

模糊信息处理方法在油气识别中的应用

钮永莉¹, 陈水利²

(1. 滁州职业技术学院, 安徽 滁州 239000; 2. 集美大学诚毅学院, 福建 厦门 361021)

[摘要] 通过对模糊理论的研究, 利用模糊信息处理方法完成对储集层的热解参数的处理, 建立了凝析油气层、气层、油层、差油层、油水同层、含油水层、水层、干层等八类储集层模式的特性指标的隶属函数、模糊识别矩阵以及最优模糊聚类中心矩阵。通过实验验证, 该方法具有较好的识别效果。

[关键词] 储集层; 热解参数; 模糊识别; 模糊聚类

[中图分类号] TP 391.4

Applied Research of Fuzzy Information Processing Methods in Distinguishing Oil-Gas Reservoir Strata

NIU Yongli¹, CHEN Shuli²

(1. Chuzhou Vocational and Technical College, ChuZhou 239000, China;

2. Chengyi University College, Jimei University, Xiamen 361021, China)

Abstract: Through the study of fuzzy theory, the fuzzy information processing method is used to deal with the pyrolysis parameters of the reservoir in this paper. Membership functions of some characteristic indexes, a fuzzy recognition matrix and an optimal fuzzy cluster center matrix of condensate oil-gas stratum, gas stratum, oil stratum, lack-oil stratum, oil-water zone, water stratum containing oil, water zone and dry stratum are established.

Keywords: reservoir stratum; pyrolysis parameters; fuzzy recognition; fuzzy cluster

0 引言

模糊数学诞生于1965年, 是由美国的L. A. Zadeh(扎德)教授所创立, 自模糊理论创建50余年来, 它的思想已广泛渗透到现代科学的许多分支, 在科技、工程等领域显示出了强大的生命力, 并在人文科学(经济、管理、社会等)领域里, 也获得了相当多的应用^[1]。包括模糊集合、模糊关系、模糊矩阵、模糊聚类、模糊变换在内的诸多方法都在信息处理领域发挥着重要作用, 很多传统工作都可以应用模糊处理的方法取得更好的效果, 如油田的地层油气储集层预测。

如何提高对含油气储集层识别的准确性, 一直是石油科技工作者关注的研究课题。目前, 识别含油气储集层常用的方法有测井解释法、地质数学法和类比法等^[2]。根据石油天然气测井结论的油气水层划分标准, 可将储集层分为八种类型: 凝析油气层(A_1)、气层(A_2)、油层(A_3)、差油层(A_4)、油水同层(A_5)、含油水层(A_6)、水层(A_7)和干层(A_8)^[2]。

在储集层含油气性的评价中, 热解与轻烃参数是非常重要的评价指标。通常是各储集层热解

[收稿日期] 2018-03-23

[修回日期] 2018-04-19

[基金项目] 安徽省高校自然科学重点项目(KJ2015A402); 安徽省职业与成人教育协会教育科研项目(azjxh17172)

[作者简介] 钮永莉(1977—), 女, 讲师, 从事模糊信息处理, 数据挖掘方向研究。

参数的取值范围进行分析归类。该取值范围由统计各类型储集层样本的热解参数值得到。由于储集层热解参数变化的复杂性,用常规的方法难以对含油气储集层进行准确划分,特别对不同类型含油气储集层之间的界限更是难以准确划分。事实上,各种不同类型含油气储集层之间的界限是不分明的。为了尽量提高对含油气储集层识别的准确性,本文将采用模糊信息处理的方法进行研究。

1 模糊识别模型的建立

1.1 特性指标的选取

由于有机碳含量 ($W_{\text{TOC}},\%$)、有效碳含量 ($W_{\text{CP}},\%$)、最高热解峰值 ($t_{\text{max}},\text{℃}$)、产油潜量 ($W_{(S_1+S_2+PG)}$)、游离烃含量 ($W_{(S_0+S_1)}$)、热解烃含量 (W_{S_2})、油质系数 (TPI)、游离烃指数 (W_{S_1}/W_{TOC})、热解烃指数 (W_{S_2}/W_{TOC}) 等 9 个热解参数在油气层的评价中起着重要的作用,故本文选取这 9 个热解参数作为模糊识别的特性指标。

1.2 储集层样本的选取

考虑由凝析油气层 (A_1)、气层 (A_2)、油层 (A_3)、差油层 (A_4)、油水同层 (A_5)、含油水层 (A_6)、水层 (A_7) 和干层 (A_8) 共 n 个储集层样本组成的样本集 $X = \{x_1, x_2, \dots, x_n\}$, 每个样本可由有机碳含量 ($W_{\text{TOC}},\%$)、有效碳含量 ($W_{\text{CP}},\%$)、最高热解峰值 ($t_{\text{max}},\text{℃}$)、产油潜量 ($W_{(S_1+S_2+PG)},\%$)、游离烃含量 ($W_{(S_0+S_1)},\%$)、热解烃含量 (W_{S_2})、油质系数 (TPI)、游离烃指数 (W_{S_1}/W_{TOC})、热解烃指数 (W_{S_2}/W_{TOC}) 等 9 个热解参数确定,构成一个特性指标向量,记第 i 个样本 x_i 的特性指标向量为 $x_i = (x_{i1}, x_{i2}, \dots, x_{i9})$ 。则样本集 X 可用特性指标矩阵表示为

$$X = \begin{pmatrix} x_{11} & \cdots & x_{19} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{n9} \end{pmatrix} = (x_{ij})_{n \times 9}, \quad (1)$$

其中, x_{ij} 表示为第 i 个样本的第 j 个特性指标值。

1.3 特性指标数据处理

由于这 9 个特性指标的量纲和数量级不相同,故为了消除特性指标单位的差别和特性指标数量级不同的影响,必须对各指标值施行数据规格化处理,从而使每一个指标值统一于某种共同的数值特性范围^[3-4]。

用极差规格化方法^[4]对式 (1) 进行数据处理,得到规格化矩阵为

$$R = \begin{pmatrix} r_{11} & \cdots & r_{19} \\ \vdots & & \vdots \\ r_{n1} & \cdots & r_{n9} \end{pmatrix} = (r_{ij})_{n \times 9}, \quad (2)$$

其中 $r_{ij} = (x_{ij} - x_{j\min}) / (x_{j\max} - x_{j\min})$ ($i = 1, \dots, n; j = 1, \dots, 9$), $x_{j\max}$ 、 $x_{j\min}$ 分别为 n 个样本中第 j 个特性指标的最大值和最小值。

1.4 构造模糊识别矩阵

将所选的 n 个样本按凝析油气层 (A_1)、气层 (A_2)、油层 (A_3)、差油层 (A_4)、油水同层 (A_5)、含油水层 (A_6)、水层 (A_7) 和干层 (A_8) 这 8 种储集层模式构造模糊识别矩阵为

$$U = \begin{pmatrix} u_{11} & \cdots & u_{18} \\ \vdots & & \vdots \\ u_{n1} & \cdots & u_{n8} \end{pmatrix}, \quad (3)$$

其中, u_{ih} 为第 i 个样本属于第 h 种模式的相对隶属度。

具体构造方法如下:

第一步: 分别建立每个样本的特性指标

每个样本的特性指标隶属于各种储集层模式中相应特性指标的隶属函数 $\mu_{A_{kj}}(x_{ij})$ ，即第 i 个样本的第 j 个指标取值为 x_{ij} 时判为第 k 类储集层模式 A_k 的第 j 个特性指标的资格。根据统计规律，可选取各种储集层模式的第 j 个特性指标的隶属函数为正态分布函数，即

这些分层的 9 个特性指标的隶属函数为：

$$\mu_{A_{kj}}(x_{ij}) = \exp[-(r_{ij} - \bar{r}_{A_{kj}})^2 / \sigma^2] \quad (j = 1, 2, \dots, 9),$$

其中： $\bar{r}_{A_{kj}} (k = 1, \dots, 8)$ 分别为选取的凝析油气层样本、气层样本、油层样本、差油层样本、油水同层样本、含油水层样本、水层样本和干层样本的第 j 个特性指标的均值，即 $\bar{r}_{A_{kj}} = \frac{1}{p} \sum_{i=1}^p r_{ij}$ ； σ_2 为各样本特性指标方差中的最大值，即 $\sigma^2 = \max\{\sigma_{A_{1j}}^2, \sigma_{A_{2j}}^2, \dots, \sigma_{A_{8j}}^2\}$ ，且 $\sigma_{A_{1j}}^2, \sigma_{A_{2j}}^2, \sigma_{A_{3j}}^2, \sigma_{A_{4j}}^2, \sigma_{A_{5j}}^2, \sigma_{A_{6j}}^2, \sigma_{A_{7j}}^2, \sigma_{A_{8j}}^2$ 分别为选取的凝析油气层样本、气层样本、油层样本、差油层样本、油水同层样本、含油水层样本、水层样本和干层样本的第 j 个特性指标的方差，即 $\sigma_{A_{kj}}^2 = \sum_{i=1}^p (r_{ij} - \bar{r}_{A_{kj}})^2 / p$ ； A_k 表示第 k 类储集层模式； p 表示该储层类型选取的样本数。

第二步：计算各储集层模式中特性指标的权重

考虑到每个特性指标对储集层识别的作用大小不同，故应计算各储集层模式中特性指标的权重。设第 k 类储集层模式中第 j 个特性指标的权重 w_{kj} ，并记 $w_k = (w_{k1}, w_{k2}, \dots, w_{k9})$ ， $k = 1, 2, \dots, 8$ 。这里 w_{kj} 可按如下方法近似计算，首先计算 $w_{kj}^* = 1 - (\sigma_{A_{kj}}^2 / \bar{r}_{A_{kj}})$ ($k = 1, 2, \dots, 8; j = 1, 2, \dots, 9$)，然后进行数据归一化，即得第 k 个储集层模式中第 j 个特性指标的权重 w_{kj} 。

第三步：求每个样本的相对隶属度

求 $u_{i1}, u_{i2}, \dots, u_{i8}$ ，即求第 i 个样本分别隶属于凝析油气层 (A_1)、气层 (A_2)、油层 (A_3)、差油层 (A_4)、油水同层 (A_5)、含油水层 (A_6)、水层 (A_7) 和干层 (A_8) 的相对隶属度。计算公式为：

$$u_{ik} = \sum_{j=1}^9 w_{kj} \mu_{A_{kj}}(x_{ij}) \quad (i = 1, 2, \dots, n; k = 1, 2, \dots, 8),$$

从而得到 n 个样本的模糊识别矩阵 $U = (u_{ik})_{n \times 8}$ 。

1.5 求最优模糊聚类中心矩阵

$$S = \begin{pmatrix} s_{11} & \dots & s_{19} \\ \vdots & & \vdots \\ s_{81} & \dots & s_{89} \end{pmatrix} = (s_{kj})_{8 \times 9}, \tag{5}$$

建立目标函数

$$\min \{ J(U, S) = \sum_{i=1}^n \sum_{k=1}^8 \{ u_{ik} \left[\sum_{j=1}^9 [w_{kj}(r_{ij} - s_{kj})] p \right]^{1/p} \}^2 \}, \tag{6}$$

其意义为，聚类中心样本对于各类别加权广义欧氏距离平方和最小。当 u_{ik}, w_{kj} 为已知时，目标函数式 (6) 中的 s_{kj} 为未知数，此时目标函数式可写成

$$\begin{aligned} \min \{ J(s_{kj}) \} &= \sum_{k=1}^8 \min \left\{ \sum_{i=1}^n \left\{ u_{ik}^2 \sum_{j=1}^9 [w_{kj}(r_{ij} - s_{kj})]^2 \right\} \right\}, \\ dJ(s_{kj}) / ds_{kj} &= 2 \sum_{i=1}^n u_{ik}^2 w_{kj}^2 s_{kj} - 2 \sum_{i=1}^n u_{ik}^2 w_{kj}^2 r_{ij} = 0, \\ s_{kj} &= \left(\sum_{i=1}^n u_{ik}^2 w_{kj}^2 r_{ij} \right) / \left(\sum_{i=1}^n u_{ik}^2 w_{kj}^2 \right). \end{aligned} \tag{7}$$

由于式 (7) 中求和循环变量 i 与 w_{kj}^2 无关，显然， w_{kj}^2 可以从式 (7) 中消去，即

$$s_{kj} = \left(\sum_{i=1}^n u_{ik}^2 r_{ij} \right) / \left(\sum_{i=1}^n u_{ik}^2 \right), \tag{8}$$

则可求得各类型储集层的最优模糊聚类中心矩阵 S 。

1.6 利用最优聚类中心矩阵 S 对样本进行识别, 检验 S 的准确性

判别原则如下, 已知所求最优聚类中心矩阵 $S = (S_1, S_2, \dots, S_8)^T$, 其中 $S_k = (s_{k1}, s_{k2}, \dots, s_{k9})$, 对任意 $R_i \in \mathbf{R}$, $R_i = (r_{i1}, r_{i2}, \dots, r_{i9})$, 如果

$$\|R_i - S_t\| = \min_{1 \leq k \leq 8} \{\|R_i - S_k\|\}, \quad (9)$$

其中 $\|R_i - S_k\| = [\sum_{j=1}^9 w_{kj} (r_{ij} - s_{kj})^2]^{1/2}$, 则将第 i 个样本划归为第 t 类模式。

2 应用实例

本文选取某油田五类储集层即油层、油水同层、含油水层、水层和凝析油气层中 80 个储集层样本 $X = (x_1, x_2, \dots, x_{80})$ 进行识别 (该样本集已包含传统地化录井软件的分类结果), 每个样本由有机碳含量 ($W_{\text{TOC}}, \%$)、有效碳含量 ($W_{\text{CP}}, \%$)、最高热解峰值 ($t_{\text{max}}, \text{°C}$)、产油潜量 ($W_{(S_1+S_2+PG)}$)、游离烃含量 ($W_{(S_0+S_1)}$)、热解烃含量 (W_{S_2})、油质系数 (TPI)、游离烃指数 (W_{S_1}/W_{TOC})、热解烃指数 (W_{S_2}/W_{TOC}) 等 9 个热解参数确定^[6-8]。按照上述算法求解这 5 类储集层的特性指标权重和最优聚类中心, 结果分别如表 1 和表 2 所示。

表 1 各类别指标权重

Tab. 1 Index weight of each category

类别 Category	有机碳 含量 $W_{\text{TOC}}/\%$	有效碳 含量 $W_{\text{CP}}/\%$	最高热 解峰值 $t_{\text{max}}/\text{°C}$	产油潜量 $W_{(S_1+S_0+PG)}$ /%	游离烃 含量 $W_{(S_0+S_1)}$ /%	热解烃 含量 $W_{S_2}/\%$	油质 系数 TPI	游离烃 指数 W_{S_1}/W_{TOC}	热解烃 指数 W_{S_2}/W_{TOC}
油层 Oil layer	0.1092	0.1070	0.1184	0.1075	0.1101	0.1064	0.1159	0.1159	0.1095
油水同层 Oil-water layer	0.1081	0.1074	0.1118	0.1078	0.1092	0.1060	0.1179	0.1151	0.1167
含油水层 Oil-bearing water layer	0.1112	0.1106	0.1126	0.1107	0.1106	0.1119	0.1121	0.1081	0.1123
水层 Water layer	0.1110	0.1120	0.1125	0.1121	0.1121	0.1112	0.1093	0.1077	0.1122
凝析油气层 Condensate oil and gas layer	0.1124	0.1064	0.1099	0.1067	0.1056	0.1124	0.1180	0.1102	0.1184

表 2 各类储集层的聚类中心

Tab. 2 Cluster centers of various reservoirs

类别 Category	有机碳 含量 $W_{\text{TOC}}/\%$	有效碳 含量 $W_{\text{CP}}/\%$	最高热 解峰值 $t_{\text{max}}/\text{°C}$	产油潜量 $W_{(S_1+S_0+PG)}$ /%	游离烃 含量 $W_{(S_0+S_1)}$ /%	热解烃 含量 $W_{S_2}/\%$	油质 系数 TPI	游离烃 指数 W_{S_1}/W_{TOC}	热解烃 指数 W_{S_2}/W_{TOC}
油层 Oil layer	0.474	0.614	0.137	0.622	0.605	0.446	0.606	0.574	0.316
油水同层 Oil-water layer	0.382	0.437	0.115	0.441	0.400	0.365	0.502	0.433	0.290
含油水层 Oil-bearing water layer	0.231	0.302	0.091	0.308	0.305	0.206	0.625	0.505	0.249
水层 Water layer	0.250	0.303	0.124	0.310	0.292	0.235	0.626	0.483	0.246
凝析油气层 Condensate oil and gas layer	0.138	0.213	0.247	0.220	0.234	0.134	0.807	0.530	0.164

利用式 (9) 通过 Matlab 编程对结果进行判别, 得到各储集层样本识别准确率为: 油层 85.6%, 油水同层 78.1%, 含油水层 76.3%, 水层 79.1%, 油气层 81.2%。该结果比传统的地化录井方法有了很大的提高, 取得了较好的实际效果。

4 结论

本文运用模糊信息处理方法, 通过对储集层的地球化学热解参数的分析, 给出对含油气储集层的一种有效识别方法。通过实例分析, 本文给出的模糊信息处理方法比测井解释法、地质数学法、类比法等识别的准确性有明显提高。由于模糊识别方法是一种多参数的模式识别方法, 识别准确率与模糊识别矩阵有关。因此, 可以通过改进初始模糊识别矩阵以应对不同环境的油气识别, 扩展该方法的使用范围。

[参 考 文 献]

- [1] 陈水利, 李敬功, 王向公. 模糊集理论及其应用 [M]. 北京: 科学出版社, 2005: 36-70.
- [2] 李阳, 吴胜和, 侯加根, 等. 油气藏开发地质研究进展与展望 [J]. 石油勘探与开发, 2017, 44(4): 569-579.
- [3] 顾丽娟, 孙慧静, 尹付梅. 基于模糊聚类分析的鱼雷战术性能综合评价 [J]. 海军航空工程学院学报, 2012, 27(6): 693-696.
- [4] 傅雨佳, 杨慧中, 陶洪峰. 基于最大矩阵元法确定聚类数的软测量建模方法 [J]. 计算机与应用化学, 2012, 29(7): 877-880.
- [5] 刘红霞. 基于模糊数学的松南地区诸盆地油气选区评价 [J]. 矿物岩石地球化学通报, 2017, 36(5): 807-812.
- [6] 余鸿飞. 塔里木盆地储层测井识别方法研究: 塔中奥陶系、天山南白垩系测井分析 [D]. 成都: 成都理工大学, 119-157.
- [7] 倪艳南. 泌阳凹陷深凹区油气层录井识别技术研究 [J]. 西部探矿工程, 2012(6): 115-119.
- [8] 滕工生. 复杂岩性油藏录井精细解释方法 [D]. 大庆: 东北石油大学, 2011: 90-113.

(责任编辑 朱雪莲 英文审校 黄振坤)