

海相碳酸盐岩岩相的模糊模式识别

刘仲衡 王硕儒 范德江

(青岛海洋大学地质系)

提要 基于模糊式识别, 本文提出聚类分析与贴近度识别相结合应用于海相碳酸盐岩岩相的分类和识别, 聚类用 F-PFS 法, 识别用本文建议的一种带参数的名为加权逆距离贴近度的新公式: $S(X_i, Y_j) = \sum_{k=1}^s W_k(1 - D_{ijk})^p$ 。其中权重 W_k 使用一种按单因素判对率确定的新技术。该权重并可用于模糊综合评判。模糊特征向量采用扩展的正则化式计算。上述方法在湖北二叠系海相碳酸盐岩地层得到了应用, 并将其划分为六个相。选用 12-21 个特征因素 (颜色、矿物成分、古生物、颗粒组分、填隙物、沉积构造等), 并量化赋值。聚类分析可分出四个相, 贴近度识别的判对率达 81.8%, 按划分为六个相的贴近度识别, 确切判对率达 79.2%。

关键词 海相碳酸盐岩岩相 模糊模式识别权重 贴近度

第一作者简介 刘仲衡 男 58 岁 教授 海洋沉积学和岩石学

引言

模糊数学自 1965 年问世以来, 在不少学科得到了广泛的应用。在海洋学和地质学中也得到了初步结合, 但在沉积学方面国内外只见有海底陆源碎屑沉积物的识别和分类的几篇文章 (王硕儒等 1987; 1989; 曹宜志等 1986)。

岩相的划分有其确定性的一面, 也有其模糊性的一面。这是因为有些岩相的识别、划分难于确定, 各相邻岩相之间往往难以找到一个明确的界限。例如, 碎屑岩剖面中, 粉砂质粘土岩与砂粉粘土岩的界限就较模糊; 在海相碳酸盐岩的剖面中, 也有类似情况, 相邻环境下的不同岩相也难以划分。

不同岩相总是和不同的特征和标志相对应, 并由它们所体现。然而由综合特征体现的不同岩相更具模糊性。

用模糊数学方法建立模型和识别, 首要的任务就是正确选取特征标志和对应的权重。这是一项基础工作, 它是建模成败的关键。

所谓识别, 实质上是一个和标准模式对比的问题。如果从建立各模式的隶属函数出发对样品直接识别, 那么这时的标准模式不必用一组数来表达。但当从各样品与各标准模式的接近程度予以识别时, 则标准模式必须用模糊向量来代表。后者正是我们所采用的识别方法和思路。

一、模糊模式识别模型

1. 建立在已知样品基础上的标准模式

对于较明确的、且具代表性的已知样品, 取相同样品同一特征标志的平均值、即可作为该岩相的标准模式。

但是, 该特征标志值应先模糊化, 即确定模型的模糊特征向量, 将一般特征向量转换成模糊特征向量可以选用一般数据正规化的方法。然而, 对于某些标志实际上只要数值超过或低于某一极限值即可确认代表某岩相, 因而可采用如下扩展的归一化式:

$$X'_{ij} = \begin{cases} 1 & X_{ij} \geq \alpha \\ & \beta < X_{ij} < \alpha \\ \frac{X_{ij} - \beta}{\alpha - \beta} & x_{ij} \leq \beta \end{cases}$$

式中, α , β 相应为上述上、下限域值 ($\beta < \alpha < X_{j\max}$, $\beta > X_{j\min}$); $X_{j\max}$, $X_{j\min}$ 分别为各样品中第 j 个标志的极长值与极小值。

所有样品 (包括未知样) 都须经上述模糊化计算。

2. 建立在聚类分析基础上的标准模式

当已知样品并不具有代表性或已知样数量不多, 这时难以直接给出标准模式, 采用已知、未知样品一起作聚类分析来建立相模式却更为客观。

在聚类方法中, 建立在模糊关系基础上的聚类是从相似矩阵出发的 (Kandel, 1982)。从相似矩阵到传递矩阵, 虽然传递闭包具有传递性可以分类, 但这种传递性通常是转换而来, 距相似矩阵较远, 难以形成一个合理的分类。实践已完全证明了这一点 (王硕儒等, 1987)。且 λ 的选定又要视分类结果而定。建立在自动迭代基础上的 F-PFS 聚类法 (张伟, 1987) 是在最小二乘基础上的最佳分类, 在很大程度上克服了系统聚类的缺点与不足, 这是一种在多次调用 Isodata 法聚类基础上确定最佳分类数及最佳分类的方法。

不过, 为获得较合理的分类, 还必须通过多次聚类结果的分析来调整特征标志和权重。对某些标志值作适当放大或缩小 (线性或非线性) 可以调整权重, 而标志的筛选则可以通过下文提出的单因素判对率的计算来进行。

事实上, 模糊聚类一般可能出现下述两种结果:

1、分类结果较好, 各类皆有意义。

2、虽然经过多次调整, 仍有个别类结果难以解释和利用。这时可对这些个别类的样品作进一步识别处理。那些较符合实际的类别可作为标准模型。

利用上述方法得到的标准类别, 都应具有较明确的岩相意义, 各类的聚类中心即代表了各标准类别 (岩相) 的模式。

3. 识别方法

聚类本身就是一种识别。然而, 当聚类中心个别类所属样品无法识别或有新的样品, 就必须将它们与标准模型 (聚类分析标准模型或给定岩相标准模型) 对比, 这时按贴适度大小作识别是一种常用的方法。

贴适度识别的关键是贴适度公式的选定, 以往的贴适度公式 (汪培庄, 1983), 由于没有考虑各特征标志在识别中的贡献大小不同这一情况, 因而不适用于岩相识别。为此, 笔者提出了加权逆距离贴适度的概念和公式。所谓加权逆距离是指: 由于贴适度识别中常用的参量是距离, 距离模糊化后, 其逆可定义为距离的逆运算, 而不同标志的逆距离又具有不同的权重。这样的贴适度具体定义如下:

$$S(X_i, Y_j) = \sum_{k=1}^n W_k (1 - D_{ijk})^p \quad (2)$$

式中 $D_{ijk} = |X_{ik} - Y_{jk}|$, $1 - D_{ijk}$ 表示该距离的“逆”; P 为调整未知样从属于各类的对比度。

无疑, 这样定义的贴适度是符合贴适度定义的一般公理的。

标志权重 W_k 的确定是实现贴适度计算的关键所在, 根据最大贴适原则, 按已知样的归属, 利用 (2) 式形成的不等式方程组可以求出权重, 但这样的权重具有一定范围, 是非确定性的。为此, 提出以下方法, 即根据单因素 (特征标志) 按择近原则判别已知样的判对率作为衡量该标志权重的尺度。

判对率:

$$W'_k = m_k / m \quad (3)$$

式中, m 为已知样品的总数, m_k 为按第 k 个标志判别的判对数, W'_k 经归一化处理即得第 k 个标志的权重 W_k 。

实践证明, 这种确定权重的方法既适用, 又简便。

二、鄂中拗陷海相碳酸盐岩岩相识别

鄂中拗陷位于扬子地台中部, 北与秦岭褶皱带相接, 南邻南华褶皱带, 东西分别与上、下扬子台地毗邻。该区内海相碳酸盐岩厚度大, 分布广, 岩相类型多, 并具广阔的油气远景。

1. 按聚类模式进行识别

笔者在拗陷内的松滋二叠系剖面集了 46 个样品。特征标志的选取立足于野外观察和室内薄片鉴定, 选用了颜色、岩石组分、古生物、颗粒组分、填隙物、沉积构造等。经聚类试算调整为 21 个或 12 个标志。

计算得 9 类为最佳分类。对应的最终确切分类是: 第一类 4 个样全为台洼相 (13、14、15、16 号样); 第二类 6 个样可解释为以后滩相为主 (4、19、20、39、40、44 号样); 第三类 8 个样全为台坡 B 相 (5、6、7、8、9、10、11、12 号样); 第四类 4 个样认为属局限台洼相 (24、41、42、43 号样); 其余五类难以确认, 这样, 确认前四类为沉积岩相的最佳分类, 各类的聚类中心构成了四种标准模式。

将余下的 24 个样与上四种模式进行贴适度识别, 结果如表 1 所示。聚类识别与贴适度识别的总判对率 $36 / 44 = 81.8\%$ (1、2 号样为潮坪相, 未参与统计)。22、25、26 号样虽属局限台洼相, 30 号属后滩相, 34 号为台坡 B 相, 它们未确切判对, 但计算数据表明, 很接近判对 (尤其 34 号样), 表明了它们的确切判属较模糊。考虑到这种因素, 上述判别效果

相当不错。

表 1 松滋剖面 24 个样 21 个变量的模式识别结果 (四模式)

Table 1 The results of fuzzy recognition with 21 variables in Songzi section (four patterns)

样品编号	(1)	(2)	(3)	(4)	确切岩相
29	0.4065	0.7893	0.7895	0.7377	(3)*
30	0.4150	0.7930	0.7954	0.7539	(3)-(2)
33	0.2887	0.7188	0.6824	0.7694	(4)*
34	0.3210	0.7644	0.7299	0.7274	(2)-(3)
35	0.3178	0.7473	0.7117	0.7206	(2)
1	0.5040	0.6446	0.6759	0.6440	(3)
2	0.3764	0.7301	0.7669	0.6867	(3)
3	0.3867	0.8063	0.7786	0.7213	(2)*
21	0.4420	0.7186	0.7382	0.7752	(4)
23	0.4370	0.7713	0.8320	0.8806	(4)*
27	0.3440	0.7826	0.7607	0.7844	(4)*
28	0.4177	0.7346	0.7689	0.8353	(4)*
36	0.4433	0.7944	0.8202	0.7828	(3)*
18	0.8348	0.4715	0.5257	0.5811	(1)*
22	0.3783	0.8073	0.7645	0.7945	(2)-(4)
25	0.3659	0.8111	0.7560	0.8032	(2)-(4)
26	0.3698	0.8171	0.7547	0.7995	(2)-(4)
31	0.3815	0.8574	0.7804	0.7281	(2)*
32	0.3984	0.8760	0.8121	0.7572	(2)*
37	0.3831	0.8470	0.7453	0.7188	(2)*
45	0.3639	0.7656	0.6389	0.7794	(4)*
46	0.3650	0.7577	0.6410	0.7816	(4)*
17	0.5287	0.7304	0.7863	0.8922	(4)*
38	0.5090	0.7458	0.7987	0.9028	(4)*

注: * 为判对样; (1) - 台洼相; (2) - 后滩相; (3) - 台坡 B 相; (4) - 局限台洼相

2.按给定模型进行识别

根据海底地形、水深、潮汐作用、水体能量、生物组合、沉积构造等特征,通过典型剖面的对比研究,可以将本区二叠系碳酸盐岩划分为六个相,按其离古海岸的远近,顺序如下:潮坪相、局限台洼相、台坡 A 相、台滩相、台坡 B 相、台洼相(其中台坡 A 相与台滩相在聚类时难以区分,故台二为一,称后滩相)。现取松滋剖面 30 个样形成六个相的标准模式。

将余下的 16 个样按(2)式识别,结果如表 2 所示,确切判对 11 个样,且有 1—2 个样接近判对。从表 2 还可看出,当把变量筛选为 12 个,标志集={颜色,方解石,白云石,石英、粘土矿物,古生物 2—6,硅质结核,瘤状体},其中古生物 2 为苔藓、海绵及类粪球粒;古生物 3 为钙球、钙藻、生物屑;古生物 4 为有孔虫、海百合;古生物 5 为腕足类、棘皮动物(海百合除外);古生物 6 为腹足类、瓣鳃类和介形虫。识别结果与 21 个变量所判相差甚微。

另取拗陷内的京山剖面 32 个样,以上述 12 个变量组成集与标准模式接贴程度(2)式计算,确切判别如表 3 所示,判对 27 个样。

表 2 松滋剖面 16 个样贴程度识别结果(六模式)

Table 2 The results of fuzzy recognition in Songzi section (six patterns)

样品编号	8	9	10	13	18	20	23	26	27	30	33	35	39	43	45	46
21 变量识别结果	(2)*	(2)*	(2)*	(3)*	(3)*	(6)-(4)	(5)*	(5)*	(6)	(6)	(6)	(6)*	(4)	(5)*	(5)*	(5)*
12 变量识别结果	(5)-(2)	(2)*	(2)*	(3)*	(3)*	(4)*	(5)*	(5)*	(6)	(6)	(6)-(4)	(6)*	(4)	(5)*	(5)*	(5)*

注: * 为判对样; (2) — 台坡 B 相; (3) — 台洼相; (4) — 台坡 A 相; (5) — 局限台洼相; (6) — 台滩相;

由此,对 48 个样总的剖对率为 79.2%。

表 3 京山剖面模糊数学识别结果

Table 3 The results of fuzzy recognition in Jingshan Section

样品编号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
模式识别结果	(4)*	(5)*	(5)*	(3)*	(2)	(5)*	(5)*	(2)*	(5)*	(5)*	(5)*	(5)*	(2)-(5)	(1)-(2)	(3)*	(3)*
综合评判结果	(4)*	(5)*	(5)*	(3)*	(5)*	(5)*	(5)*	(5)	(5)*	(5)*	(3)*	(5)*	(5)*	(5)*	(2)*	(3)*
样品编号	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
模式识别结果	(3)*	(3)*	(3)*	(3)*	(3)*	(2)*	(4)*	(4)*	(4)*	(4)*	(4)*	(4)*	(4)*	(3)	(6)-(2)	(3)*
综合评判结果	(3)*	(3)*	(3)*	(3)*	(5)	(6)	(4)*	(6)-(4)	(4)*	(4)*	(4)*	(4)*	(5)	(3)	(6)-(2)	(3)*

注:同表 2

为进一步验证方法的有效性,笔者使用模糊综合评判数学模型正问题解法

$$B = A \circ R$$

对京山剖面 32 个样识别, 其中 $A = (a_1, a_2, \dots, a_{12})$ 为 12 个标志的权重。模糊关系由二元对比排序加定量化的方法 (王硕儒等, 1989) 算得。算 (o) 取乘, 加运算, 获得的评判结果 B , 按最大隶属原则确切评定。

表 3 中也列出了模糊综合评判确定结果。对比可看出两点: 一是模式识别与综合评判绝大多数吻合; 二是前者结果稍优于后者。

综合上述讨论可知, 本文提出的具体的模糊模式识别方法是可行的、有效的。文对本区海相碳酸盐岩岩相的划分和识别提供了一个辅助的, 然而必要的定量化手段。

如果综合应用模式识别与综合评判的结果, 取其并集作为总的判别标准, 则确切判对率可高达 93.7%。

收稿日期: 1990 年 11 月 1 日

参 考 文 献

- (1) 王硕儒, 刘仲衡, 曹钦臣, 1987, 海洋通报, 3卷, 1期, 31-37页。
- (2) 王硕儒, 曹钦臣, 1989, 黄渤海海洋, 7卷, 2期, 27-33页。
- (3) 汪培庄, 1983, 模糊集合论及其应用, 上海科技出版社。
- (4) 吴宜志等, 1986, 美国勘探地球物理学家学会55届年会议文集 (中译), 598-601页, 石油工业出版社。
- (5) 张伟, 1987, 模糊数学, 3-4卷, 52-56页。
- (6) Kandel A. 1982, Fuzzy techniques in pattern recognition p.59-64. A Willey-Interscience Publication.

Fuzzy Pattern Recognition of Marine Carbonatite Facies

Liu Zhongheng Wang Shuoru Fan Dejian

(Ocean University of Qingdao)

Abstract

On the basis of fuzzy pattern recognition, a practical method of recognition of marine carbonatite facies is presented in this paper. The standard pattern can be constructed in the situation of cluster analysis and known samples. The cluster centre is obtained by the F-PFS method or the characteristic vector. Besides of the using of cluster analysis, the recognition of unknown sample mainly uses the calculation of similarity degree and the recognition is according to the nearest neighbour criterion. Then, a new similarity degree of complement distance with the different weights is presented in the paper. That is

$$S(X_i, Y_j) = \sum_{k=1}^n W_k (1 - D_{ijk})^p \quad (2)$$

Where the weight W_k is determined by the technique which is presented by the authors and is according to the rate of right judgement of the single factor.

Four facies can be identified by using fuzzy cluster analysis, other samples can be recognized by using the fuzzy recognition method. The percent of right judgment is 81.8%.

Six standard patterns were built based on the known facies samples of Shouzi Section. Both the samples of Jingshan Section and the rest samples of Shouzi Section were recognized with this model, the percent of right judgment is 79.2%.

The results of fuzzy recognition show that fuzzy cluster analysis and similarity degree are a kind of new effective method for the classification of carbonatite facies. Furthermore, cluster centres based on the fuzzy cluster or known facies samples could become the basis of later fuzzy recognition marine carbonatite facies.

* * * * *

经济实用

GIA-1 颗粒图象分析仪

测试原理 图象信号由摄像头变成电信号后, 未经数字处理, 而采用时间测量法, 将显示图象信息的监视屏“划分”成细小的点面阵, 利用人的智能作用, 在监视屏上识别图象边界及特征点, 以光笔、内游标无视差测量技术, 即人操作光笔和游标定位, 将图象特征座标实时输入计算机进行处理, 得到不同学科的分析结果

用途 广泛用于地质学、矿物学、土壤应用学及材料、冶金、电子、医药、机械等领域, 对显微图象、照(图)片、透明片等进行测试, 完成粒度分析、铸体薄片分析、裂缝测量、方位测试、物质含量统计、面孔率测试、裂变径迹测量、形状分析、粉末测定、精密检测及动态图象研究等, 即对点、面、角、维参数的测量和处理分析, 达到对图象信息的研究和开发利用。

配置及组成 摄像头、监视器、主机箱、计算机及打印机、用户自选显微镜(带摄影装置)。

仪器特点:

- 1. 积木式结构, 可灵活组建黑白和彩色系统, 并可随意选配。
- 2. 实用面广, 对复杂样品的测试较用自动图象仪灰度检测更具优越性, 测试数据的可信度及精确度高。
- 3. 全汉化交互式用户软件, 操作简便、实用, 不需二次开发, 并且系统测试功能软件模块全部开放

主要技术参数:

- 1. 系统分辨率 512 × 512
- 2. 光标配置 512 × 512 × 1Bit
- 3. 显微图象分辨率 0.63 / K (mm)
 $K = K_1 \cdot K_2$ (总放大倍数) K_1 为显微镜放大倍数 K_2 为摄像系统放大倍数, 4-5 倍
- 4. 测量方式: 光笔单点、连续, 游标单点
- 5. 测量误差 (相对大于 1/2 视域):
 线值相对误差 < 1%
 面积相对误差 (标准图形) < 1.5%
 角度绝对误差 < 1

系统参考售价: 人民币 3-5 万元, 随用户配置而定。

地址: 四川·成都 金河街 75 号 单位: 成都电子研究所 联系: 图象工程部 陈 实
邮编: 610013 电话: (028) 637880-29 电挂: 5939