

国家“863”计划“地球观测与导航系统领域项目”(2009AA12Z117)资助

智能算法在高光谱 遥感数据处理中的应用

*Intelligent Algorithms and Their Applications in
Data Processing of Hyperspectral Remote Sensing*

龚文引 蔡之华 杨鸣 著



中国地质大学出版社
CHINA UNIVERSITY OF GEOSCIENCES PRESS

智能算法在高光谱遥感数据 处理中的应用

Intelligent Algorithms and Their Applications in Data Processing of
Hyperspectral Remote Sensing

龚文引 蔡之华 杨 鸣 著



中国地质大学出版社
CHINA UNIVERSITY OF GEOSCIENCES PRESS

图书在版编目(CIP)数据

智能算法在高光谱遥感数据处理中的应用/龚文引,蔡之华,杨鸣著.—武汉:中国地质大学出版社,2014.11

ISBN 978-7-5625-2587-5

I. ①智…

II. ①龚… ②蔡… ③杨…

III. ①人工智能-算法理论-应用-光谱分辨率-光学遥感-遥感数据-数据处理

IV. ①TP722 ②P183

中国版本图书馆 CIP 数据核字(2014)第 013975 号



智能算法在高光谱遥感数据处理中的应用

龚文引 蔡之华 杨 鸣 著

责任编辑:王凤林 张晓红

责任校对:张咏梅

出版发行:中国地质大学出版社(武汉市洪山区鲁磨路 388 号)

邮政编码:430074

电 话:(027)67883511

传 真:67883580

E-mail: cbb@cug.edu.cn

经 销:全国新华书店

<http://www.cugp.cug.edu.cn>

开本:787 毫米×1 092 毫米 1/16

字数:500 千字 印张:19.25

版次:2014 年 11 月第 1 版

印次:2014 年 11 月第 1 次印刷

印刷:武汉市教文印刷厂

印数:1-500 册

ISBN 978-7-5625-2587-5

定价:78.00 元

如有印装质量问题请与印刷厂联系调换

前 言

进入 21 世纪后,世界范围内对地观测战略中同一地区不同时相、不同分辨率、不同成像机理的遥感数据呈指数增长。高光谱已细分到数百个甚至上千个波段,如何从这些海量的数据中获取有用的信息是遥感应用的重要问题。

笔者在书中以承担的国家“863”计划项目“基于演化数据挖掘的铀矿床高光谱遥感信息自动提取技术研究”的研究成果为基础,探讨智能算法(包括基于实例的学习方法、决策树、贝叶斯分类算法、人工神经网络算法、演化算法、支持向量机算法、模型树等)在高光谱数据处理中的应用。

全书共包含两个部分的内容,第一部分是高光谱数据处理基础,包括高光谱遥感基本知识、高光谱数据的获取与分析、铀矿物的光谱特征、高光谱数据的处理模型等。第二部分是本书的重点,对常见的数据挖掘方法:基于实例的学习方法、决策树算法、贝叶斯分类算法、人工神经网络、演化算法、支持向量机算法、模型树算法及其在高光谱数据处理中的应用进行了全面探讨,对演化算法及其改进进行了重点研究,研究成果对演化算法应用具有重要的启示作用。

本书是课题组研究成果的总结,具体由下列人员完成。

第一部分:第 1 章、第 2 章由龚文引、蔡之华、吴佳、赵玲编写;第 3 章由龚文引、蔡之华、杨鸣、张方编写;第 4 章由龚文引、蔡之华、杨鸣、胡振邦编写;第 5 章由龚文引、蔡之华、吴亦奇、敖霜编写。

第二部分:第 6~9 章由龚文引、蔡之华、杨鸣、吴佳、陈晓琳、高哲超编写;第 10 章由龚文引、蔡之华、杨鸣、吴佳、余超、姜爱晓、童益萍、南丽玲编写;第 11 章由龚文引、蔡之华、杨鸣、余小陆编写;第 12 章由龚文引、蔡之华、吴佳编写;第 13 章由吴佳、张耀东编写,并且龚文引、蔡之华、杨鸣对全书进行了统一编审。

尽管本书是课题组精心研究的成果,但由于水平有限,不免存在许多不足之处,欢迎同行指正。

作 者

2013 年 12 月 28 日

目 录

第一部分 高光谱数据处理基础

1 绪论	(3)
1.1 高光谱遥感原理	(3)
1.2 高光谱遥感的发展历史	(3)
1.3 高光谱遥感数据的分类	(4)
1.4 高光谱卫星遥感光谱测量	(5)
1.5 高光谱地面光谱测量	(6)
1.6 高光谱遥感数据处理方法	(6)
2 高光谱数据的获取与分析	(8)
2.1 地物光谱数据库的建立	(8)
2.2 高光谱地面实测数据的预处理	(10)
2.3 地物光谱特征参数	(12)
2.4 地物光谱建模分析	(14)
2.5 卫星遥感高光谱数据的获取	(14)
2.6 卫星遥感数据的存储格式	(15)
2.7 卫星遥感图像的预处理	(17)
3 常见铀矿物(实验数据)的光谱特征.....	(19)
3.1 常见原生铀矿物的光谱特征	(19)
3.2 常见次生铀矿物的光谱特征	(21)
4 遥感数据的可视化.....	(24)
4.1 视觉感知要素	(24)
4.2 图像增强简介	(25)
4.3 基本灰度变换	(27)
4.4 直方图处理	(31)
4.5 图像局部增强	(34)
4.6 二阶微分空间域滤波	(36)
4.7 图像融合技术	(38)

5 高光谱数据的处理模型·····	(45)
5.1 高光谱地面实测数据矿物识别模型·····	(45)
5.2 高光谱地面实测数据物谱关联模型·····	(51)
5.3 高光谱地面实测数据特征子集提取·····	(54)
5.4 高光谱卫星遥感图像波段选择模型·····	(59)
5.5 高光谱卫星遥感图像分类模型·····	(65)

第二部分 高光谱数据处理的智能算法

6 基于实例的学习方法·····	(75)
6.1 引言·····	(75)
6.2 K 近邻分类器·····	(75)
6.3 K 近邻回归预测器·····	(76)
6.4 K 近邻算法在高光谱遥感数据处理中的应用·····	(76)
7 决策树·····	(83)
7.1 引言·····	(83)
7.2 决策树的表示方法·····	(83)
7.3 决策树的属性度量问题·····	(83)
7.4 决策树方法在高光谱地面实测数据矿物识别中的应用·····	(86)
7.5 决策树方法在高光谱卫星遥感图像分类中的应用·····	(89)
8 贝叶斯分类·····	(93)
8.1 引言·····	(93)
8.2 贝叶斯法则与分类器·····	(93)
8.3 贝叶斯在高光谱遥感数据处理中的应用·····	(95)
8.4 贝叶斯与决策树的混合模型及其遥感应用·····	(99)
9 人工神经网络·····	(102)
9.1 引言·····	(102)
9.2 反向传播网络基本原理·····	(102)
9.3 神经网络的基本特性·····	(104)
9.4 BP 网在高光谱遥感数据处理中的应用·····	(105)
10 演化算法·····	(114)
10.1 引言·····	(114)
10.2 演化算法的基本原理·····	(115)

10.3	基本演化算法	(119)
10.4	差分演化算法	(127)
10.5	遗传规划	(136)
10.6	基于基因表达式的演化算法	(147)
11	支持向量机	(158)
11.1	引言	(158)
11.2	支持向量机分类	(158)
11.3	支持向量机回归	(162)
11.4	SVM 在高光谱遥感数据处理中的应用	(164)
11.5	改进的支持向量机及其遥感应用	(168)
12	其他智能回归算法	(172)
12.1	引言	(172)
12.2	模型树算法	(172)
12.3	线性回归算法	(172)
12.4	遥感应用	(174)
13	高光谱数据处理与分析系统	(177)
13.1	高光谱数据的处理系统的组成	(177)
13.2	国外高光谱遥感数据处理软件	(178)
13.3	国内高光谱遥感数据处理软件	(184)
13.4	基于智能算法的高光谱数据处理软件功能介绍	(186)
附 录		(188)
附录 1:	GEP 回归实现	(188)
附录 2:	GP 回归实现	(216)
附录 3:	SMO 回归实现	(234)
附录 4:	DESMO 分类实现	(252)
附录 5:	特性指标测试结果明细	(281)
附录 6:	地物分类表	(286)
附录 7:	光谱数据库表结构	(289)
参考文献		(297)

第一部分

高光谱数据处理基础

1 绪论

1.1 高光谱遥感原理

遥感(Remote Sensing)是20世纪60年代发展起来的一种从远处探测、感知物体或事物的对地观测技术,即不直接接触物体本身,而是从远处通过各种传感器探测和接收来自目标物体的信息,经过信息的传输和处理分析,来识别物体的属性及其分布等特征。作为一门新兴的、建立在现代物理学、电子计算机技术、数学方法基础上的学科,遥感经过几十年的发展,已经在地理学、地质学、生态学、环境科学以及农业和海洋科学等学科领域得到了广泛的研究和应用,成为当今最活跃的科技领域之一。

高光谱遥感是20世纪80年代兴起的新型对地观测技术。高光谱遥感具有不同于传统遥感的新特点。

- (1) 波段多。成像光谱仪在可见光和近红外光谱区内有数十甚至数百个波段。
- (2) 光谱分辨率高。成像光谱仪采样的间隔小,一般为10nm左右。
- (3) 数据量大。随着波段数的增加,数据量呈指数增加。
- (4) 信息冗余增加。由于相邻波段的相关性高,信息冗余度增加。

一般认为,光谱分辨率在 $10^{-1}\lambda$ 数量级范围内的遥感称为多光谱遥感,在 $10^{-2}\lambda$ 数量级范围内的遥感为高光谱遥感。高光谱遥感的基础是波谱学,当电磁波入射到物质表面时,物质内部的电子跃迁,原子和分子的振动、转动等作用使物质在特定的波长形成特有的吸收和反射特征,能够通过物质的反射或吸收光谱反映出物质的组成成分与结构的差异。

1.2 高光谱遥感的发展历史

自20世纪80年代以来,美国已经研制了三代高光谱成像光谱仪。1983年,第一幅由航空成像光谱仪(AIS-1)获取的高光谱分辨率图像的正式出现标志着第一代高光谱分辨率传感器面世。第一代成像光谱仪(AIS),由美国国家航空和航天管理局(NASA)所属的喷气推进实验室设计,分为AIS-1和AIS-2两种,有128个通道,光谱覆盖范围为 $1.2\sim 2.4\mu\text{m}$ 。1987年,由NASA喷气推进实验室研制成功的航空(可见光)红外光成像光谱仪(AVIRIS)成为第二代高光谱成像光谱仪的代表。它是首次测量全部太阳辐射光谱范围的成像光谱仪,共有224个通道。与此同时,加拿大、澳大利亚、日本等国家也竞相投入力量研究成像光谱仪。第三代高光谱成像光谱仪为克里斯特里尔傅立叶变换高光谱成像仪(FTHSI),其质量仅为35kg,采用256通道,光谱范围为 $400\sim 1\,050\text{nm}$,光谱分辨率为 $2\sim 10\text{nm}$,视场角为 150° 。在国内,成像光谱仪的研制工作也取得了长足的发展,紧跟国际前沿技术,目前已跻身国际先进

行列。先后研制成功了专题应用扫描仪 IR/UV 和 VIS/MIR/IR、6 波段红外细分光谱扫描仪 FIMS、热红外多光谱扫描仪 ATIMS、19 波段多光谱扫描仪 AMSS、71 波段的模块化航空成像光谱仪 MAIS、128 波段实用型模块化机载成像光谱仪 OMIS 以及我国第一台 224 波段推扫式高光谱成像仪 PHI 等。此外,中国科学院上海技术物理研究所研制的中分辨率成像光谱仪于 2002 年随“神舟”三号飞船发射升空,使得中国成为继美国 1999 年发射 EOS 平台之后第二次将中分辨率成像光谱仪送上太空的国家。

1.3 高光谱遥感数据的分类

在遥感的众多应用领域中,利用遥感数据进行地物的分类是其中一个重要组成部分,即如何根据获取的遥感信息对地物目标的物理属性进行识别。虽然经过几十年的研究和发展,遥感数据分类具备了十分丰富的理论和方法,并在多光谱遥感的应用中获得成功,但仅仅依靠多光谱遥感提供的信息是远远不够的。高光谱遥感的出现为我们开展地物理化特性的深层探索,进行地物间微小差异的精细识别,以及发现自然界的知识,都提供了前所未有的丰富信息。目前,应用高光谱遥感数据进行地物分类的方法主要分为两大类。

(1) 直接在原始高维光谱数据上采用一些特殊的处理方法和技術,如光谱匹配技术(OSP)等。光谱匹配技术主要是通过研究两条光谱曲线的相似度来判断地物的归属类别,主要分为两种方法:①利用像元的光谱曲线与用于参照的标准地物的光谱曲线进行对比;②利用像元之间的曲线相似度进行聚类分析,得出大致相加的地物分类。其特点为不对高光谱数据进行特征选择,只提取中间某些有用的信息来进行比较,利用不同地物高光谱数据之间的异同来区别地物,如图 1-3-1 所示。

(2) 通过某种特征提取算法对原始数据进行特征提取,在低维空间内利用已经成熟的多



图 1-3-1 遥感影像光谱匹配分类结果

光谱分类方法进行提取。特征提取算法主要分为过滤和封装两类,其中过滤式特征提取算法是使用与分类算法无关的评价标准(如信息熵)和搜索算法结合对候选空间中的特征进行智能提取,封装式特征提取算法是利用与分类算法分类结果有关的评价标准(如分类精度)和搜索算法进行特征的智能提取。过滤式特征提取算法速度较快,但是有一定的盲目性,如果没有合适的评价标准则无法选择出较好的特征子集;封装式特征提取算法由于是基于某种分类算法来进行评价的,因此能得到较好的分类结果,但是需多次进行分类建模,所以效率较低。

1.4 高光谱卫星遥感光谱测量

1.4.1 成像光谱仪

成像光谱就是在特定光谱域以高光谱分辨率获得连续的地物光谱图像,这使得遥感应用可以在光谱维上进行空间展开,定量分析地球表层生物、物理、化学过程和参数。它在结构和成像原理方面与光机扫描仪、推扫式扫描仪有很多相似点。成像光谱有以下4个方面的突出特点:①高光谱分辨率,成像光谱仪能获得整个可见光、近红外、短波红外、热红外波段的多而窄的连续光谱波段;②图谱合一,成像光谱仪在获得数十、数百个光谱图像的同时,可以显示影像中每个像元的连续光谱;③空间分辨率,航空成像光谱仪均有较高的空间分辨率;④辐射分辨率与信噪比,研究表明,信噪比的高低直接影响成像光谱图像对地物的识别能力。

高光谱成像光谱仪的成像方式主要有两种类型。第一种是线阵列探测器推扫式扫描方式。它通过旋转平面镜沿垂直飞行方向的来回摆动,将地面不同目标的辐射能量,通过物镜传到色散装置,经过色散元件的分色,产生不同波长的辐射能量,该能量再经过像镜汇聚,由线阵列探测器记录下来,形成高光谱数据。第二种是面阵列探测器推扫式扫描方式。它一次扫描同时收集垂直飞行方向的一行地面数据,经过物镜透射到色散元件上,经过分色,产生不同波长的数据,再经过像镜汇聚,由面阵列探测器记录下来。

目前使用比较广泛的成像光谱仪主要有中等光谱分辨率成像光谱仪 MODIS、ARIES、实用型模块化成像光谱仪 OMIS、推扫式光谱成像仪 PHI 以及航空可见光/红外成像光谱仪 AVIRIS 等。

1.4.2 卫星遥感数据的格式

卫星遥感数据常以不同的数据格式存储于磁带或光盘中。主要有以下4种数据格式。

(1) HDF 格式(Hierarchical Data Format):HDF 是一种不必转换格式就可以在不同平台间传递的新型数据格式。HDF 常指代 HDF4,它是该数据格式的第四个版本。美国国家航空航天局的地球观测系统中,记录全球温度变化的数据主要采用的就是这种数据存储格式。

(2) BSQ 格式(Band Sequential Format):按波段顺序记录图像数据,即在连续的存储空间中存储着一整幅图像的某一波段的相应数据,最适合用于单波段数据读取。

(3) BIL 格式(Band Interleaved by Line):按扫描行顺序记录图像数据,即在连续存储空间中存储某一行数据在所有波段上的相应数据,适合单行多波段数据读取。

(4) BIP 格式(Band Interleaved by Pixel):按像元顺序记录图像数据,即在连续存储空间中存储某一像元在所有波段上的相应数据,适合单像元波谱存储和读取。

1.5 高光谱地面光谱测量

1.5.1 野外光谱仪

在遥感领域中,为了研究各种不同地物或环境在野外自然条件下的可见和近红外波段反射光谱,需要适用于野外测量的光谱仪器,对野外地物光谱进行测量。目前使用的是美国 ASD 公司 FieldSpec Pro,它是一种测量可见光到近红外波段地物波谱的有效工具。下面主要介绍便携式地物光谱仪 FieldSpec Pro FR。

功能及应用范围:作为野外光谱采集仪器,对各种地物,如植被、建筑物表面、水体、岩石等进行光谱数据采集,同时也能为相关研究领域的光谱数据库的建立提供原数据等。

光谱仪波长范围为 $0.35\sim 2.5\mu\text{m}$,波长精度为 $\pm 1\text{nm}$,且具有较高的抽样间隔和光谱分辨率、采样速度快、操作简单、存储容量大等特点,是野外光谱测量最理想的选择。技术指标:光谱范围 $350\sim 2\,500\text{nm}$, $350\sim 1\,050\text{nm}$;最快采集速度 100ms , 17ms ;光谱分辨率 $3\text{nm}@350\sim 1\,000\text{nm}$, $10\text{nm}@1\,000\sim 2\,500\text{nm}$;采样间隔 $1.4\text{nm}@350\sim 1\,000\text{nm}$, $2\text{nm}@1\,000\sim 2\,500\text{nm}$;数据间隔 1nm ;波长精度 $\pm 1\text{nm}$;波长重复性 $\pm 0.02\text{nm}$;重复性优于 0.3% ;杂散光优于 $0.02\% @ 350\sim 1\,000\text{nm}$, 优于 $0.1\% @ 1\,000\sim 2\,500\text{nm}$;全自动重量大约 8.5kg (1.2kg 含电池)。

1.5.2 地面数据的测量方法

美国 ASD 公司设计制造的 FieldSpec 系列光谱仪在中国的遥感应用已经相对普及,应用范围已经扩展到包括精准农业、林业、海洋与内陆水体、冰雪、环境污染监控、气象、地质与矿产等领域。下面简单介绍 FieldSpec JR 光谱仪进行地物数据测量的过程。

- (1) 光谱仪、计算机充电。
- (2) 光谱仪电量不足时红灯闪亮,充满电后绿灯亮;如果黄灯闪亮则说明过热,需等待一段时间后继续充电。
- (3) 安装适当的镜头或其他附件(如 GPS、余弦接受器等),准备好白板。
- (4) 打开光谱仪电源,然后打开计算机电源,并启动 RS3 软件。
- (5) 在软件上选择相应的镜头。
- (6) 点击 OPT 优化(注意:工作过程中特别是开始工作的前半个小时内每隔一定时间做一次优化)。
- (7) 在软件中选择或填写需要存储数据的路径、名称和其他内容。
- (8) 开始测量。

1.6 高光谱遥感数据处理方法

1.6.1 传统高光谱数据处理方法

传统的高光谱数据处理方法主要应用在两个方面。一种是高光谱分类,高光谱分类的目标是根据测得的光谱数据,实现对地物类别的计算机自动识别。传统的分类方法是从模式识

别中发展起来的一些分类技术,如最小距离分类、最大似然分类、神经网络分类等。另一种是高光谱物谱建模,高光谱物谱建模就是对蚀变矿物含量与光谱吸收特征参数(光谱吸收特征参数是反射光谱线的吸收峰所对应的各参数)进行相关分析,建立蚀变矿物含量与光谱吸收特征深度、面积等参数之间的物谱关联模型,从而可以根据光谱数据预测矿物成分的含量。传统的高光谱物谱建模技术主要是统计回归分析、散点图等。

1.6.2 高光谱数据智能处理方法

由于高光谱具有波段数众多、信息冗余度高等特点,一些适用于多光谱遥感数据的处理分析技术已不能适应高光谱数据的要求,因此迫切需要一种高效率、智能化的高光谱遥感数据处理技术。数据挖掘技术就是从大量的、不完全的、有噪声的高光谱数据中,去提取隐含在其中的、人们事先不知道的、但又是潜在有用的信息和知识的过程。

数据挖掘技术经过几十年的发展,已经形成了以关联分析、分类分析、聚类分析等为主要形式的应用技术,各种分类挖掘方法也不断得到完善和改进,如以信息论为基础的决策树分类、以概率论为基础的贝叶斯分类、以神经科学为基础的神经网络方法等。研究和运用适合于高光谱数据的分类挖掘技术和方法,建立物谱关联模型、光谱分类等模型来实现对铀矿床高光谱数据的高效处理分析,通过对地物混合成分丰度预测估计、岩石矿物物理化学成分与光谱特征参数之间的定量关系研究,进而从大量的高光谱遥感数据中提取岩石和蚀变矿物的光谱信息,然后根据测得的光谱数据,实现对地物类别的计算机自动识别。

2 高光谱数据的获取与分析

2.1 地物光谱数据库的建立

2.1.1 地物光谱数据库数据的组成

中国铀矿床光谱数据库中的光谱数据主要来自于核工业北京地质研究院承担的“遥感卫星中国铀矿床光谱特征研究与最佳谱段选择”项目的支持,其主要包括矿物、矿石、岩石、土壤、植被、人工、水体、冰雪、大气共9个专业子数据库。数据库系统中的任何数据库建设都遵循系统数据采集与录入规范、建库规范,遵循相应的国家标准和行业标准,计算机软件开发标准。这些子数据库既是整个光谱数据库的有机组成部分,又能各司其职。详细各表之间的关系参阅附件。

光谱数据库涉及的内容主要包括以下方面。

- (1) 样本基本属性(包括样本编码、样本名称、样本类别、数据来源方式等)。
- (2) 样本光谱数据(包括测量日期、测量时间、总波段数、波长及相应的反射率等)。
- (3) 样本特征(包括样本岩石特性或植被特性等)。
- (4) 样本化学成分(包括样本所含的化学成分名称及其含量等)。
- (5) 样本图像(包括样本的遥感图像、实地图像或显微镜下图像等)。
- (6) 测量条件(包括测量时的天气状况、采集人员、质检人员、测量仪器、测量距离、测量倾角、测量方位角等)。
- (7) 测量位置(包括测量地点的测线编号、测线起始经纬度、测点经纬度等;对于矿床相关数据还包括矿床的相关信息)。
- (8) 光谱测量仪器的各类参数(包括仪器名称、仪器型号、标定日期等)。
- (9) 光谱特征数据(根据样本地物类别的不同,包括不同的光谱特征数据,如对于岩矿样本,包括吸收峰位置、面积、宽度、对称度等)。
- (10) 各地物分类的编码、名称等基本信息。
- (11) 各个用户的基本信息(包括用户名、用户密码、用户权限等)。

2.1.2 地物光谱数据库建立的标准

中国铀矿床光谱数据库的建立是“基于演化数据挖掘的铀矿床高光谱遥感信息处理软件”的基础,系统所用到的所有的(包括原始的和已经处理分析后的)数据都将保存在数据库中,以增强对各类光谱及相关数据的管理和处理分析。针对“基于演化数据挖掘的铀矿床高光谱遥感信息处理软件”的要求和特点,我们提出了“规范化、专业化、可扩充性、并行处理性”4个标准建立了“中国铀矿床光谱数据库”。

- (1) 规范化:光谱数据库的建立要符合国家及行业规范,数据库的构造方式、结构和设计

要严格按照各类规范进行,数据库的专业术语、编码方式等要符合相关行业的通用标准。

(2) 专业化:根据数据库的应用模式,光谱数据库的设计凸显花岗岩型铀矿光谱数据方面的特色,能提供更好的数据库结构和内容来支持花岗岩型铀矿光谱数据,更有利于针对花岗岩型光谱数据的查询、处理和分析。

(3) 可扩充性:在光谱数据库的设计上,除较好地支持现有功能外,还提供更灵活的数据库结构以便于光谱数据库内容以及应用领域的扩充,使之能适应更多的变化和更广泛的应用领域。

(4) 并行处理性:光谱数据库支持多个用户同时对其中的各种光谱及其相关数据进行管理和处理分析,给用户提供更灵活、便捷的数据管理和数据处理分析方式。

2.1.3 地物光谱实测数据的基本数据库

(1) 实测岩矿光谱数据子库(包括岩石、矿石和矿物)。用于保存各类实测岩矿光谱数据(包括野外实测、馆藏、单矿物以及混合粉样等多类型的数据)及相应岩矿样本的基本属性和特征(包括实测岩矿光谱波长、反射率,实测岩矿样本编号、名称及相关各类物性描述和化学成分参数)。

(2) 参考岩矿光谱数据子库。能将其他各类地物光谱库中相关数据转化为一定形式后保存到库中(包括参考光谱的波长、反射率,参考岩矿样本的名称及相关部分物理特征)。

(3) 植被光谱数据子库。能保存各类植被光谱数据及相应植被的基本属性和特征(包括植被光谱波长、反射率,植被样本编号、名称及各类物理特征)。

(4) 岩矿及植被光谱特征表。能保存各类岩矿和植被光谱数据处理分析后提取出来的各类光谱特征参数(包括吸收波段位置、反射率、深度、宽度、斜率等各类特征)。

(5) 岩矿及植被光谱测量辅助信息表。能保存各类岩矿和植被样本进行光谱测量时的测量地点、测量时间、天气、测量人员等相应光谱测量辅助信息。

(6) 其他相关信息子库。包括矿床信息情况表、测量仪器情况表等,能保存与光谱数据相关的其他信息。

(7) 土壤、水体、人工、冰雪、大气以及航天航空光谱数据子库的大体框架,为日后光谱数据库的扩充做好基础工作。

2.1.4 常见的光谱库

当前国际上常用的光谱库有 6 个,其中公开提供电子版的机构有 USGS、JPL、JHU、IGCP-264、ASTER。

(1) USGS 光谱库是美国国家地质调查局 USGS(United State Geological Survey)光谱实验室在 1993 年建立的波长为 $0.2\sim 3.0\mu\text{m}$ 的光谱库,目前光谱库版本是 Splib04,包含 444 种矿物的反射波谱,还有几十种植被和水体的反射光谱,光谱分辨率为 4nm (可见光波段 $0.2\sim 0.8\mu\text{m}$)和 10nm (近红外波段 $0.8\sim 2.35\mu\text{m}$)。

(2) JPL 光谱库是美国喷气推进实验室(JPL,Jet Propulsion Laboratory)用 Beckman 系列光谱仪对 160 种不同粒度的常见矿物进行测试,并同时进行 X 光测试分析,最后按照小于 $45\mu\text{m}$ 、 $45\sim 125\mu\text{m}$ 、 $125\sim 500\mu\text{m}$ 3 种粒度,分别建立了 3 个光谱库:JPL1、JPL2、JPL3。JPL 光谱库中有 1 348 种矿物的反射光谱,24 种岩石的反射光谱,另外,还包括土壤、植被、水、雪、冰和人造物质的反射光谱。

(3) 约翰霍普斯大学(JHU)提供了包含 15 个子库的光谱库,针对不同的地物类型选用了不同的分光计,并且每种地物都给出了详细的文本介绍。其中 $2.03\sim 2.5\mu\text{m}$ 的光谱数据是用

Beckman UV-5240 仪器测试得到的, $2.08 \sim 15 \mu\text{m}$ 的光谱数据是用 FTIR 仪器测试得到的。

(4) 美国在 IGCP-264 工程实施过程中, 为了比较光谱分辨率和采样间隔对光谱特征的影响, 对 26 种样本采用 5 种分光计测试, 并同时进行了 EDS、SEM、XRD 分析测试, 最后建成了 5 个光谱库。

(5) ASTER 光谱库是 2000 年 5 月加利福尼亚技术研究所建立的, 该库包括了 USGS、JPL、JHU 3 个光谱库的所有光谱。ASTER 光谱库目前具有近 2 000 条光谱, 包括矿物、岩石、土壤、人造材料、水和雪, 许多光谱覆盖了 $0.4 \sim 1.4 \mu\text{m}$ 的波长范围。这个库通过网络 <http://speclib.jplnasa.gov> 就可交互使用。

2.2 高光谱地面实测数据的预处理

以下内容均以样本编号为 Co33-V-2-10、样本名称为沥青铀矿的样本来进行说明。

2.2.1 反射率光谱数据的校验

为了确保获得精确可靠的光谱数据, 通常对每一测量样本要获取 5~6 次采样数据, 再取其平均值作为该测量样本的光谱数据。这组数据中有时会出现个别离散现象, 采用均方差检验方法将它剔除, 并使它不参加平均值计算。以同一样本的一组原始光谱数据(5 或 6 条光谱曲线)作为输入, 以光谱数据的平均值作为输出(图 2-2-1)。

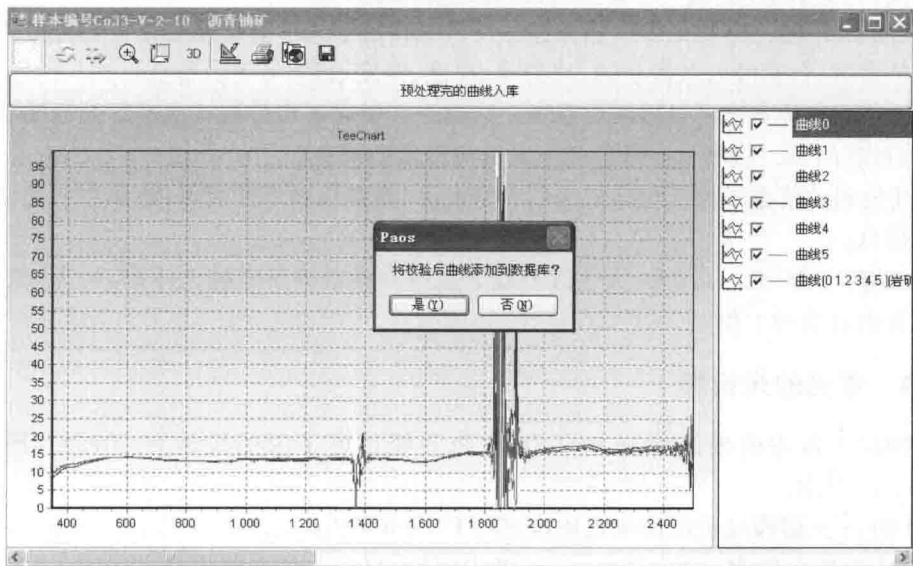


图 2-2-1 地面实测光谱数据库反射率光谱曲线的校验

2.2.2 反射率光谱曲线的光滑

实测的光谱数据往往包含 3 种误差: ①由于光谱仪是由三个不同光谱范围的探头组成, 它们之间会出现台阶; ②随机误差, 导致反射光谱曲线出现一些小抖动的现象; ③水汽吸收带的杂乱跳动。通常多项式是既简单又有效的光滑实测数据的方法, 选用五点距离平方加权法对光谱数据进行光滑; 对于杂乱跳动的水汽吸收带, 根据变化趋势采用二次曲线拟合, 即用二次多项式插值的方法实现曲线光滑; 对于“台阶”, 关键是识别出“台阶”的位置, 再求出“台阶”的