

模糊聚类分析法在油气评价中的应用

刘世翔¹, 薛林福²

(1. 中海油研究总院, 北京 100027; 2. 吉林大学 地球科学学院, 长春 130061)

摘要: 文章介绍了模糊聚类分析在石油勘探决策中的应用, 选用了烃源岩 9 个评价参数, 利用模糊聚类方法对烃源岩进行分类评价, 并对评价过程进行了详细的论述。通过对评价结果与实际地质情况进行比较, 评价结果与实际情况完全符合, 证明模糊聚类方法用于油气资源评价是完全可靠的。

关键词: 油气评价; 模糊聚类; 烃源岩; 决策分析

中图分类号: O21; P618.13 **文献标识码:** A **文章编号:** 1001-1412(2011)03-0300-04

0 引言

油气资源地质评价研究的目标就是通过对油气在地下赋存规律的研究, 来圈定油气在地下的聚集场所, 预测探区未发现的远景资源量。该研究是一个复杂的过程, 涉及到制约油气成藏的多种因素, 每种因素的好坏将直接影响油气最终评价结果。

模糊数学的理论与方法引入石油勘探决策分析系统中不仅为石油勘探决策开辟了理论研究广阔的前景, 而且为石油勘探开发各阶段的决策分析增加了新的内容和手段^[1]。模糊聚类分析就是应用模糊数学的基本理论对受多因素影响的事物或者对象按照一定的指标特性进行分类的一种决策方法^[2]。本文在烃源岩评价中利用已经获得的烃源岩评价参数, 通过聚类方法对烃源岩参数进行分类, 从而确定烃源岩生烃能力的级别, 为下一步勘探确立目标。

1 模糊聚类原理和方法

1.1 方法概述

经典的聚类分析是一种数学分类方法, 其基础

是数理统计的多元分析方法^[2]。但在现实世界中, 一组事物根据其亲疏程度和相似性是否形成一个类群, 或一个事物是否属于某一个类别, 其界限往往是不分明的, 具有很大程度的模糊性, 模糊聚类正是刻画和解决这类聚类问题的数学方法。模糊聚类分析是依据客观事物间的特征、亲疏程度和相似性, 通过建立模糊相似关系对客观事物进行分类的数学方法^[4]。用模糊聚类分析方法处理带有模糊性的聚类问题要更为客观、灵活、直观和计算简捷。

1.2 方法原理

设样本空间 $X = \{X_1, X_2, \dots, X_N\}$, 其中 $X_i = (x_{i1}, x_{i2}, \dots, x_{ip})$, $x_{ip} \in R$, P 为特征指标数。模式聚类即将 X 划分成若干个子类: $F_1, F_2, \dots, F_K$, 使之满足 $F_i \cap F_j = \varnothing (i \neq j)$, $\bigcup_{i=1}^K F_i = F$; 同时满足在某种相似准则下类内距离最小, 而类间距离最大。

对于 X , 设所有可能的模糊 K -划分组成的集合为:

$$M_k = \left\{ U \in V_{P \times N} \mid \mu_{ik} \in [0, 1], \forall i, k; \sum_{k=1}^N \mu_{ik} = 1, \forall i; 0 \leq \sum_{k=1}^N \mu_{ik} \leq N, \forall k \right\} \quad (1)$$

式中, μ_{ij} 为第 i 个样本对第 k 个类的模糊隶属度, $V_{K \times N}$ 是 $k \times N$ 维的实矩阵集合:

收稿日期: 2010-03-17; 改回日期: 2011-01-26

基金项目: 国家油气专项项目(编号: XQ-2004-07)资助。

作者简介: 刘世翔(1978-), 男, 河北冀州人, 工程师, 博士, 主要从事油气资源评价方面的研究, 通信地址: 北京市东城区东直门外小街 6 号, 中海石油大厦 602 室; 邮政编码: 100027; E-mail: lsxiang123@163.com

$$U = \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_N \end{bmatrix} = \begin{bmatrix} U_{11} & U_{12} & \cdots & U_{1N} \\ U_{21} & U_{22} & \cdots & U_{2N} \\ \vdots & \vdots & & \vdots \\ U_{P1} & U_{P2} & \wedge & U_{PN} \end{bmatrix}。$$

在 R^P 空间中, d_{SK} 记表示样本 X_S 和 X_K 之间的加权距离, 定义为

$$d_{SK}^2 = \|X_S - X_K\|^2 = \sum_{p=1}^P w_p (x_{sp} - x_{kp})^2, 0 \leq w_p \leq 1, 0 \leq \sum w_p \leq 1 \quad (2)$$

式中, w_p 为第 p 个特征的权重。

定义聚类准则函数 J :

$$J(U, V, M) = \sum_{k=1}^n \sum_{i=1}^k (\mu_{ik})^\alpha (d_{ik})^2 \quad (3)$$

式中, $U = (\mu_{ik}) \in M_K$ 为 X 的一个模糊 K -划分; $V = (v_1, v_2, \dots, v_K) \in R^{K \times P}$ 为类中心矢量; d_{ik} 为样本 X_k 到中心 v_i 的距离; μ_{ik} 为样本 X_i 对类 k 的隶属度; α 为指数权重。

对样本空间 X , 按照类内距离最小, 类间距离最大的分类准则, 最佳模糊 K -划分应满足: $J = (U^*, V^*, M^*) = (J(U, V, M))_{\min}$ 即确定 U^*, V^*, M^* , 使 $J(U, V, M)$ 最小。

若 μ_{ik}^* 是最佳隶属度, 则可求出第 k 类中心 V_k , 可得 $V = (V_1, V_2, \dots, V_K)^T$, 即

$$V = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1N} \\ v_{21} & v_{22} & \cdots & v_{2N} \\ \vdots & \vdots & & \vdots \\ v_{k1} & v_{k2} & \wedge & v_{kp} \end{bmatrix}$$

根据 V_k 各分量的不同重要程度, 定义第 p 个分量的权重因数 w_p 为

$$w_p = \frac{\sum_{K=1}^K (v_{kp} - \bar{v}_p)^2}{\sum_{p=1}^P \sum_{K=1}^K (v_{kp} - \bar{v}_p)^2} \quad (4)$$

式中, $\bar{v}_p = \frac{1}{K} \sum_{K=1}^K v_{kp}$, w_p 反映各类中心 V_k 的第 P 个特征相对方差的大小, 即在聚类分析中的重要程度, X_i 到类中心 V_k 的加权模糊距离为

$$d_{ik}^2 = \|X_i - X_K\|^2 = \sum_{p=1}^P w_p (x_{ip} - v_{kp})^2 \quad (5)$$

1.3 基本步骤

(1) 归一化处理。在实际材料处理中, 由于烃源岩评价参数的量纲不同, 实际数据值量级相差较大, 如果直接使用原始数据计算, 就会突出数量级大的指标作用, 而对数量级小但敏感性较高的指标估计不足, 所以需要对原始数据做归一化预处理^[5]。

设 x_{ij} 为第 i 个样本点第 j 个参数的原始测量

值, 标准化后的 x'_{ij} 为

$$x'_{ij} = \frac{x_{ij} - (x_{ij})_{\min}}{(x_{ij})_{\max} - (x_{ij})_{\min}}, 1 \leq i \leq n$$

式中, n 为样本总数。

(2) 初始化。根据烃源岩划分标准, 给定聚类数 k , 迭代最大次数 $T_{\max}$, 误差精度 ϵ ; 用随机发生器产生初始隶属度 $\mu_{ik}(1)$, 先产生 $N \times K$ 个随机数 p_{ik} , 确定 $\mu_{ik}(1) = p_{ik} / \sum_{k=1}^K p_{ik}$ 。

(3) 自组织迭代。由 $V_K(t)$ 按式(1)计算 $w_p(t)$, t 为迭代次数;

按照步长 $\alpha (0 < \alpha \leq K/N)$ 和加权模糊距离 d_{ik}^2 修改最近节点及其邻域节点的类中心向量 $V_K(t+1) = V_K(t) + \alpha(X_K(t) - V_K(t))$;

计算误差函数值: $E_d = \sum \|V_K(t+1) - V_K(t)\|^2$, 若 $E_d \leq \epsilon$ 则则停止, 给出分类数。

2 烃源岩评价

本文利用模糊聚类分析法对某盆地烃源岩的生烃潜力进行了系统的评价, 并将模糊聚类结果与实际地质情况进行对比, 从而验证模糊聚类分析法在油气评价中的可行性。

2.1 烃源岩评价参数

通过多年的油气勘探实践, 人们总结出了描述烃源岩特征的多种参数, 这些参数可作为烃源岩综合评价的指标体系, 本文用于评价烃源岩生烃能力的参数包括暗色泥岩、暗色泥岩分数、有机质含量、镜质反射率、氯仿沥青“A”、干酪根类型。每一个参数称为烃源岩评价的单因素, 这些参数都是离散的点数据, 在进行模糊聚类之前还需要将这些点数据通过插值方法生成区域性的面数据。

2.2 评价参数赋值

单项成藏地质条件存在发生的概率, 取决于其子项地质因素的好坏^[6]。根据该盆地烃源岩的特征, 将烃源岩评价参数按照表1的取值标准进行打分, 可以对各项成藏地质因素赋予定量的评价值(0~1), 表示不同等级的优劣。

2.3 单因素评价数据合成

经过处理后的每个评价参数图层都成为一个赋值的格网文件, 但在进入模糊聚类分析系统进行聚类之前还必须将这些格网文件合成, 使之成为一个文本文件, 在这个文件中能表示所有图层中每个格

表1 烃源岩评价参数取值标准

Table 1 source rocks evaluation parameters value standard

参数名称	分 值			
	一级(1~0.75)	二级(0.75~0.5)	三级(0.5~0.25)	四级(0.25~0)
烃源岩厚度/m	>500	500~200	200~100	<100
烃源岩沉积相	深湖—半深湖	半深湖	半深湖—浅湖	浅湖、湖沼
泥岩分数/%	>30	30~15	15~5	<5
干酪根类型	I	I—II ₁	II ₂ —III	III
有机碳/%	>2.0	2.0~1.0	1.0~0.4	<0.4
氯仿沥青“A”/%	>0.2	0.2~0.1	0.1~0.05	<0.05
镜质体反射率 R_o	0.8~1.2	1.2~2.0	0.5~0.8	>2 或 <0.5

网点上的数据,这些数据构成了一个样本空间矩阵,数据合成示意图见图1。

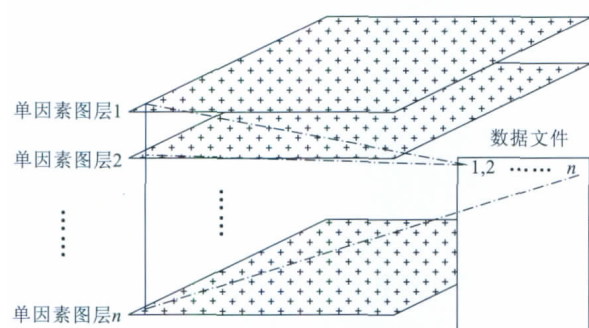


图1 数据合成示意图

Fig. 1 Schematic diagram showing data synthesis

2.4 模糊聚类的烃源岩评价

将合成后的数据文件输入到模糊聚类系统中,并对格网数据文件的每个格网结点逐个输入模糊聚类分析模块中进行分类。从而对烃源岩的生烃能力进行分类,进而确定分类的级别。

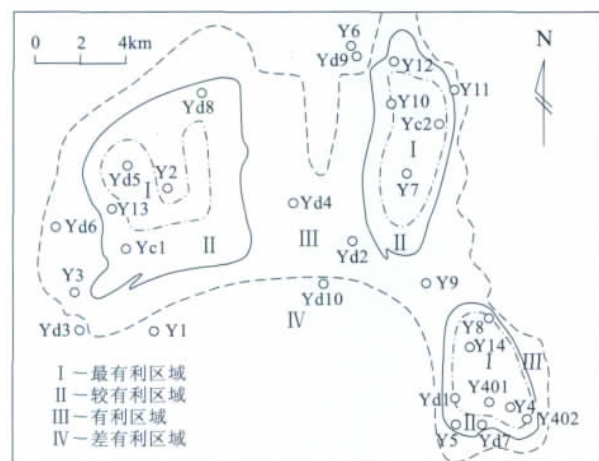


图2 模糊聚类结果图

Fig. 2 Diagram showing fuzzy clustering results

2.5 模糊聚类结果

在此基础上,对烃源岩参数进行模糊聚类分析,将该盆地烃源岩评价分成4种类型,其结果见图2。通过分析结果与实际地质进行对比,认为模糊聚类结果与实际完全相符。充分证明了模糊聚类方法对烃源岩进行评价是完全可行的。

3 结语

模糊聚类分析法是依据客观事物间的特征、亲疏程度和相似性,通过建立模糊相似关系对客观事物进行分类的数学方法。用模糊聚类分析方法分析客观事物则更为客观、灵活、直观,且计算更加简捷。

通过对烃源岩聚类结果分析表明,将模糊聚类分析法应用于油气评价中,得到的聚类结果能够有效反映出待聚类样本的真实情况,能为油气评价找出一个合理的分类体系,对油气勘探开发具有一定指导作用。

参考文献:

- [1] 王涛,王艳平,唐剑涛. 模糊数学及其应用[M]. 沈阳:东北大学出版社,2005:60-90.
- [2] 刘育骥,耿新宇,肖辞源. 石油工程模糊数学[M]. 成都:成都科技出版社,1994.
- [3] 诸克军,冯光华,冯刚项. 一种石油勘探决策的模糊聚类分析法[J]. 江汉石油学院学报,1999,21(3):16-20.
- [4] Ng M K, Wong J C. Clustering categorical data sets using tabu search techniques[J]. Pattern Recognition, 2002, 35: 2783-2790.
- [5] 郭秀娟,袁月,范小鹏. 模糊聚类算法分析及应用[J]. 吉林建筑工程学院学报,2009,26(4):79-81.
- [6] 夏明忠,张蓉,董广华,等. 油气资源评价及决策系统[M]. 北京:石油工业出版社,1993.

The application of fuzzy clustering analysis to oil and gas evaluation

LIU Shi-xiang¹, XUE Lin-fu²

(1. CNOOC Research Institute, Beijing 100027, China;

2. College of Earth Sciences, Jilin University, Changchun 130061, China)

Abstract: This paper introduces the application of fuzzy clustering analysis method to decision-making of the oil exploration. 9 evaluation parameters of source rocks are adopted and fuzzy clustering analysis applied to classification and evaluation of the source rocks. Comparison of the evaluation with the actual geology shows that the results are fully consistent with the actual conditions. Fuzzy clustering analysis is feasible to oil and gas resource evaluation.

Key Words: oil and gas evaluation; fuzzy clustering; source rocks; decision-making analysis

欢迎订阅 2012 年《地质找矿论丛》

《地质找矿论丛》为国家新闻出版总署批准,由中钢集团天津地质研究院有限公司主管、主办的地学科技期刊,创刊于 1986 年。

《地质找矿论丛》为中国科技核心期刊,被美国《化学文摘》(CA)、俄罗斯《文摘杂志》、《中国学术期刊文摘(中文版)》、《中国地质文摘》等著名文摘刊物收录,是《中国科技论文统计》、《中国学术期刊综合评价数据库》和《中国科学引文数据库》的来源期刊,期刊全文入编《中国学术期刊(光盘版)》、《中国知网》(CNKI)、《万方数据科技期刊群》、《中文科技期刊数据库》(维普)和《华艺 CEPS 中文电子期刊》等电子出版物及数据库,以多种媒体方式向读者提供服务。

《地质找矿论丛》主要报道矿产成矿理论与成矿预测、物质成分及综合利用、矿产地质勘查新技术新方法及其应用、地学信息技术、水文地质与工程地质、环境地质调查与治理、资源勘查工程、矿产品深加工技术、地质矿产技术经济等方面的科研成果、进展评介、研究简报,并不断开拓报道领域与深度。

《地质找矿论丛》刊物信息丰富、学科跨度大、兼顾理论与应用,是地学研究、矿产勘查、矿业开发领域专业人员和院校师生的良师益友。热忱欢迎地矿行业、地学院校、文献信息部门的单位和个人踊跃订阅并投稿。

《地质找矿论丛》为季刊;国际标准连续出版物号 ISSN 1001-1412,中国标准连续出版物号 CN 12-1131/P;铜版纸彩色印刷,国际开本(A4),每季末月 25 日出版;每期定价 20.00 元,全年定价 80.00 元;公开发行,国内邮发代号:6-283。

订阅办法:

(1)通过各地邮局订阅:

邮发代号:6-283,全国各地邮局均可办理订阅。

(2)通过“全国非邮发报刊联合发行部”订阅:

地址:天津市大寺泉集北里别墅 17 号,全国非邮发报刊联合征订服务部

邮政编码:300385

电话:022-23973378;23962479 传真:022-23973378

E-mail:LHZD@public.tpt.tj.cn “刊林网”网址:www.LHZD.com

(3)向《地质找矿论丛》编辑部订阅:

编辑部地址:天津市河东区友爱东道平房 4 号,中钢地质院《地质找矿论丛》编辑部

邮政编码:300181 电话:022-84283083 联系人:任光义

E-mail:luncong@163.com;luncong@yeah.net