

基于数据挖掘的M地区低渗透储层评价

陈玉雪,常安定,强玲娟

(长安大学理学院,西安 710064)

摘要: 文章以M地区为研究对象,针对该区实际地质情况,基于孔隙度、渗透率等测井资料,运用多元统计中的R型因子分析、判别分析及数据挖掘中的聚类分析,对储层进行合理的分类,得出储层空间的展布规律,并建立定量的评价标准。结合实际产能资料验证,认为该方法在储层评价中是可行的。

关键词: R型因子分析;Fisher判别;凝聚层次聚类;储层评价;鄂尔多斯

中图分类号: P618.13,TE348 **文献标识码:** A

0 引言

储层研究的目的是对储层做出符合地质实况的分类与评价^[1-3],故首要任务就是对储层做出合理的分类评价,它的结果能够直接指导油气的勘探与开发。因此,储层评价工作一直是国内外研究者关注的焦点。储层分类的关键是合理选择评价参数及方法,目前研究方向主要集中在运用各种数学方法、地质学与实验方法等对渗透率和孔隙度等储层物性参数进行评价,其中包括:地质经验法、模糊数学法,权重分析法以及各种地震方法等等^[1]。根据各个地区不同的地质特征,每种评价方法均有各自不同的优缺点。

本次工作在对M地区的研究中发现,M地区的储层非均质性较强,溶蚀孔洞发育较多,现有方法对该区评价时其结果不太理想。由此,本研究以M地区的地质资料为例,引入多元统计分析中的R型因子分析优化评价参数,然后使用数据挖掘中的凝聚层次聚类算法对储层分类,并建立评价标准,从而得到更精准的评价结果。相比较而言,比传统定性分类更加合理、全面^[4-5]。

1 方法和原理

因子分析和判别分析都是多元统计中的典型方法^[6]。因子分析的目的是降维和简化数据,它通过研究多个变量间的内部依赖关系,用少数几个能反映原来众多变量主要信息的‘因子’来表示其基本的数据结构。原始变量是可观测的显在变量,而因子是不可观测的潜在变量^[7]。因子分析在此用来确定变量权重。

判别分析是利用已经得到的变量数据,找到一种判别函数,使得此函数具有某种最有性质,能够把属于不同类别的样本点尽量区别开来^[6]。判别分析在此用来建立评价标准并验证评价结果。

数据挖掘的目的是从大量、随机的数据中通过算法探寻隐藏的潜在信息的过程,主要包括分类、聚类、关联分析与异常检测^[8]。其中,聚类分析是数据挖掘中的一个重要方法,它是一种在没有任何先验信息条件下,将现有的无标记数据进行归类的数据分析过程。其目标是,组内的对象相互之间是相似的,而不同组中的对象是不相关的。组内的相似性越大,组间差距越大,聚类效果就越好^[9]。

收稿日期: 2016-04-28; 收稿日期: 2016-07-21; 责任编辑: 王传泰

基金项目: 陕西省科技计划项目(编号:2015JM1022)资助。

作者简介: 陈玉雪(1970—),女,在读研究生,主要从事统计最优化方向研究工作。通信地址:陕西省西安市南二环中段,长安大学理学院;邮政编码:710064;E-mail:332575096@qq.com

1.1 R 型因子分析

设有 n 个样本,每个样本有 m 项参数,则原始数据集可用矩阵表示为:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix}$$

(1)将样本标准化,去除原始量纲及异常点。

(2)判断原始变量是否适合于因子分析(求变量间的相关系数矩阵,若矩阵中大部分相关系数均大于 0.3,且通过球形检验即可)。

(3)将矩阵线性变化为公因子 $F_i (i < m)$ 的表达式,求 F_i 的特征根 $\lambda_i (\lambda_1, \lambda_2, \cdots, \lambda_m > 0)$ 及其对应的标准正交化特征向量 e_{ij} 。其中,公因子 F_i 的方差贡献 $g_i^2 = \sum_{j=1}^m r_{ij}^2$,而 $r_{ij} = \sqrt{\lambda e_{ij}}$ ($i, j = 1, 2, \cdots, m$); 方差贡献率 $T = \frac{\lambda_i}{\sum_{i=1}^m \lambda_k} (i = 1, 2, \cdots, m)$;最后选择 g_i^2 或 $T > 0.8$ 的公因子。

(4)因子旋转。

(5)计算公共因子变量得分^[9-10]。

1.2 凝聚层次聚类算法

层次聚类常常使用称作树状图的类似于树的图显示。凝聚层次聚类算法的基本思想是:将数据对象看成一棵树状图,运用递归的方法发现潜在的簇,在自下而上的凝聚模型中,从个体点作为本身的簇开始,相继合并两个最接近的簇,直到只剩下一个簇^[8]。

算法描述如下:

第一步:根据需要,计算邻近度矩阵。

第二步:repeat。

第三步:合并最接近的两个簇。

第四步:更新邻近性矩阵,以反映新的簇与原来的簇之间的邻近性。

第五步:until 只剩下一个簇。

1.3 Fisher 判别

从 m 个总体中抽取具有 n 个指标的样品观测数据,然后借助方差分析的思想构造一个线性判别函数:

$$U(X) = u_1 X_1 + u_2 X_2 + \cdots + u_n X_n = u'X \quad (1)$$

其中,系数 $u = (u_1, u_2, \cdots, u_n)$ 确定的原则是使得总体之间区别最大,每个个体内部离差最小。有了线性判别函数后,对于一个新的样品,将它的 n 个指标值代入(1)式中求出 $U(X)$ 值,然后根据判别一定的规则,即可判别出新的样品属于哪个总体^[6,11]。

2 储层类型划分

2.1 地质背景

鄂尔多斯盆地是一个位于中国西部,在华北地台上发育起来的多旋回中、新生代沉积盆地。盆地具有构造稳定,整体抬升、持续下降、地层平缓的特征^[12]。在晚三叠世延长期,发育了大型的鄂尔多斯湖盆,成藏了大量的油气资源^[9],其中“长 4+5—长 2”是湖盆的萎缩期,面积逐渐变小^[12]。

M 地区位于鄂尔多斯盆地伊陕斜坡中部,地层为一个单斜,倾角不足 1° 。M 地区长 2 期以三角洲平原分流河道沉积为主,尤其是长 21 期,三角洲发育进入鼎盛时期,广泛发育的河道砂体为油气聚集提供了必须的储集空间,由于岩性的横向变化和储集砂体沉积之后的差异压实作用形成构造—岩性圈闭。

三叠纪末期,鄂尔多斯盆地受印支运动的影响整体抬升。地表遭受长时期的风化剥蚀,形成起伏不平、沟谷纵横的古地貌景观。研究区位于前侏罗纪甘陕古河谷南侧的子午岭斜坡带。延长组长 2 保存完整,长 1 也不同程度的有所保存,保存厚度可达 30~50 m,有利于油气成藏。

2.2 参数优选和因子分析

在 4 口取心井产段选,取样本 32 个,优选了孔隙度、渗透率、非汞饱和度、中值压力、结构系数和进汞迂曲度作为分析指标,并对数据进行因子分析,分析结果见表 1。

从表 1 可知,本例中有两个因子对应的特征值大于 1,因此应提取相应的两个公因子,且从中可以看出,前两个因子已经可以解释 79.093% 的方差。

从表 2 可以看出,第一个公因子在指标 x_1, x_2, x_6 上有较大载荷,说明这三个指标有较强的相关性,可以归为一类;其他指标归为另一类。

根据表 2 可得到旋转后的因子得分表达式为:

$$F_1 = 0.696x_1 + 0.658x_2 + 0.834x_3 + 0.923x_4 + 0.59x_5 + 0.762x_6$$

$$F_2 = 0.102x_1 + 0.217x_2 + 0.378x_3 + 0.29x_4 + 0.946x_5 + 0.035x_6$$

式中: x_1 为孔隙度(%); x_2 为渗透率($\times 10^{-3} \mu\text{m}$); x_3 为非汞饱和度(%); x_4 为中值压力(MPa); x_5 为结构系数; x_6 为进汞迂曲度。

表 1 各公因子方差贡献表

Table 1 variance contribution for each common factor

因子	起始特征值			提取平方和载入			旋转平方和载入		
	特征根	方差 贡献率/%	累加方差 贡献率/%	特征根	方差 贡献率/%	累加方差 贡献率/%	特征根	方差 贡献率/%	累加方差 贡献率/%
1	3.049	50.820	50.820	3.049	50.820	50.820	3.048	50.803	50.803
2	1.096	28.273	79.093	1.096	28.273	79.093	1.097	28.291	70.093
3	0.866	4.428	83.521						
4	0.517	8.619	92.140						
5	0.398	6.636	98.776						
6	0.073	1.224	100.000						

注:提取方法为主体元件分析。

表 2 旋转后因子载荷阵

Table 2 The factor loading matrix after rotating

	元件	
	1	2
VAR00001	0.696	0.102
VAR00002	0.658	0.217
VAR00003	-0.834	0.378
VAR00004	-0.923	0.029
VAR00005	0.059	0.946
VAR00006	0.762	0.035

2.3 凝聚层次聚类

根据 2 个公共因子的得分对该区储层进行聚类分析,可得到分类系谱图 1。由图 1 可知:当距离系数为 1.5 时,所有样品明显的分为 4 类(A、B、C、D)。

结合系谱图,对划分的四类储层进行分析,其结果如表 3 所述。

为验证分类的合理性,与现场储层分类结果对比发现,A 类为好储层,B 类为较好储层,C 类为较差储层,D 类为非储层。

2.4 储层类型判别

确定了储层划分的 4 个类型以后,将标准样品按照以上四类储层类型分组,建立判别需要的输入文件,利用 Fisher 判别分析法建立每一类储层的判别函数模型,判别函数分别为:

$$F_1 = -64.98 - 1.680x_1 + 0.656x_2 + 0.877x_3 + 0.097x_4 + 0.017x_5 + 0.579x_6$$

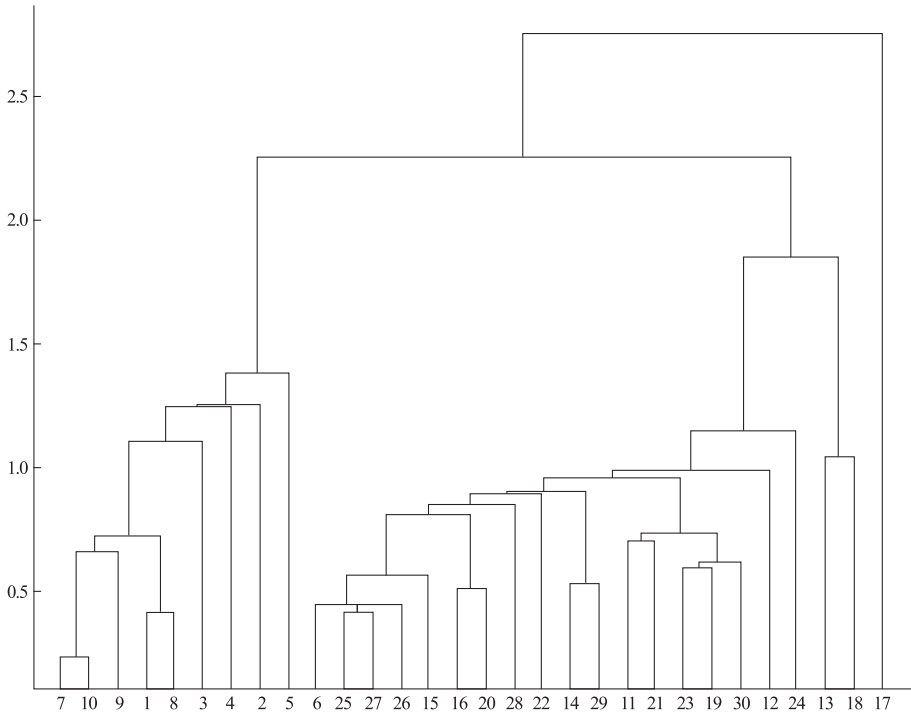


图 1 M 地区储层分类系谱图(横坐标为样本,纵坐标为距离系数)

Fig. 1 Family tree of reservoir classification in area M

表 3 储层分类参数统计

Table 3 Parametric statistics of reservoir classification

储层类别	物性分析			毛管压力特征参数			铸体		个数	
	孔隙度/%	渗透率/ 10 ⁻³ μm ²	中值压力/ MPa	非汞饱和度/ %	结构 系数	进贡 迂曲度	面孔率/%	孔隙组合		
A	平均值	12.76	9.9	1.65	23.11	3.07	2.55	8.5	溶孔-粒间孔	18
	范围值	16.4~19.1	6.11~20.6	1.3~2.3	19.7~26.2	1.77~3.70	2.2~2.7			
B	平均值	16.0	2.8	3.0	22.80	3.44	2.60	4.5	粒间孔-溶孔	9
	范围值	15.2~19.0	1.1~4.7	2.2~3.6	18.1~24.8	2.43~3.89	2.45~2.71			
C	平均值	16.5	2.4	5.1	25.48	2.64	2.46	4.5	粒间孔-溶孔	2
	范围值	14.6~18.1	0.75~4.3	4.0~6.4	22.97~27.91	2.39~2.99	2.09~2.79			
D	平均值	12.8	0.29	9.5	36.3	3.08	2.04		少量可见孔	1
	范围值									

表 4 正 2 井区相对渗透率分析数据表

Table 4 Analysis of relative permeability data from well +2

井号	层位	束缚水饱和度/%	最大含水饱和度/%	无水期驱油效率/%	等渗点饱和度/%	最终驱油效率/%	空气渗透率/ $10^{-3} \mu\text{m}^2$	孔隙度/%
5028	长 ₂ ¹	34.6	67.2	23.8	58.0	49.8	1.94	10.9
5032	长 ₂ ¹	35.0	67.0	23.5	58.0	49.3	0.88	14.3
5033	长 ₂ ¹	36.2	68.7	23.7	59.1	50.9	2.81	14.0
	平均值	35.3	67.6	23.7	58.4	50.0	1.88	13.1

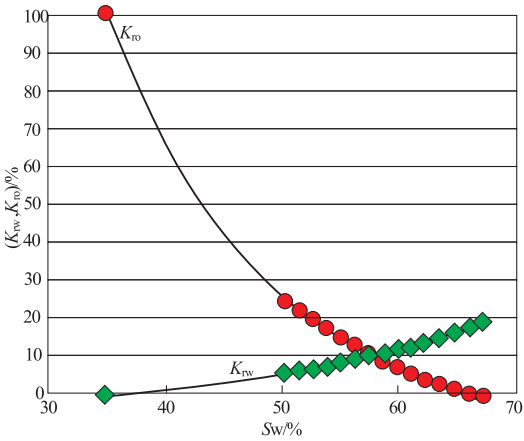


图 2 5032 井相对渗透率曲线
Fig. 2 Permeability saturation curve of well 5032

$$F_2 = -28.767 - 0.867x_1 + 1.155x_2 + 0.357x_3 + 0.038x_4 + 0.054x_5 + 0.698x_6$$

式中: x_1 为孔隙度(%); x_2 为渗透率($\times 10^{-3} \mu\text{m}^2$); x_3 为非汞饱和度(%); x_4 为中值压力(MPa); x_5 为结构系数; x_6 为进汞迂曲度。

从中可以得到,孔隙度和第一个判别函数关系较密切,渗透率和第二个判别函数关系较密切。实际上这两个函数式计算的是各观测值在每个维度上的坐标,这样可以通过这两个式子求出各个样品观测值的具体空间位置。

通过验证,随机选择的样本点与上述聚类分析

结果吻合率达到 94% 以上,说明所建立的 Fisher 判别函数与其评价指标的分类标准符合了精度上的要求。通过 spss 判别分析,可计算其他井样品所属储层类型。

2.5 应用效果分析

通过对正 2 井区 5032 井岩心样品模拟地层条件进行的储层润湿性试验,结果为亲水储层。水润湿指数为 0.646,油润湿指数为 0。相对渗透率分析见图 2、表 4 所述,等渗点含水饱和度 58.4%,束缚水饱和度为 35.3%。水驱油试验,无水驱油效率为 25%,最终驱油效率为 43%。渗流分析成果表明,正 2 井区长 2 油藏的储层性质对油田开发较为有利。

综上所述,该储层分类评价标准的结论与实际生产情况吻合,这一新的评价标准达到了较高的精度,适用于该区的储层评价分析。

3 结语

采用因子分析优化评价参数,解决了参数多、数据不规整的问题;通过凝聚层次聚类分析,将该区储层分为四类,每类储层各参数值均是规律递变;最后,通过 Fisher 判别函数建立了储层分类的综合判别函数表达式,以及定量的分类标准。

参考文献:

- [1] 陈欢庆, 丁超, 杜宜静, 等. 储层评价研究进展[J]. 地质科技情报, 2015, 34(5): 66-74.
- [2] 杨俊杰. 鄂尔多斯盆地构造演化与油气分布规律[M]. 北京: 石油工业出版社, 2002.
- [3] 于兴河. 油气储层地质学基础[M]. 北京: 地质出版社, 2009: 362-382.
- [4] 黄渊, 廖明光, 李斌, 等. 数据挖掘技术在碳酸盐岩储层评价中的应用[J]. 特种油气藏, 2014(5): 37-42.
- [5] 谭锋奇, 李洪奇, 孟照旭, 等. 数据挖掘方法在石油勘探开发中的应用研究[J]. 石油地球物理勘探, 2010, 45(1): 85-91.
- [6] 张润楚. 多元统计分析[M]. 北京: 科学出版社, 2010.
- [7] 张林泉. 基于因子分析的应用研究[J]. 哈尔滨师范大学学报(自然科学版), 2009, 25(5): 60-63.
- [8] 范明, 范宏建, 等. 数据挖掘导论[M]. 北京: 人民邮电出版社, 2014.
- [9] 李文厚, 庞军刚, 曹红霞, 等. 鄂尔多斯盆地晚三叠世延长期沉积体系及岩相古地理演化[J]. 西北大学学报(自然科学版), 2009(3): 501-506.
- [10] 吕红华, 任明达, 柳金诚, 等. Q型主因子分析与聚类分析在柴达木盆地花土沟油田新近系砂岩储层评价中的应用[J]. 北京大学学报(自然科学版), 2006, 42(6): 740-745.
- [11] 付殿敬, 徐敬领, 王贵文. 基于Q型聚类分析和贝叶斯判别算法研究储层分类评价[J]. 科技导报, 2011, 29(3): 29-33.
- [12] 刘池洋, 赵红格, 桂小军. 鄂尔多斯盆地演化-改造的时空坐标及成藏(矿)响应[J]. 地质学报, 2006, 80(5): 617-633.

The data mining-based evaluation of low permeability reservoir in area M

CHEN Yuxue, CHANG Anding, QIANG Lingjuan

(Changan University Faculty of Science, Xi'an 710064, China)

Abstract: According to practical geological situation, logging data, such as porosity, permeability of area M multivariate statistical R factor analysis, discriminant analysis and cluster analysis for data mining are used to make reasonable classification of reservoirs, then is summarized the spatial distribution pattern of reservoirs and set up evaluation criterion. The actual production data show that the method is feasible.

Key Words: type R factor analysis; Fisher discrimination; condensation hierarchical clustering; reservoir evaluation; Erduosi