔就等于 $2/\|\boldsymbol{w}\|$,因此,使间隔最大等价于使 $\|\boldsymbol{w}\|^2$ 最小;而要求分类线对所有样本正确分类,就是要求它满足约束条件

$$y_i[(\boldsymbol{w} \cdot x_i) + b] - 1 \geqslant 0 \tag{2}$$

式中,$i = 1, \cdots, n$。考虑到有些训练样本是线性不可分的,Cortes 和 Vapnik 等人[5] 引入了非负的松弛变量,将(2)式放宽为

$$y_i[(\boldsymbol{w} \cdot x_i) + b] \geqslant 1 - \xi_i \tag{3}$$

式中,当划分出现错误时,ξ_i 就会大于零。因此,在求分类平面的同时,$\sum_i \xi_i$ 的值也希望愈小愈好。所以,原本目标函数是求 $\|\boldsymbol{w}\|^2/2$ 的最小值,会变成求目标函数 $\|\boldsymbol{w}\|^2/2 + C(\sum_i \xi_i)$ 的最小值。其中,$C > 0$ 是一个常数,是可调的参数,控制对错分样本惩罚的程度,C 越大表示对错误的惩罚越重。这是一个二次规划问题,其对偶形式为

$$\max \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \tag{4}$$

$$0 \leqslant \alpha_i \leqslant C \tag{5}$$

求解这个对偶问题得到最优分类函数

$$f(\boldsymbol{x}) = \operatorname{sgn}\{\sum_{i=1}^{n} \alpha_i y_i (x_i \cdot x_j) + b\} \tag{6}$$

式中,α_i 为 Lagrange 乘子,α_i 不为零的样本点就称作支持向量,即位于 H_1、H_2 上的样本。这些向量充分描述了整个训练样本集数据的特征,使得对支持向量集的线性划分等价于对整个数据集的分类。

对非线性可分样本,支持向量机构造分类决策函数的方法是,首先,将训练数据从原始模式空间经过特定核函数的非线性变换,映射到高维特征空间[6] ;然后,在特征空间中,寻找最优分类超平面,该超平面实际上对应原始模式空间中的非线性分类面。因此,支持向量机在处理非线性情况时,仅比线性情况多了一个非线性映射环节。其对偶形式变为

$$\max \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \tag{7}$$

式中,$K(x_i, x_j)$ 为满足 Mercer 条件的核函数。目前,应用较多的核函数有 3 种:即多项式核函数、径向基核函数和 Sigmoid 核函数。

对于多类分类问题,可采用多个二分类 SVM 组合的办法解决。

1.2　SVM 算法实现

首先,利用光谱角制图法(SAM)[7] 选取足够数量的训练样本;然后,利用 SVM 构造最优分类器模型;最后,利用该分类器模型对遥感图像进行信息提取,其算法描述如下:

第 1 步,对每个影像像元光谱计算它与各类实测光谱间的光谱角,并将该角度和所设定的初始阈值比较,若小于某类设定阈值,则选为该类的训练样本。这里的技术关键是初始阈值的设置,既要保证采集到足够的样本,又要保证训练样本与提取目标的最大相似性;

第 2 步,如果某类的训练样本数未达到所需的样本数,则重新调整阈值,执行步骤 1;

第 3 步,利用提取的训练样本集,应用 cross - validation 算法确定最优 SVM 模型参数;

第 4 步,根据最优 SVM 模型参数训练 SVM 分类器模型;

第 5 步,采用训练好的 SVM 模型对整个遥感图像进行信息提取。

2　实验结果及分析

2.1　实验数据

试验场选择在青海芒崖金(多金属)成矿区。范围长 22 km,宽 10.5 km,面积为 231 km^2。该区位于阿尔金南缘断裂带东段,地层主要为奥陶系中基性火山岩、变质岩和海相碎屑岩,岩浆岩主要为晚元古代中晚期花岗岩、闪长岩和华力西期中酸性岩。在该区已发现的金(多金属) 矿床(点) 中,近矿围岩蚀变发育,蚀变带宽 20 ~ 200 m,长 2 000 m。围岩蚀变多为褐铁矿化、黄铁矿化、黄钾铁矾化、绿泥石化、绿帘石化及绢云母化等,这些蚀变矿物(组合) 都是提取遥感找矿信息异常的重要目标。实验所用的数据为 2001 年 11 月 7 日的 Landsat7 ETM 数据,选用其中的 6 个波段(1 ~ 5 和 7 波段)。通过地面调查和实测光谱数据分析结果将研究区内岩石类别划分为 6 类,即 花岗岩、花岗闪长岩、未蚀变凝灰岩、石英砂岩、铁化蚀变岩及泥化蚀变岩。对图像进行大气校正、几何校正和地理配准等数据预处理。因为,TM 图像的每个像元值是灰度值,需将它转换为反射率值,具体转换公式为[5]

$$L = gain \times DN + bias \tag{8}$$

$$\rho = \pi L d_s^2 / (E_0 \cos\theta) \tag{9}$$

式中,L 是地物在大气顶部的辐射亮度;DN 是

像元灰度值；增益（*gain*）和偏移值（*bias*）可从头文件中得到；ρ 是地物反射率；d_s 为日地天文单位距离；E_0 为大气顶部的太阳辐照度；θ 为成像时的太阳天顶角，也可从图像的头文件中读取。

2.2 训练和测试样本采集

由于支持向量机法是根据训练样本来选择特征参数，建立判别函数进行分类的，因此，训练样本选择的质量数量很大程度上关系到整个分类结果的精度。有研究表明，训练样本选择比分类算法的选择对分类精度的影响更大[8]。本次训练和测试样本采集先是利用光谱仪对已知矿（化）区与金（多金属）矿化有关的蚀变岩及围岩进行实地光谱测试，根据实测光谱数据分析结果将研究区内岩石类别划分为6类，取其在对应 ETM1 ETM2 ETM3 ETM4 ETM5 和 ETM7 波段的光谱均值建立6维空间的矢量样本，然后在遥感影像上，利用 SAM 算法，对每个类别分别提取 200 个像元组成样本集。

2.3 SVM 模型选择及实验结果

选择合适的核函数参数和误差惩罚因子 C 对 SVM 的性能至关重要。统计学理论目前对这些问题给出了一些建议和解释，但还没有给出实际可行的方案，因此，在使用支持向量机进行分类和预测时，如何选择适当的参数就成了一个非常重要的问题。Vapnik 等人的研究表明，SVM 的性能与所选用的核函数的类型关系不大，而核函数的参数和误差惩罚因子 C 是影响 SVM 性能的主要因素[9]。因此，在笔者设计 SVM 分类器时，选择 *RBF* 函数作为核函数，原因主要有以下几点：首先，*RBF* 函数可以将样本非线性地规划到更高维的空间中，从而解决类标签和属性间非线性的关系问题，这是线性核函数无法解决的。实际上，线性核函数是 *RBF* 核函数的特例。另外，Sigmoid 核函数取某些特定参数时性能和 *RBF* 相同。因此，对本文算法来说，需要选择的参数即为 *RBF* 核参数 γ 和惩罚因子 C。本文采用交叉比对算法（cross－validation）来选取这两个参数。交叉比对法过程是：将样本集分成 n 个子集，每次将其中 $n-1$ 个子集代入 SVM 训练，针对 SVM 的惩罚因子 C 和不同的核函数，计算剩余 1 个子集的分类正确率。系统根据误差自动调整惩罚因子的取值。通常，n 的取值一般为 10。通过上述实验选择的最佳 *SVM* 模型参数为 $\gamma=0.5$，$C=32$。最后，应用该模型对研究区整个 ETM 影像进行了信息提取，提取结果如插页彩片 6。本文实验中采用的 SVM 算法来自台湾大学 Chih－Jen Lin 开发的 LibSVM 软件[10]，LibSVM 中采用“一对一”的策略解决多类别分类问题。LibSVM 已广泛用于 SVM、回归和分类估计，并且支持多类分类，它的基本算法结合了 Platt 提出的 SMO[11] 和 Joachims 提出的 SVMLight[12] 算法思想。SVM 信息提取具体实现步骤如下：

（1）将数据转换成 SVM 软件要求的格式；

（2）使用交叉比对法找到最佳参数 C 和 γ；

（3）用 C 和 γ 训练整个训练集，得到 SVM 分类器模型；

（4）应用 SVM 分类器模型对整个图像进行信息提取。

2.4 精度分析

精度分析是遥感数据分类过程中一项不可缺少的工作，通过精度分析，分类者能确定分类模式的有效性。通过野外调查和与已知的矿（化）区资料对比，本文应用基于 SAM 的 SVM 方法提取的矿化蚀变信息与已知的矿（化）点位置基本吻合，新发现的一些矿化蚀变异常点，均不同程度存在矿化蚀变现象。实地验证如插页彩片 7（红线圈为已知金矿体，绿线圈范围为经实地检查新发现的金矿化体或蚀变岩），吻合率达 95%，见矿率 40%。

另外，本文还将基于 SVM 方法提取的结果与传统的比值法提取结果（插页彩片 8）进行了比较，通过对比可以看出，利用比值法进行蚀变信息提取出现了许多假蚀变信息，效果很不理想。

3 结论

本文提出的基于 SVM 遥感矿化信息提取方法，既减少了野外实地调查工作量，同时又保证了较高的分类精度，得到了比较满意的效果。克服了传统的提取遥感矿化蚀变信息方法应用中受样本数目以及维数的限制等局限。由于受光谱波长的限制，采用 TM 和 ETM 的光谱数据进行矿化信息提取，对同物异谱与同谱异物的问题尚不能很好解决，在一定程度上会影响信息提取的精度，如果采用高光谱数据可能会提高提取的精度。这方面的工作还有待开展。

参考文献

[1] Yu X，Reed I S，Stocker A D. Comparative performance analysis of adaptive multispectral detectors[J]. IEEE Transaction on Signal Processing，1993，41（8）：2639－2656.

[2] Ruiz－armenta J R. Prol－ledesma. Techniques for enhancing the spectral response of hydro－thermal alteration minerals in Thematic

Mapper images of Central Mexico[J]. Int. J. Remote Sensing, 1998, 19 (10) :1981 – 2000.

[3] Vapnik V N. The Nature of Statistical Learning theory[M]. New York :Springer Verlag,1995.

[4] Vapnik V N. Statistical Learning theory[M]. New York :Wiley, 1998.

[5] Cortes C, Vapnik V N. Support – Vector Networks[J]. Machine Learning,1995,20(3) :273 – 297.

[6] Scholkopf B, Mika S, Burges C. Input space vs. feature space in kernel – based methods[J]. Transactions on Neural Networks, 1999,10(5) :1000 – 1017.

[7] Kruse F A, et al. The Spectral Image Processing System (SIPS) – Interactive Visualization and Analysis of Imaging Spectrometer Data [J]. Remote Sens. Environ,1993, 44 :145 – 163.

[8] Hixson M, Scholz D, Fuhs N et al. Evaluation of several schemes for classification of remotely sensed data[J]. Photogrammetric Engineering and Remote Sensing, 1980,46 :1547 – 1553.

[9] Vapnik V N. 统计学习理论的本质[M]. 张学工译. 北京 :清华大学出版社. 2000.

[10] Chih – Chung Chang, Chih – Jen Lin. LIBSVM :A library for support vector machines,2001. software available at[EB/OL]. http ://www. csie. ntu. edu. tw/ ~ cjlin/libsvm

[11] Platt J. Fast training of support vector machines using sequential minimal optimization[M]. Advances in Kernel Metheds – support vector Learning. Cambridge. MA :MITpress,1999.

[12] Joachims T. Making Large – scale SVM Learning Practiceal[M]. In Advances in Kernel Methods – support Vector Learning, MIT Press,Cambridge, MA, 1998.

THE EXTRACTION OF MINERALIZED AND ALTERED ROCK INFORMATION FROM REMOTE SENSING IMAGE BASED ON SVM

FU Wen – jie[1,2], HONG Jin – yi[1], ZHU Gu – chang[3]

(1. *School of Geoscience and Environmental Engineering, Central South University, Changsha* 410083, *China* ;2. *Putian University, Putian* 351100, *China* ;3. *Nonferrous Metals Resource Geological Survey of China, Beijing* 100814, *China*)

Abstract :A new method for extracting mineralization information from remote sensing image based on Support Vector Machines (SVM) is presented in this paper. According to the field measured spectral data of mineralized alteration rocks and wall rocks, the authors first extracted the training examples by Spectral Angle Mapper (SAM), and then selected the RBF as the kernel function. After that, cross – validation algorithm was applied to seek superior SVM model parameters. This model was used to extract mineralization information from remote sensing image in Mangya area, Qinghai province. Practice has proved that this method is effective in extracting mineralization information.

Key words :SAM ;SVM ;Mineralied and altered rock information ;Extraction ;Remote sensing data

第一作者简介：傅文杰(1967 –),男,中南大学国土资源信息工程博士研究生,研究方向：遥感技术及 GIS 应用。

(责任编辑：肖继春)

消息报道

未来 10 年我国计划发射 18 颗陆地观测卫星

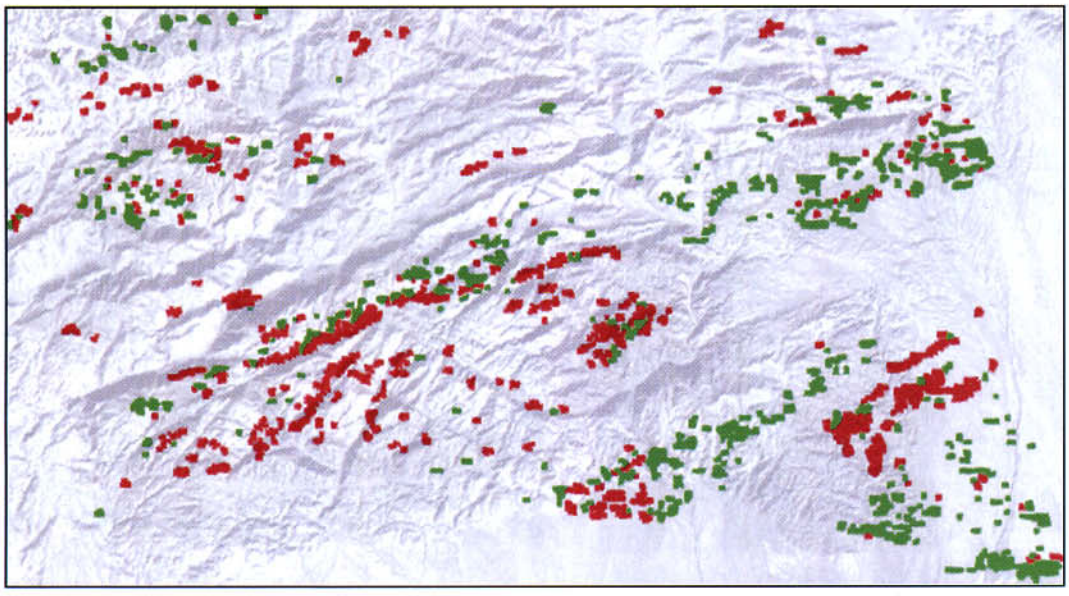
中国资源卫星应用中心 3 月 17 日在北京发布中巴地球资源卫星 02 星数据免费发放政策。记者从发布会上获悉,中巴地球资源卫星 02B 星将于 2007 年发射,03、04 星研制计划已经国务院批准,各项研制工作全面展开。“十一五”期间至后续 5 年内,我国将发射除海洋和气象卫星外的 18 颗陆地观测卫星。

这 18 颗陆地观测卫星,包括资源卫星、环境减灾小卫星星座等,将与气象卫星系列和海洋卫星系列组成长期稳定运行的中国卫星对地观测体系。

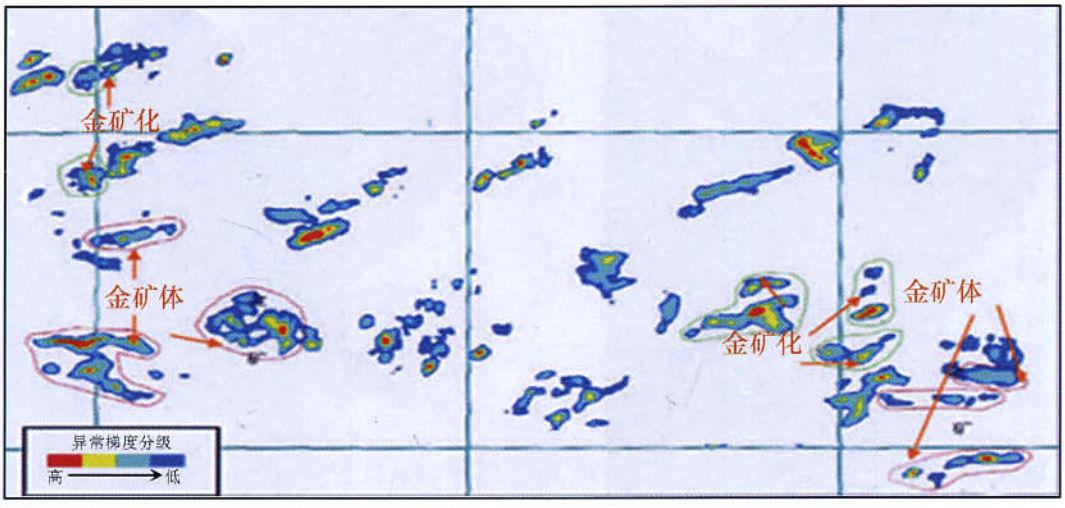
我国资源卫星应用取得了长足发展。截止 2005 年年底,中国资源卫星应用中心接受并存档了资源卫星获取的 80 余万景数据,根据用户需求分发数据产品 24 000 景,相当于覆盖我国全部国土 18 遍。这些资源卫星数据广泛应用于农业、林业、水利、国土资源、城市规划、环境保护、灾害监测和国防建设等众多领域。

为了扩大资源卫星数据的应用,更好地发挥资源卫星的社会经济效益,国防科工委宣布自 2006 年 4 月 1 日起,中巴地球资源卫星 02 号数据对国内用户免费网上发放政策,首批 70 家国内用户获得授权。

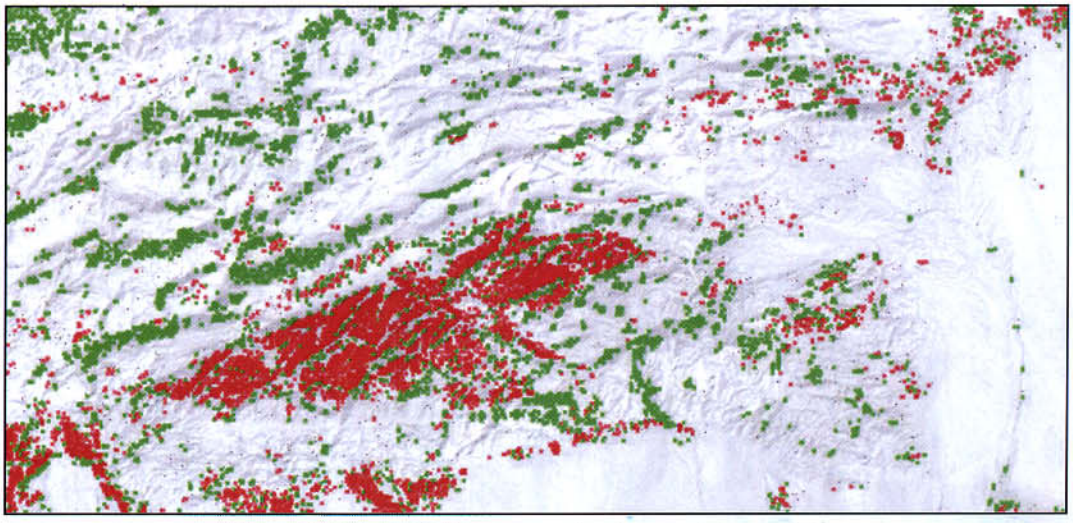
《摘自北京青年报》



彩片6　SVM提取结果叠加ETM原图（7、4、1，红色为铁化蚀变，绿色为泥化蚀变）



彩片7　实地验证图



彩片8　比值法提取结果（红色为铁化蚀变，绿色为泥化蚀变）