

岩石实验室光谱对应分析

刘庆生 王志刚 荆林海

(中国科学院遥感应用研究所 北京 100101)

摘要 成像光谱图像具有几十个窄波段和高的光谱分辨率,使直接识别岩石岩性成为可能,但其海量数据又为数据的处理、分析和应用带来了不便。因此就需要寻求一种在不损失信息或最大限度保持原有信息的前提下又压缩或最小化数据量的方法。在对岩石实验室光谱按成像光谱仪(MAIS)通道重采样后,应用了一种特征信息提取的方法(对应分析法),对岩石光谱进行了分析,达到了压缩数据与岩石分类的目的,同时也探讨了其处理大数据量成像光谱图像数据的意义。

关键词 岩石实验室光谱,对应分析

1 引言

当所用的光谱范围较宽、波段数多时,由于波段之间的相关性,人脑已不足以区分不同的矿物和岩石,要想区分它们,就必须知道它们各自的诊断性光谱波段。而波段间相关性强,导致许多分类和识别算法很费时且精度降低,为了解决这一问题,人们发展了许多不同的最佳波段选择算法^[1],主要有:最优指示因子法^[2];主成分分析法^[3];最大方差-协方差矩阵决定法^[4];转换扩散指示法^[5];空间自相关分析法^[6];分类精度比较法^[7]。其中主成分分析法是一种基于方差、协方差矩阵和距离的统计方法,该方法能很好地压缩数据,同时能对图像进行信息的集中^[8-11]。例如,TM主成分变换后,第一主成分主要包含了亮度信息,主成分变换后的各主成分与原始波段的关系可从主成分矩阵得到,但由于变换后各主成分的物理意义不太明确,因而人们对它的使用并不乐观。正是基于上述原因,作者试图利用对应分析方法对光谱数据进行特征提取。本文旨在通过对岩石实验室光谱的分析,并提出用于分类岩石的波段范围,为将来成像光谱图像数据的处理提供参考。

2 基本原理

对应分析又称R-Q型因子分析,是由法国人Benzeci于1970年首次提出的,这种方法是在因子分

析的基础上发展起来的^[12]。前述的主成分分析方法只考虑了各个变量(波段)间的关系,并没有考虑各个变量(波段)与图像中地物(样品)的关系,因而各个主成分的物理意义不太明确。对应分析克服了这一缺点,并考虑到了进行样品(像元)因子分析时需要大的计算机内存和时间这一问题。由于在变量空间(R^p)的 N 个样品点的主成分 $PC1, PC2, \dots$,与样品空间(R^N)的 p 个变量点的主成分 $G1, G2, \dots$,在各自的总方差中的贡献(占的百分比)是相同的,因此可用同样的主成分轴去同时表示变量(波段)和样品,并可由R型分析(对变量(波段)进行变换)则很容易地获得Q型分析(对样品进行变换)的结果^[12]。因此,对应分析可提供以下信息:

(1) 变量(波段)间的关系:在主成分空间,邻近的一些变量(波段)点,表示这些变量(波段)紧密相关。

(2) 样品间的关系:在主成分空间,邻近的样品点具有相似的光谱属性,属于同一种地物类型。

(3) 变量(波段)与样品之间的关系:同一类型的样品点将为邻近的变量(波段)点所表征。具体到图像就是说,这些波段是这些样品的特征波段。同时各个主成分的物理含义也得到明确。

对应分析的数学原理^[12]如下:

设有 p 个波段的 n 个样品,其原始数据矩阵记为:

$$X = \{x_{ij}\}_{p \times n} \quad (1)$$

变换矩阵 X 为矩阵 Z :

$$Z = \{z_{ij}\}_{p \times n} \quad (2)$$

其中:

$$z_{ij} = (x_{ij} - x_i * x_j / T) / \text{sqrt}(x_i * x_j) \quad (3)$$

$$x_i = x_{i1} + x_{i2} + \dots + x_{in} \quad (4)$$

$$x_j = x_{1j} + x_{2j} + \dots + x_{pj} \quad (5)$$

$$T = x_1 + x_2 + \dots + x_p \quad (6)$$

可以证明,波段间的协方差矩阵 F 和样本之间的协方差矩阵 G 分别为:

$$F = ZZ' \quad (7)$$

$$G = Z'Z \quad (8)$$

从(7)、(8)两式可看出,矩阵 F 和 G 之间存在着简单的对应关系,故对于样本的分析,不是采用求解矩阵 G 的特征值和特征向量,而是利用 F 的结果导出

G 的特征值及特征向量,这样就克服了当样品数量太大进行 Q 型因子分析的困难,这一点对于处理图像尤为重要,因为图像处理通常要对整幅图像中的每一个像元进行分析。

3 试验分析

岩石实验室光谱是采用 GARY-5 光谱仪,测试 0.4—2.5 μm 的岩石反射光谱,选取研究区中有代表性的岩石光谱,并在地形图上精确定位,以期能与将来图像处理结果相对应。所用样品的具体数据见表 1,岩石光谱选择与 MAIS 成像光谱仪光谱通道的中间波长所对应的岩石反射率值,中间波长数值见表 1¹⁾。

表 1 内蒙大青山实验室岩石光谱分析数据

Table 1 Rock laboratory spectral analytical data of Daqingshan Mountain, the inner Mongolian Autonomous Region

Km421	Kmh21	Kmh31	Kmh9	Qm111	Qmh41	Qmh71	Sbh11	Sbh51	波长/ μm
5.71	2.59	3.22	4.20	4.02	4.26	3.51	4.17	7.21	0.45
6.41	2.60	3.53	4.40	4.52	4.19	3.79	4.39	7.90	0.51
6.82	2.74	3.67	4.60	4.85	4.29	4.05	4.41	8.05	0.53
7.42	2.98	3.96	5.03	5.22	4.51	4.40	4.39	8.14	0.55
8.45	3.35	4.67	5.94	5.68	4.88	4.87	4.40	8.21	0.57
10.41	3.91	5.89	7.58	6.35	5.40	5.56	4.39	8.24	0.63
10.55	3.89	6.02	7.65	6.35	5.40	5.64	4.30	8.20	0.65
10.71	3.88	6.17	7.76	6.37	5.43	5.69	4.21	8.17	0.67
10.90	3.92	6.28	7.95	6.40	5.52	5.78	4.16	8.16	0.69
11.63	4.28	6.71	8.79	6.70	6.12	6.37	4.21	8.53	0.77
11.75	4.19	6.72	8.77	6.77	6.11	6.47	4.26	8.66	0.79
11.77	4.10	6.72	8.76	6.79	6.06	6.45	4.27	8.69	0.81
11.66	3.94	6.60	8.60	6.72	5.87	6.39	4.22	8.59	0.83
11.41	3.79	6.38	8.28	6.53	5.59	6.25	4.08	8.37	0.85
11.39	3.76	6.30	8.20	6.43	5.37	6.26	3.99	8.32	0.87
11.50	3.82	6.24	8.24	6.44	5.28	6.29	3.93	8.39	0.89
11.58	3.88	6.18	8.29	6.47	5.23	6.31	3.92	8.43	0.91
11.65	3.95	6.16	8.34	6.50	5.20	6.35	3.89	8.43	0.93
11.73	4.30	5.90	8.75	6.47	5.35	6.45	3.78	8.10	1.03
11.48	9.04	5.30	9.46	5.98	6.33	7.40	4.97	8.43	1.630
11.43	8.72	5.27	9.36	5.93	6.27	7.39	5.03	8.39	1.660
9.91	6.77	4.67	8.21	4.93	5.52	6.31	4.80	7.05	2.030
10.07	7.02	4.64	8.25	4.92	5.62	6.43	4.91	7.07	2.070
10.10	7.11	4.57	8.25	4.84	5.65	6.47	4.85	6.92	2.100

1) 王志刚等,遥感地质岩性填图应用示范系统研究报告. 中国科学院遥感应用所. 1997.

表 1(续)

Km421	Kmh21	Kmh31	Kmh9	Qm111	Qmh41	Qmh71	Sbh11	Sbh51	波长/ μm
10.10	7.14	4.54	8.24	4.80	5.66	6.48	4.79	6.85	2.110
10.10	7.16	4.50	8.22	4.74	5.67	6.44	4.74	6.80	2.130
10.04	7.10	4.34	8.17	4.47	5.65	6.30	4.55	6.36	2.150
9.92	6.96	3.92	8.13	4.15	5.56	5.95	3.69	5.08	2.190
9.82	6.62	4.13	8.04	4.24	5.59	5.97	3.62	6.10	2.230
9.51	6.25	4.20	7.85	4.33	5.50	5.85	3.87	6.05	2.260
9.42	6.20	4.16	7.81	4.35	5.23	5.84	3.95	5.82	2.290
9.46	6.16	4.01	7.80	4.23	5.11	5.81	3.60	5.37	2.320
8.35	5.20	3.64	7.13	3.62	4.79	4.82	3.13	4.43	2.470

表中：Km421：绢云母化钾长伟晶岩脉；Kmh21：风化钾长石脉；Kmh31：黄铁绢英钾长石脉；Kmh9：石英钾长石脉；Qm111：绢云母化含角闪黑云石英钾长石脉；Qmh41：矿化石英钾长石脉；Qmh71：绿泥绢云石英钾长石脉；Sbh111：矿化绿泥绿帘绢云石榴片麻岩；Sbh51：黄铁绢英岩。

表 1 中 9 种岩石的实验室光谱曲线如图 1。

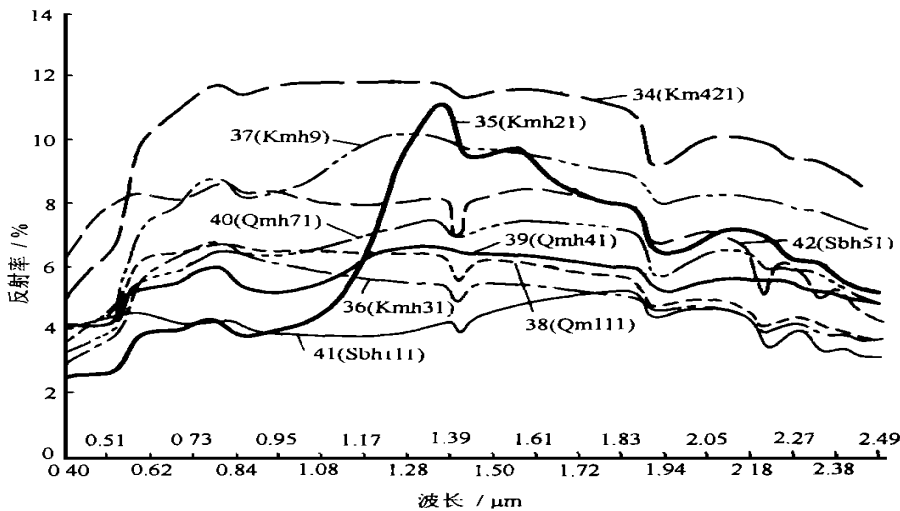


图 1 9 种岩石的实验室光谱曲线

Fig. 1 Laboratory spectral curves of the nine rocks

对原始数据进行对应分析，取所含原始数据信息量达 96% 的主成分，结果只用两个主成分就可表示原来的 33 个波段，从而可在一张图上清楚地区分各类岩石(图 2)，分类的依据就是根据二维平面图上样本的从聚性，彼此之间距离的远近程度，同时要考虑样本在图上分布的自然分类趋势和分类后样本的实际意义，因此对图 2 结果的分析必须是客观和主观的结合。

基于上述原则，分析图 2，我们可以看出：

(1) 33 个波段大致可分成 3 类：1—5 波段为 1 类，该类的特点是在第 1 和第 2 主成分上均为正值，

且值较大。从波长范围来看为 0.4—0.6 μm ，为典型的可见光波段。其中，第 1(0.45 μm)、第 5 波段(0.57 μm)在第 1 主成分上值一致，较其它 3 个波段小，正好反映 Cr, Mn, Fe⁺²在 0.55—0.57 μm 波长范围的电荷迁移吸收特征¹⁾；6—19 为第 2 类，该类在第 2 主成分上值为负，在第 1 主成分上值为正，波长范围从 0.63—1.03 μm 。其中，9—18 波段相对集中，从第 1 主成分可将它们与第 19 波段区分开，从第 2 主成分可将它们与 6, 7, 8 波段分开。9—18 波段(0.69—0.93 μm)主要反映出铁氧化矿物的吸收特征，第 19 波段反映 Fe⁺² 的吸收特征；20—33 波段为第 3 类，

波长范围(1.630—2.470 μm)该类以第 1 主成分上为负值而与前两类不同。其中,20—27 波段(1.630—2.160 μm)主要反映了水汽和 CO_3^{2-} 的吸收特征,

28—33 波段(2.190—2.470 μm)主要反映了粘土和云母等矿物的吸收特征,这两小类的区分在第 2 主成分上很容易实现。

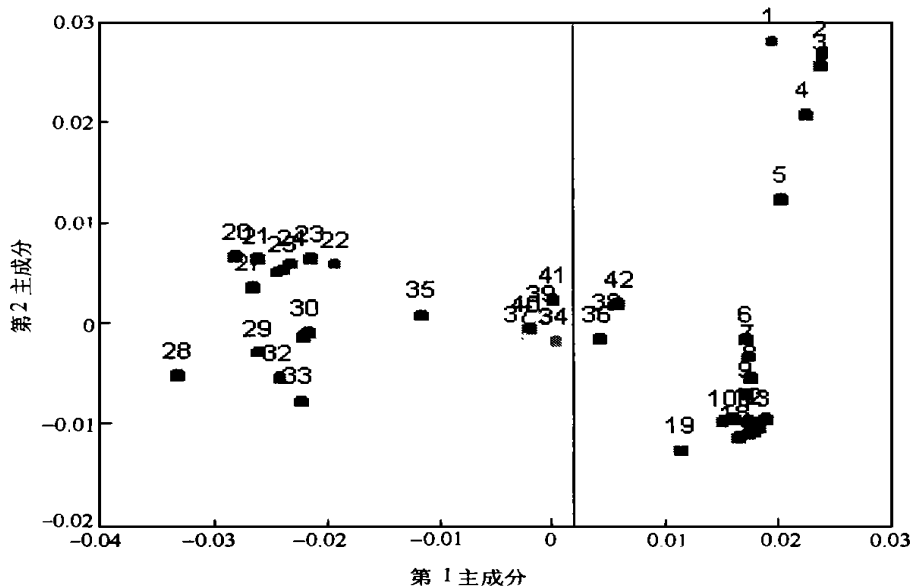


图 2 对应分析结果图

Fig. 2 The results of correspondence analysis

图中:1,2,3, ..., 32, 33 为岩石光谱曲线的 33 个波段,具体数值见表 1。34 为 Km421;35 为 Kmh21;36 为 Kmh31;37 为 Kmh9;38 为 Qm111;39 为 Qmh41;40 为 Qmh71;41 为 Sbh111;42 为 Sbh51

(2) 9 个岩石样品可分成两大类:35(Kmh21)风化钾长石脉独为一类,其余的为另一类,这从岩石光谱曲线图 1 上可得到证实。另一类中,34(Km421)可分为一类,其岩性为绢云母化钾长伟晶岩脉。由第 1 主成分很容易将其余的岩石分为两类,图 2 上 34 与 36 之间的直线,将 36(Kmh31)、38(Qm111)和 42(Sbh51)分为一类,在直线右方;37(Kmh9)、39(Qmh41)、40(Qmh71)和 41(Sbh111)在直线左侧,为另一类。由野外和室内岩矿鉴定分析,这样的区分是合理的。

(3) 波段分类与岩石分类间的关系。1—5 波段对区分上述 9 种岩性作用不大;35(Kmh21)风化钾长石脉以 20—33 波段为特征,在 6—19 波段无特殊表现,其余 8 种岩石在 6—33 波段都有表现。其中,36(Kmh31)、38(Qm111)和 42(Sbh51)以 6—19 波段为特征;37(Kmh9)、39(Qmh41)、40(Qmh71)和 41(Sbh111)以 20—33 波段为主要特征;34(Km421)在 6—33 波段上都有较大值,在图 1 中为标有 34 的粗虚线。

(4) 第 1 主成分主要代表的是 9 种岩石在 0.63—0.93 μm 内的光谱特征,36、38 和 42 号岩石在

该波段范围内有较大的值,是第 1 主成分的代表。第 2 主成分主要体现 9 种岩石在 1.630—2.470 μm 内的光谱特征,35 号岩石在此范围内具有强的光谱特征,是第 2 主成分的代表。

图 3 是对应分析后的 33 个波段和 9 个样品(图 2 中的 42 个点)的简单聚类谱系图,相似度系数为欧氏距离,利用了组间简单链接。当取相似度水平为 10 时,9 个岩石样品可分成两类 35 与其它样品分开;当取相似度水平为 5 时,其它样品又可分为两类,(34、36、38、42)和(37、39、40、41);再取较小的相似度水平,还可细分成多类。显然,图 3 和图 2 的结果是一致的。

4 结 论

综上所述,实验室岩石光谱的对应分析不仅给出了波段之间的关系,而且也给出了岩石样品之间的关系,更重要的是给出了波段和岩石样品之间的相互关系。试验中所得两个主成分集中了原始数据 96% 的信息,现在只用两个主成分就能产生原来 33 个波段的作用而将 9 种岩石样品区分开来。与主成

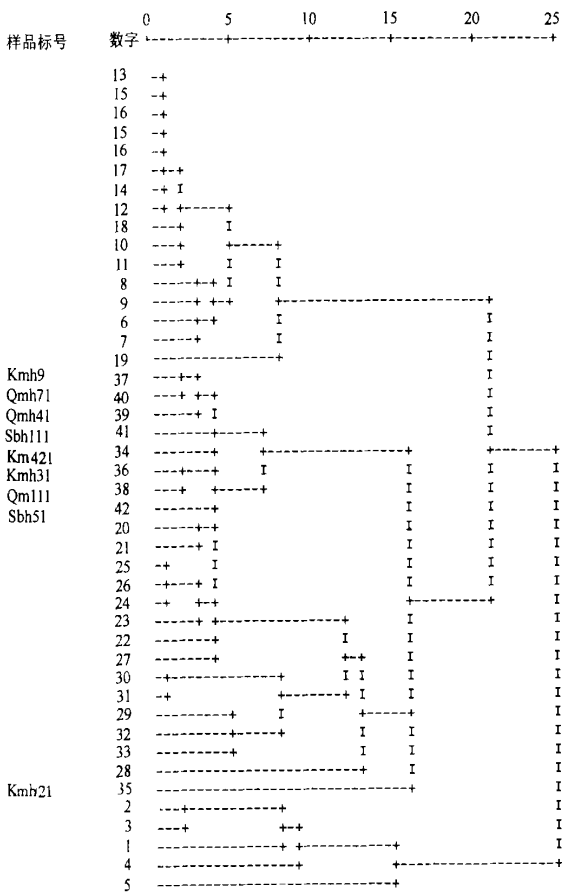


图 3 对应分析结果聚类谱系图

Fig. 3 Dendrogram of the results of correspondence analysis

分析不同的是，对应分析后主成分的含义非常明确。试验中，第 1 主成分主要代表了 9 种岩石样品在 $0.63-1.03\mu\text{m}$ 的光谱特征，36、38、42 号岩石在此主成分为正值，是第 1 主成分的主要代表。第 2 主成分主要体现出 9 种岩石在 $1.630-2.470\mu\text{m}$ 的光谱特征，其中 35 号岩石的光谱特征主要表现于此，34、37、39、40 和 41 号岩石在此波长范围比在 $0.63-1.03\mu\text{m}$ 波长范围有较强的特征，因此第 2 主成分主要代表了这 6 种岩石的光谱特征。 $0.45-0.57\mu\text{m}$ 波长范围对区分这 9 种岩石基本无效。上述分析结果表明，当利用 MAIS 成像光谱仪图像区分这 9 种岩石时， $0.40-0.70\mu\text{m}$ 范围基本无效，可不用，这样就减少了分类计算的时间和空间。另外，区分 35 号岩石，可从 20—33 波段入手，其它岩石的区分，也可从对应分析图上得到。具体到成像光谱数据，就是要有目的地反差扩展它们在主成分轴上的差别，达到图像岩石的分类，对应分析所提供的试验样品的特征波段为其它图像处理方法的实施节约了时间和空间。对应分析设计的另一优点是处理大

批量的样品节约时间，因而对基于像元的图像分类处理方法将会发挥出更大的作用。但当处理差异较小的岩石时，与其它分类方法相同，同样会遇到分多少类和临界值选择的问题。譬如图 2 和图 3，如果粗略的划分，分成 35 号和其余岩石两类即可，但这并不能满足实际的需要，因此除了依据变换后主成分空间里各样品的空间分布从聚性外，还要有具体的实际考虑，进行合理解释，这样也就难免主观因素的加入，所以进一步发展更好的定量的分类标准和合理的临界值确定方法是今后要深入研究的问题。

参 考 文 献 (References)

- 1 Meng Li, M Antrop, L Daels. Optimum band selection techniques evaluation for information extraction of the multispectral data. Proceedings of the Eleven Thematic Conference. *Geologic Remote Sensing*, 1996, 2: 119—128.
- 2 D. Greenbaum. Lithological Discrimination in Central Snowdonia Using Airborne Multispectral Scanner Imagery. *Int. J. Remote Sensing*, 1987, 8(6): 799—816.
- 3 P. W. Mausel et. al. Optimum Band Selection for Supervised Classification of Multispectral Data. *Photogrammetric Engineering and Remote Sensing*, 1990, 56(1): 55—60.
- 4 C. Sheffield. Selecting Band Combinations From Multispectral Data. *Photogrammetric Engineering and Remote Sensing*, 1985, 51(6): 681—687.
- 5 P. Gong, P. J. Howarth. The Use of Structural Information for Improving Land Cover Classification Accuracies at the Rural-Urban Fringe. *Photogrammetric Engineering and Remote Sensing*, 1990, 56(1): 67—73.
- 6 P. A. Brivio, et. al. Aspect of Spatial Autocorrelation of Landsat TM Data for the Inventory of Waste Disposal Sites in Rural Environments. *Photogrammetric Engineering and Remote Sensing*, 1993, 59(9): 1377—1382.
- 7 J. M. Stenback, Congalton. Using Thematic Mapper Imagery to Examine Forest Understory. *Photogrammetric Engineering and Remote Sensing*, 1990, 56(9): 1285—1290.
- 8 Zhang Manlang. Principal Component Analysis Used in Gold Alteration Extraction. *Remote Sensing Technology and Application*, 1996, 11(3): 1—6. (In Chinese) [张满郎. 金矿蚀变信息提取中的主成分分析. 遥感技术与应用, 1996, 11(3), 1—6].
- 9 R. S. Dwivedi, T. Ravi Sanker. Principal component analysis of LANSAT MSS data for delineation of terrain features. *Int. J. Remote Sensing*, 1992, 13(12): 2309—2318.
- 10 L. Eklundh, A. Singh. A comparative analysis of standardised and unstandardised Principal Components Analysis in remote sensing. *Int. J. Remote Sensing*, 1993, 14(7): 1359—1370.
- 11 W. P. Loughlin. Principal Component Analysis for Alteration Mapping. The Eighth Thematic Conference on Geologic Remote Sensing. Denver, Colorado, USA, 1991, 293—306.
- 12 Zhao Pengda, Li Zijing, Hu Wangliang. Statistical Inference of Mineral

Deposit· Geological publishing house, Beijing, 1983, 234—242. (In Chinese)[赵鹏大, 胡旺亮, 李紫金· 矿床统计预测· 北京: 地质出版社, 1983, 234—242].

(北京)研究生院, 获数学地质硕士学位, 现在中国科学院遥感应用研究所攻读地图学与遥感博士学位, 发表论文 1 篇。

作 者 简 介

刘庆生, 男, 1971 年 2 月生, 1996 年毕业于中国地质大学

Correspondence Analysis of Laboratory Spectra of Rock

LIU Qing-Sheng WANG Zhi-Gang JING Lin-Hai

(*Institute of Remote Sensing Applications, Chinese Academy of Sciences Beijing 100101*)

Abstract Imaging spectrometer has tens or hundreds of bands. Thus, it provides the opportunity to acquire more volume of information about the surface targets. At the same time it is difficult to process so huge volume of data, thus, it is important to find a solution to this problem. Principal Component Analysis (PCA) is a good method to reduce the volume of data at the cost of a little loss of information. But one of disadvantages of PCA is that each PCA transformed component has a blur meaning. Correspondence Analysis (CA) overcomes this shortcoming. This paper uses CA to process the Laboratory spectra of nine rocks, which have been resampled on the channels of MAIS, and get a good result. It is also a meaningful attempt to the processing of imaging spectrometer data.

Key words Laboratory spectra of rock, Correspondence analysis