

成像光谱数据建模与分类识别技术研究

——以云南腾冲矿区为试验区

张杰林^{1,2}, 岑长华¹, 牛 铮³, 陆书宁⁴, 曹代勇²

(1. 核工业北京地质研究院, 北京 100029; 2. 中国矿业大学, 北京 100083;

3. 中国科学院遥感应用研究所, 北京 100101; 4. 国家卫星应用中心, 北京 100055)

[摘 要]在系统分析成像光谱数据特征及岩石矿物具有诊断意义的吸收光谱特征形成机理的基础上, 采用基于相关系数测度的光谱匹配技术、基于高斯改进型模型的光谱建模技术及人工神经网络分类算法, 实现了岩石矿物光谱特征波形对比分析及诊断光谱信息提取与建模, 提高了光谱分类识别算法的计算速度和分类精度。采用上述技术对云南腾冲铀矿区进行实验研究, 取得良好效果。

[关键词]成像光谱 光谱建模 光谱匹配 人工神经网络 分类识别

[中图分类号]TP79; P575.4 **[文献标识码]**A **[文章编号]**0495-5331(2003)02-0052-04

0 引言

成像光谱测量技术与成像雷达技术一起并列为当代遥感发展的两大前沿技术^[1]。成像光谱测量技术具有图谱合一的特征, 成像光谱测量获得的图像包含了大量的空间、辐射及光谱三重信息^[2], 构筑成地物光谱测量的海量数据库^[8]。如何有效地充分利用这一海量数据进行地面目标的定量化、自动化识别, 从而为地质矿物分类与制图、工程地质、环境地质及军用领域的研究服务, 成为当前遥感领域的研究热点。本文重点开展岩石矿物高光谱数据建模、光谱匹配与分类识别技术的研究。

1 光谱建模技术

1.1 高斯改进型模型的理论基础

利用成像光谱测量技术直接识别岩矿类型的主要依据是岩石、矿物表面与太阳电磁波相互作用而形成的可诊断的光谱吸收特征。大量的理论和实验表明, 这种光谱吸收特征的成因主要是在太阳电磁能量的激发下, 矿物、岩石内部的电子跃迁和分子振动过程对电磁能量的吸收作用^[3,9,10], 具有稳定化学组分和物理结构的岩石矿物则具有较稳定的本征光谱吸收特征。

在太阳电磁能量的激发下, 岩石矿物表面的吸收能量与高斯分布的平方根成正比, 即高斯分布作为物质表面能量的函数, 与反射率具有一定的比例关系^[3,4]; 在晶体场理论中, 吸收能量(e)与平均键长(r)存在幂定律关系: $e \propto r^n$ 。因此, 平均键长的高斯分布能映射成吸收能量的高斯分布^[1,3]。

$$R = G \times \exp\left[\frac{-(r^n - \mu^n)^2}{2\omega^2}\right] \quad (1)$$

其中: G 为吸收峰强度, μ 为吸收峰波长值, ω 为吸收峰宽度。

(1) 式中指数 n 的变化, 相当于改变光谱吸收峰的对称性, 即吸收峰两翼相对斜率的变化。采用该模型, 可以将具体的矿物吸收光谱曲线分解及拟合成高斯分布曲线。

1.2 算法实现及模拟结果分析

为了消除系统误差的影响, 突出地物吸收峰特征, 在光谱数据建模时, 要首先进行数据预处理。主要包括: 样本光滑处理、包络线消除及光谱数据对数变换等。

本文采用云南腾冲 381 矿区砂岩岩芯的高光谱测量数据进行实验分析研究, 模拟目标是砂岩铀矿在 2.205 μm 位置的吸收峰特征。从误差曲线图(图 1)和误差分析表(表 1)中可以看出, 在吸收峰

[收稿日期]2002-03-01; **[修订日期]**2002-04-22; **[责任编辑]**曲丽莉。

[基金项目]核工业遥感信息与图像分析技术重点实验室基金项目(编号: 遥 U66.6)资助。

[第一作者简介]张杰林(1970年-), 男, 1995年毕业于中国矿业大学, 获硕士学位, 在职博士研究生, 高级工程师, 主要研究成像光谱图像处理及其应用。

位置(2.205 μm)及左右肩部(带宽30 μm 之内,2.19~2.22 μm),误差值接近于0,实测光谱数据点与模拟曲线非常吻合,表明利用高斯改进型模型模拟地物光谱吸收特征是一种行之有效的技术手段。利用该技术可以将地物光谱曲线分解成一系列的高斯模拟曲线,用确定的函数形式定量化描述地物吸收光谱特征,为光谱数据库编码、建库和查询提供便利工具。

表1 高斯改进型模型模拟结果误差分析表

波长/mm	自然对数反射率	模拟结果	误差值
2.18	-0.129807	-0.170641	0.040834
2.185	-0.159274	-0.183988	0.024714
2.19	-0.194242	-0.204329	0.010087
2.195	-0.222241	-0.224041	0.0018
2.2	-0.23851	-0.238134	-0.000376
2.205	-0.240048	-0.240433	0.000385
2.21	-0.226521	-0.226677	0.000157
2.215	-0.200378	-0.199431	-0.000947
2.22	-0.166123	-0.167483	0.00136

测试时间:2001年9月20日。

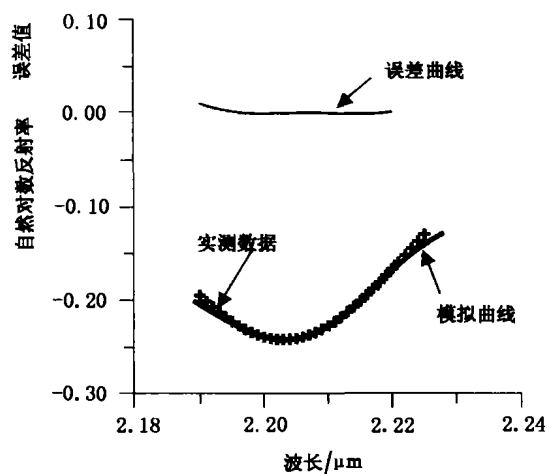


图1 光谱建模结果图

2 光谱匹配模式

2.1 光谱二值编码匹配

光谱二值编码技术是一种较为常用的光谱匹配方法,其基本原理就是对预处理的光谱数据,首先指定一个阈值 T ,大于此阈值的数据赋1,否则赋0,这样被测光谱可用简单的0、1序列来表述,完成编码后,可利用基于最小汉明距离^[2]的算法来进行匹配识别(图2)。

其基本算法为:

$$H(n) = \begin{cases} 0, & x(n) \leq T \\ 1, & x(n) > T \end{cases} \quad n = 1, 2, 3, \dots, N \quad (2)$$

其中: $x(n)$ 为第 n 波段的反射率值; T 为阈值; $H(n)$ 为编码值。

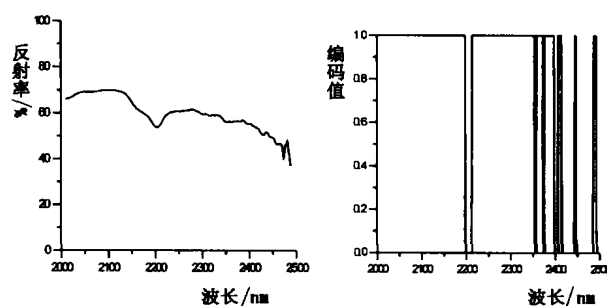


图2 云南腾冲381矿中粒砂砾岩岩芯的二值编码曲线图

2.2 基于相关系数测度的光谱匹配技术

二值编码技术使光谱匹配过程快速、有效,有助于提高光谱数据分析处理的效率。但由于该技术在编码过程中忽略了许多细微的吸收光谱信息,因此只适合于较粗略地识别岩石矿物光谱,而不能应用于精细的高光谱数据的匹配识别研究中。为克服此弊端,本文采用相关系数测度进行光谱匹配的分析处理研究。光谱匹配技术主要是依据距离来度量模式的相似性,对于成像光谱数据匹配而言,相关系数测度同时兼顾平均值和协方差两个统计参量,是计算成像光谱超维空间中两类统计距离的理想测度。

为了消除各种干扰因素的影响,保证各变量在分布上的同等地位,光谱数据匹配时,要首先对数据进行标准化变换和导数变换。

1) 利用标准差进行属数据标准化变换^[5]:

$$x'_i = \frac{x_i - \bar{x}}{S_i} \quad (3)$$

$$\text{其中: } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (4)$$

$$S_i = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (5)$$

2) 一阶导数计算:

$$X'_{(i)} = \frac{[X_{(i+1)} - X_{(i-1)}]}{2 \times \Delta\lambda} \quad (6)$$

3) 二阶导数计算:

$$X''_{(i)} = \frac{[X_{(i+1)} - 2X_{(i)} + X_{(i-1)}]}{2 \times \Delta\lambda^2} \quad (7)$$

其中: $\Delta\lambda$ 为波段间隔。

4) 相关系数计算:

$$c_{ij} = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2 \cdot \sum_{k=1}^n (x_{kj} - \bar{x}_j)^2}} \quad (8)$$

其中: $\bar{x}_i, \bar{x}_j$ 的计算同公式(4)。

2.3 实例分析

本文利用 VC++ 编程技术设计完成了光谱匹配软件,其主要功能包括数据的多维显示,匹配区选择,光谱数据的多参数匹配及匹配结果的动态显示与存储。

云南腾冲铀矿区同种岩性(砂岩)不同地点的岩芯高光谱测量数据匹配结果较好(图3),但匹配区间不同,匹配系数略有差异(表2),表明匹配区的选择在整个匹配计算中非常重要。因此应尽量选择地物诊断光谱特征区数据作为匹配区间,这样才能客观反映匹配结果。

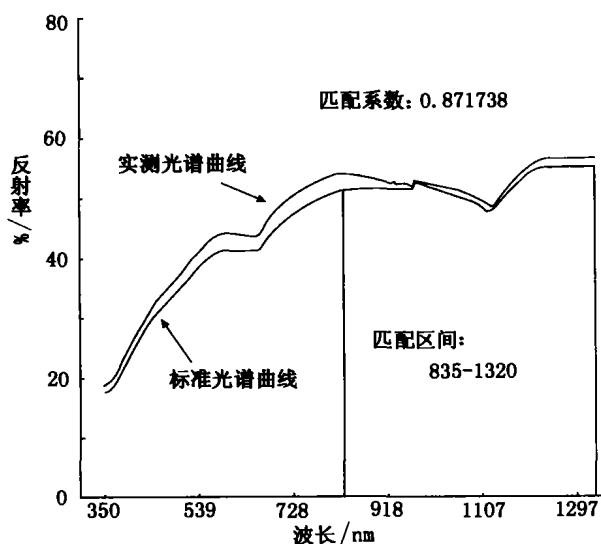


图3 光谱匹配结果图

表2 不同匹配区间匹配系数计算结果表

匹配区间/nm	匹配系数
500 ~ 748	0.999572
350 ~ 1000	0.995822
379 ~ 1291	0.984153
835 ~ 1320	0.871738
682 ~ 1238	0.723582

测试时间:2001年9月20日。

3 基于人工神经网络的分类识别模式

基于模式识别的分类技术,如统计模式识别、人工神经网络算法等,在遥感数据处理领域应用日益广泛^[4,6-8,11],本文采用人工神经网络分类模式进行成像光谱数据分类识别研究。

3.1 基本的神经元模型

人工神经网络是近十多年来迅速发展起来的新型的信息处理技术。其网络模型多样,如感知器模

型、RBF模型、HOPFIELD模型、BAM模型以及多层映射反向传播算法(BP)模型等^[7]。应用于光谱分类识别研究的人工神经网络技术,主要是以一定数量的具有已知特征的光谱数据为训练样本,采用反向传播算法(BP),训练神经网络,然后根据最佳的网络权重系数和阈值对未知样本进行分类处理研究。另外,由于神经元的变换函数为S型函数,因此可以实现从输入到输出的任意的非线性映射,这对于处理超高维空间的成像光谱数据非常有效。

基本的神经元模型具有N个光谱样本($R(1), R(2), R(3) \dots R(n)$)输入(图4),每个样本输入都通过一个适当的权重 $W(n)$ 与下一层网络相连,网络输出可表示为 $Y = \sum_{i=1}^n F(w * R(i), b)$, 其中 F 为S型变换函数, b 为阈值。

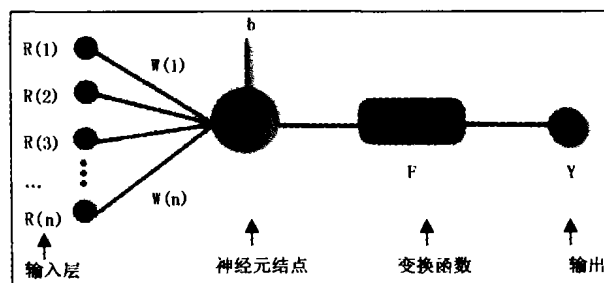


图4 基本神经元模型

3.2 网络结构设计及算法实现

人工神经网络是一个并行的分布式信息处理网络系统,它由多个基本神经元模型组成^[7]。在网络结构研究方面,本文设计了具有隐层的人工神经网络结构。而在算法研究工作中,考虑到传统的BP算法存在一些局限性,如:累积误差可能会陷入局部极值、学习步长不易确定及噪声影响大等。本文提出了分批样本学习、动态调整步长及惯性量的加入等策略,优化网络权重和阈值,以提高算法的可靠性。

3.3 实例分析

本次工作利用 ENVI 系统附带的 AVIRIS 成像光谱数据(50个波段)进行分类识别试验研究。由于图像为未知区,在选取训练区时,主要根据图像的自身特征进行选取,从技术角度来讲,作为分类算法的研究,该种方法是可行的。分类结果(图5)表明,相比于统计分类算法,神经元分类结果影像信息丰富、识别的信息量大,如一些地貌景观、道路有很好的体现。

4 结论与讨论

1)在系统分析岩石矿物具有诊断意义的吸收

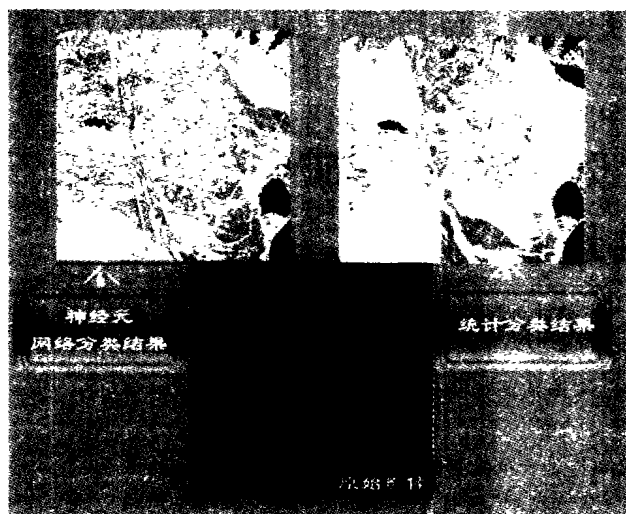


图5 光谱分类结果图像

光谱特征形成机理的基础上,采用高斯改进型模型,成功地模拟了云南腾冲铀矿区的砂岩岩芯在 $2.2\mu\text{m}$ 附近的光谱吸收特征。实验研究表明该技术可以量化描述地物吸收光谱特征,并为光谱数据建库及查询提供了便利工具。

2)在光谱匹配模式研究方面,利用二值编码匹配和基于相关系数测度的光谱匹配两种方式,系统研究试验区的地物光谱特征。结果表明:相关系数测度同时兼顾平均值和协方差两个统计参量,是计算高光谱超维空间中两类统计距离的理想测度,是典型目标波形对比分析强有力的工具。

3)在神经网络分类算法研究中,系统分析了传统的反向传播算法(B-P算法)的局限性,提出分批样本学习、动态调整步长和惯性量加入等策略,优化网络权重和阈值,以提高算法的计算速度和分

类精度,取得实效。

由于本项目工作时间有限,一些技术问题有待进一步研究,如:开展基于光谱和空间双重信息的分类技术研究,进一步完善人工神经网络分类算法;充分利用成像光谱测量系统提供的超高维空间数据,开展多源信息挖掘技术研究等。

[参考文献]

- [1] 浦瑞良,宫 鹏. 高光谱遥感及其应用[M]. 北京:高等教育出版社,2000.
- [2] 王晋年,张 兵,刘建贵,等. 以地物识别和分类为目标的高光谱数据挖掘[J]. 中国图形图像学报,1999,11(4A版):957~964.
- [3] 甘肅平,王润生,郭小方,等. 高光谱遥感信息提取与地质应用前景[J]. 国土资源遥感,2000(3):38~44.
- [4] 袁震东,洪 渊,林武忠. 数学建模[M]. 上海:华东师范大学出版社,1999.
- [5] 施吉林,刘淑珍,陈桂芝. 计算机数值方法[M]. 北京:高等教育出版社,1999.
- [6] 郭得方. 遥感图像计算机处理和模式识别[M]. 北京:电子工业出版社,1987.
- [7] 楼顺天,施 阳. 基于matlab的系统分析与设计[M]. 西安:西安电子科技大学出版社,1999.
- [8] 刘素红,蔺启忠,燕守勋. 面向成矿分析的黄金遥感数据库归一化模型[J]. 地质与勘探,2000,36(3):40~43.
- [9] Haskett, H Hanna Tran. Spectral - Spatial automatic target detection of small targets using hyperspectral imagery[D]. PhD paper, Gerge Mason University, 2001.
- [10] Du, Qian. Topic in hyperspectral image analysis[D]. PhD paper, University of Marland Baltimore County, 2000.
- [11] Swayze G A, R N Clark, K S Smith, et al. Using Imaging Spectroscopy to Cost - Effectively Locate Acid - Generating Minerals at Mine Sites: An Example From the California Gulch Superfund Site In Leadville, Colorado[J]. Geological Society of America Abstracts with Programs, October 20 ~ 23, 1997, 29(6): A ~ 322.

STUDY ON SPECTRAL MODELING, CLASSIFICATION AND DETECTION TECHNOLOGY OF THE IMAGING SPECTRUM DATA——URANIUM MINERALIZING AREA OF YUNNAN TENGCHONG AS A TESTING SAMPLE

ZHANG Jie - lin^{1,2}, CEN Chang - hua¹, NIU Zheng³, LU Shu - ning⁴, CAO Dai - yong²

(1. Beijing Geological Institute of Nuclear Industry, Beijing 100029; 2. China University of Mining & Technology, Beijing 100083; 3. Institute of Remote Sensing Application, Chinese Academy of Science, Beijing 100101; 4. National Satellite Application Center, Beijing 100055)

Abstract: Imaging spectrum technology which characterizes by high resolution and quantitative analysis is one of front areas and research hotspots in the remote sensing. It can improve remote sensing technology from qualitative analysis to quantitative detection phase. On the base of the characteristic of imaging spectrum data, and adopting correlation coefficient spectral matching methods, Guassian modified model, artificial neural net (ANN) classification algorithm, this article succeeded in analyzing spectral signature of rock mineral, detecting and modeling the diagnosis spectral information, and improving the speed and precision of classification algorithm. Using the models and methods, uranium mineralizing area of Yunnan Tengchong was studied, and better processing results were achieved.

Key words: Imaging spectrum, spectral modeling, spectral matching, artificial neural net (ANN), classification and detection algorithm